

**UNIVERSIDADE ESTADUAL PAULISTA JULIO DE MESQUITA FILHO - UNESP -
FACULDADE DE ENGENHARIA DE BAURU - PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA DE PRODUÇÃO**

ALEXANDRE FRUCTUOSO DA COSTA

**PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS POR MEIO DE REDES
NEURAIS DINÂMICAS E PROCESSOS DE TRANSFORMAÇÃO DE DADOS: UMA
ABORDAGEM EMPÍRICO-COMPARATIVA**

**BAURU (SP)
2012**

ALEXANDRE FRUCTUOSO DA COSTA

**PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS POR MEIO DE REDES
NEURAIS DINÂMICAS E PROCESSOS DE TRANSFORMAÇÃO DE DADOS: UMA
ABORDAGEM EMPÍRICO-COMPARATIVA**

Dissertação apresentada à Faculdade de
Engenharia da Universidade Estadual
Paulista “Júlio de Mesquita Filho” para
obtenção do título de Mestre em
Engenharia de Produção

Orientador: Prof. Dr. Antonio Fernando Crepaldi

**BAURU (SP)
2012**

Costa, Alexandre Frutuoso.

Previsão de séries temporais financeiras por meio de redes neurais dinâmicas e processos de transformação de dados: uma abordagem empírico-comparativa / Alexandre Frutuoso da Costa, 2012
94 f. : il.

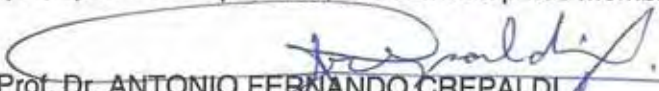
Orientador: Antonio Fernando Crepaldi

Dissertação (Mestrado)-Universidade Estadual Paulista. Faculdade de Engenharia, Bauru, 2012

1. Previsão de séries temporais. 2. Redes neurais.
3. Transformação de dados I. Universidade Estadual Paulista. Faculdade de Engenharia. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE ALEXANDRE FRUCTUOSO DA COSTA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO, DO(A) FACULDADE DE ENGENHARIA DE BAURU.

Aos 21 dias do mês de dezembro do ano de 2012, às 09:30 horas, no(a) ANFITEATRO DA SEÇÃO TÉCNICA DE PÓS-GRADUAÇÃO DA FACULDADE DE ENGENHARIA DE BAURU, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. ANTONIO FERNANDO CREPALDI do(a) Departamento de Engenharia de Produção / Faculdade de Engenharia de Bauru - UNESP, Prof. Dr. ROGÉRIO ANDRADE FLAUZINO do(a) Departamento de Engenharia Elétrica e Computação / Escola de Engenharia de São Carlos - USP, Prof. Dr. JOSE ALFREDO COVOLAN ULSON do(a) Departamento de Engenharia Elétrica / Faculdade de Engenharia de Bauru - UNESP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de ALEXANDRE FRUCTUOSO DA COSTA, intitulado "PREVISÃO DE SÉRIES TEMPORAIS FINANCEIRAS POR MEIO DE REDES NEURAI DINÂMICAS E PROCESSOS DE TRANSFORMAÇÃO DE DADOS: UMA ABORDAGEM EMPÍRICO-COMPARATIVA". Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADO. Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.


Prof. Dr. ANTONIO FERNANDO CREPALDI


Prof. Dr. ROGÉRIO ANDRADE FLAUZINO


Prof. Dr. JOSE ALFREDO COVOLAN ULSON

RESUMO

A previsão de séries temporais financeiras é uma das questões mais pesquisadas no campo das finanças, sobretudo, no que diz respeito ao mercado acionário e à análise de riscos. Para tanto, essas pesquisas envolvem desde modelos estatísticos e econométricos até modelos de inteligência artificial, como as redes neurais dinâmicas. Nesse sentido este estudo tem o propósito de desenvolver e aplicar dois modelos de redes neurais artificiais dinâmicas, a rede neural focada atrasada no tempo - FTDNN (*focused time delay neural network*) e a rede neural auto-regressiva com entradas exógenas - NARX (*nonlinear autoregressive network with exogenous inputs*) para previsão de séries temporais financeiras, tendo como padrão de referência de desempenho mínimo um modelo estatístico tradicional do tipo ARMA-GARCH. Essa abordagem comparativa também considera três modalidades diferentes de transformação de dados na fase de pré-processamento das redes: as diferenças de primeira ordem, os retornos logarítmicos e a transformação Box-Cox, buscando analisar o impacto de cada uma delas no desempenho preditivo das redes neurais. Também propõe uma abordagem neural para o processo de reversão dos dados previstos e uma métrica de erro capaz de verificar o desempenho preditivo das redes neurais e sua capacidade de captar tendências de curto prazo e eficiência negocial. Em sentido amplo, os resultados obtidos indicam que a rede NARX apresenta melhor desempenho preditivo que a rede FTDNN, sobretudo, no que diz respeito à captura de tendências; que as transformações de dados podem melhorar o nível de acurácia das previsões em ambas as redes e que a transformação por retornos logarítmicos gera os melhores desempenhos. Quanto ao processo de reversão dos dados previstos para a escala da série original, o método neural proposto foi bem sucedido apenas para a transformação Box-Cox.

Palavras-chave: Previsão de séries temporais. Redes neurais. Transformação de dados.

ABSTRACT

Financial time-series forecasting is one of the most researched issues in finances, mainly with regard to the stock market and risk analysis. Therefore, these studies involve from statistical and econometric models up to artificial intelligence models, such as dynamic neural networks. In this sense, this study aims to develop and apply two models of dynamic artificial neural networks, FTDNN (focused time delay neural network) and NARX (nonlinear autoregressive network with exogenous inputs) for financial time series forecasting, with a traditional statistical model such as ARMA-GARCH as the benchmark for minimum performance. This comparative approach also considers three different types of data transformation in the pre-processing phase of the networks: first order differences, logarithmic returns and Box-Cox transformation, and tries to analyze the impact of each on the predictive performance of the neural networks. It also proposes a neural approach to the process of reversing the predicted data set, and an error metric that could be able to verify the predictive performance of neural networks and its ability to capture short-term trends and negotiation efficiency. In a broad sense, the results indicate that the NARX network performs better than the FTDNN, especially with regard to capturing trends; that data transformations may improve the forecasting accuracy in both networks, and that the logarithmic returns transformation generates the best prediction performance. Regarding the process of reversing the predicted data for the scale of the original series, the neural method proposed succeeded only for Box-Cox transformation.

Keywords: Time series forecasting. Neural networks. Data transformation.

LISTA DE FIGURAS

Figura 1 - Preços médios diários das ações ordinárias do Banco do Brasil no período de 01/01/2009 a 27/08/2012.....	17
Figura 2 - Série de preços mínimos da ação BBDC4.....	19
Figura 3 - O modelo biológico e o modelo artificial.....	21
Figura 4 - Topologia de uma rede FTDNN com ordem de predição p	24
Figura 5 - Topologia básica da rede NARX com ordem de predição p e q reentradas.....	25
Figura 6 - Fluxo metodológico	43
Figura 7 - Definição da ordem de Predição.....	47
Figura 8 - Séries Propostas.....	49
Figura 9 - Previsões realizadas pela rede FTDNN	50
Figura 10 - Previsões realizadas pela rede NARX	50
Figura 11 - Previsões realizadas pelo modelo ARMA-GARCH.....	51
Figura 12 - Rotina geral de previsão com $N=725$ e $p=8$, para 10 passos à frente.....	55
Figura 13 - Rotina geral da reversão neural de dados	57
Figura 14 - Transformação por diferenças de primeira ordem da Série BBDC3med.....	66
Figura 15 - Transformação por retornos logarítmicos da Série BBDC3med	67
Figura 16 - Transformação Box-Cox da Série BBDC3med	67
Figura 17 - Melhores e piores performances de previsão das redes NARX e FTDNN para séries não transformadas.....	70
Figura 18 - Comparações Gráficas entre as transformações para a série BBDC4-preços mínimos - Rede NARX	74
Figura 19 - Comparações Gráficas entre as transformações para a série BBDC4-preços mínimos - Rede FTDNN	77

LISTA DE TABELAS

Tabela 1 - Alguns trabalhos sobre comparações de modelos para previsão séries temporais.....	10
Tabela 2 - Tempo médio de convergência até o número máximo de iterações (épocas).....	26
Tabela 3 - Códigos e especificações de ações negociadas no mercado a vista além das ON e PN	38
Tabela 4 - Carteira de ações	45
Tabela 5 - Performance das topologias eleitas	49
Tabela 6 - Performance do modelo ARMA-GARCH.....	51
Tabela 7 - Correlação entre as métricas adotadas.....	60
Tabela 8 - Variação das métricas e escores relativos	61
Tabela 9 - Desempenho de previsões para as modalidades de preços - Anova: fator único $\alpha = 5\%$	63
Tabela 10 - Estatística descritiva das séries pré e pós transformação	64
Tabela 11 - Desempenho de previsão 10 passos a frente para séries não transformadas	68
Tabela 12 - Médias das métricas por modelo	69
Tabela 13 - Teste-t: duas amostras em par para médias	70
Tabela 14 - Reversões - Teste t - duas amostras em par para médias $\alpha = 5\%$	73
Tabela 15 - Transformações entre si - NARX - Teste t - duas amostras em par para médias bi-caudal $\alpha=5\%$	74
Tabela 16 - Transformações entre si - FTDNN - Teste t - duas amostras em par para médias bi-caudal ($\alpha=5\%$).....	77
Tabela 17 - Transformações entre redes - Teste t - duas amostras em par para médias $\alpha = 5\%$	80

SUMÁRIO

1. INTRODUÇÃO	8
1.1. Justificativa	9
1.2. Objetivos	12
1.2.1. Objetivo Geral	12
1.2.2 Objetivos Específicos	12
1.3. Organização do Trabalho	13
2. PREVISÃO DE SÉRIES TEMPORAIS	15
2.1. Estacionariedade das Séries	18
3. REDES NEURAS ARTIFICIAIS.....	20
3.1. Rede FTDNN	23
3.2. Rede NARX	24
3.3. Treinamento das redes neurais	26
4 - TRANSFORMAÇÃO E REVERSÃO DE DADOS	28
4.1. Diferenças de primeira ordem	28
4.2. Retornos logarítmicos	29
4.3. Transformação Box-Cox.....	30
4.4 Transformações, reversões e previsões	31
5. MODELO ESTATÍSTICO DE REFERÊNCIA	35
6. ESSENCIALIDADES SOBRE MERCADO DE CAPITALIS	37
6.1 Estratégia negocial.....	41
7. METODOLOGIA	43
7.1. Base de dados	43
7.2. Carteira de ações.....	44
7.3. Desenvolvimento dos modelos de previsão	45
7.3.1. Redes Neurais	45
7.3.2. Modelo ARMA-GARCH	50

7.4. Análise estatística pré e pós transformação de dados	51
7.5. Transformação de dados.....	52
7.6. Treinamento das redes de previsão e reversão	53
7.7. Previsão de dados transformados e não transformados	54
7.8. Reversão dos dados Transformados	55
7.8.1. Reversão algébrica dos dados transformados por diferenças de primeira ordem	55
7.8.2. Reversão algébrica dos dados transformados por retornos logarítmicos	56
7.8.3. Reversão algébrica dos dados transformados por Box-Cox	56
7.8.4. Reversão neural dos dados transformados.....	57
7.9. Avaliação de desempenho	57
7.9.1 Métricas de erro	57
7.9.2 Taxa de eficiência negocial (TEN) - "voando por instrumentos"	59
7.9.3 Escore de desempenho de previsão (EDP)	60
8. ANÁLISE DE RESULTADOS	63
8.1. Análise estatística das séries	63
8.2 Resultados por modelo para séries não transformadas	68
8.3. Comparação entre os processos de reversão em cada rede.....	72
8.4. Avaliação de desempenho de previsão entre as transformações em cada rede.....	73
8.5. Comparação entre as redes para dados transformados.....	80
9. CONCLUSÕES	81
9.1. Trabalhos futuros	82
10. REFERÊNCIAS	83

1. INTRODUÇÃO

Uma das questões essenciais da análise econômico-financeira é a previsão de eventos com a maior exatidão possível para servir como suporte à tomada de decisões por parte dos agentes econômicos. Uma grande parte desses eventos se materializa por meio de preços, independentemente do tipo de produto, serviço ou ativo a que se refira. E há evidente interesse dos agentes envolvidos em projetos de investimento, sejam eles grandes ou pequenos, em prever e estudar o comportamento desses preços, que são analisados na forma de séries temporais, ou seja, uma série de preços indexados temporalmente.

A grande dificuldade para realização dessa tarefa está no fato de que séries temporais econômico-financeiras se apresentam como sistemas dinâmicos com comportamento não-linear, ou seja, são sistemas em que as variáveis do ambiente possuem um complexo padrão de inter-relacionamento que se altera ao longo do tempo (COELHO et al., 2008). Segundo Coelho et al. (2008) e Clements et al. (2004), de forma geral, os modelos econométricos tradicionais, como os modelos auto-regressivos de médias móveis (ARMA) e auto-regressivo de médias móveis supondo heterocedasticidade condicional auto-regressiva (ARMA-GARCH), não trazem resultados satisfatórios em termos de previsão. Daí o fato de que nos últimos vinte anos houve grandes avanços na aplicação de sistemas inteligentes no estudo de séries temporais financeiras, embora, como se pode verificar em Crone et al. (2011), os modelos estatísticos complexos ainda foram os que obtiveram os melhores resultados nessa área.

Por sistemas inteligentes entendem-se as técnicas, computacionalmente intensivas, relacionadas a Sistemas Nebulosos (Lógica Fuzzy), Redes Neurais Artificiais (RNA), Algoritmos Genéticos, Agentes Inteligentes, entre outros, bem como modelos híbridos entre esses modelos mencionados.

A literatura mostra que a discussão quanto à precisão dos modelos para previsão de séries temporais financeiras está longe de um consenso (COELHO et al., 2008); embora já esteja bastante claro que as técnicas não lineares, como as baseadas em

sistemas inteligentes, sejam mais adequadas à previsão de séries temporais financeiras do que as técnicas econométricas ou estatísticas tradicionais, existem estudos que também mostram que os ganhos obtidos, em termos de precisão, não são significativamente melhores para que justifiquem a complexidade adicional na implementação desses modelos (CLEMENTS et al., 2004).

O desenvolvimento da capacidade computacional, no entanto, mostra que a complexidade de manipulação e o custo de processamento desses modelos tende a se tornar pouco relevante diante de seus resultados, porém, dentre os modelos não lineares, intensivos computacionalmente, quais são os mais eficientes? Este trabalho não pretende abordar todos os modelos, mas quer contribuir com as discussões por meio do aprofundamento do estudo da capacidade preditiva das redes neurais dinâmicas, sobretudo quanto ao impacto do pré-processamento no seu desempenho preditivo.

1.1. Justificativa

Uma análise da literatura sobre previsão de séries temporais, mostra que essa é uma área de grande interesse e intensivamente estudada. Conforme Crone et al. (2011), nas duas últimas décadas foram publicados mais de 5000 artigos em revistas acadêmicas e congressos somente sobre redes neurais artificiais, cobrindo uma ampla gama de áreas de estudo. Acrescentando-se a isso todas as publicações envolvendo outras técnicas de inteligência computacional, infere-se o quão importante é essa área de estudos. Pode-se, a título de exemplo, mencionar os trabalhos de Cherif et al. (2011), que discorrem sobre a previsão de séries temporais por meio de redes neurais recorrentes e redes auto-organizáveis de Kohonen; Coelho et al. (2007), que comparam a performance de previsão de várias topologias da rede NARX com modelos estatísticos tradicionais; Figueiredo; Chaves Neto (2008), que comparam a performance preditiva de um modelo ARIMA e de uma rede neural de base radial para séries de vendas como suporte ao planejamento da produção; Gómez-Gil et al. (2011), que propõem uma arquitetura complexa de rede neural para previsões de longo prazo de séries caóticas estacionárias; Katagiri et al. (2011), que propõem um método evolutivo de otimização

para treinamento e topologia de redes neurais recorrentes para previsão de séries temporais; Meade (2002), que estuda previsões de séries temporais de curto prazo com variações de modelos estatísticos; Rapach; Wohar (2006), que realizam previsões de séries temporais de taxas de câmbio por meio de variações de modelos de alisamento exponencial; Villanueva (2006), que estuda a previsão de séries temporais por meio de um comitê de máquinas, ou uma junção de modelos; Stekler (2007), que discorre sobre o processo de previsão de séries temporais em geral na macroeconomia, entre outros.

A literatura relativa a comparações de modelos para previsão de séries temporais, sejam eles estatísticos ou baseados em ferramentas de sistemas inteligentes, como as redes neurais artificiais, é farta, porém, de difícil compilação e comparação, pois, quase que invariavelmente, em cada trabalho os modelos utilizados são diferenciados, as séries são diferentes e as métricas de erros também o são, conforme pode ser observado na Tabela 01, que traz alguns estudos nessa área.

Tabela 1 - Alguns trabalhos sobre comparações de modelos para previsão séries temporais

Trabalho	Comparação	Séries	Métrica	Horizonte de previsão	Melhor desempenho
ADEODATO et al., 2009	FTDNN e "ensembles" de FTDNNs	111 séries das competições NN3 e NN5	SMAPE e MAPE	18 a 56 passos a frente	Ensembles
CHEN, 2005.	Hebbian Linear Associator Network, Perceptron e ADALINE	Crescimento econômico e crescimento agrícola	MAE, MAPE e RMSE	1 passo a frente	Perceptrons
CORRÊA et al. 2007	Diferentes topologias de FTDNNs e uma arquitetura recorrente	Produção brasileira de gás natural e petróleo	MAPE	Não específica	Não foi conclusivo
FIRAT et al., 2010	GRNN, CCNN e FFNN	Consumo de água	AARE e NRMSE	6 passos	CCNN
FLORES, 2009.	FTDNN, RNN, ARIMA e SARIMA	Produção industrial e manchas solares	NMSE e MAPE	12 passos	Não foi conclusivo
GIORGI et al., 2011.	ARMA FTDNN, ELMAN e ANFIS	Força eólica	NMSE	5 horizontes definidos em horas: 1, 3, 6, 12 e 24h	FTDNN para 1 e 3h e ELMAN para 6, 12 e 24h
GOMES, 2005	RNN com diferentes topologias e ARMA	Séries de inflação de diferentes países	MAE	de 1 a 12 meses a frente	Não foi conclusivo.
HERAVI et al., 2004	Modelos lineares e FTDNN	Produção industrial da Europa	RMSE	1,3,6,12 meses a frente	FTDNN

KARA et al., 2011	FFMLP e SVM	Direção do preço de ações	Percentual de acertos	1 passo a frente	FFMLP
KHASHEI, 2010	FTDNN e ARIMA	Taxas de câmbio, manchas solares e lincas canadenses caçados por ano	MAE e MSE	de 1 a 67 passos a frente	FTDNN

FTDNN: Focused time delay neural network; RNN: Recurrent neural network; ANFIS: Adaptive neural fuzzy inference system; FFMLP: Feedforward multi-layer Perceptron; GRNN: Generalized Regression Neural Network; CCNN: Cascade Correlation Neural Network; FFNN: Feed Forward Neural Network; ARMA: Autoregressive Moving Average; ARIMA: Autoregressive integrated moving average; AARE: Average Absolute Relative Error; NRMSE: Normalized Root Mean Square Error; SMAPE: Square mean absolute percentage error; MAPE: Mean absolute percentage error; RMSE: Root mean square error; NMSE: Normalized mean square error; MAE: Mean absolute error.

Fonte: o autor

Especificamente quanto aos modelos neurais comparados nesta dissertação, NARX e FTDNN, podem ser mencionados os trabalhos de Menezes Jr; Barreto (2008), que utiliza as séries *chaotic laser* e *variable bit rate* (VBR) com vários horizontes de previsão e conclui pela superioridade da rede NARX e Coelho et al.(2008), que faz previsões um passo a frente para séries temporais relativas à taxa de câmbio brasileira frente ao dólar americano comparando variações de topologia da rede NARX com modelos estatísticos, concluindo também pela superioridade da rede NARX. Nessa linha de pesquisa se encaixa este trabalho, que pretende discutir a performance desses modelos neurais e sua possível melhoria por intermédio de métodos de transformação de dados.

Prever o futuro com acurácia suficiente para tomar decisões acertadas no mundo financeiro é o estado-da-arte em termos de análise e gestão financeira. Diante da complexidade do mundo financeiro, poder obter a melhor previsão para a variação da taxa de câmbio ou do preço de uma ação nos próximos dias e decidir sobre, por exemplo, o fechamento de um contrato de câmbio envolvendo alguns milhões de dólares, ou a venda das ações, cujo valor deve cair, é, sem dúvida, a grande motivação dos muitos estudos sobre o assunto.

Dessas razões deriva a justificativa para uma abordagem comparativa entre redes neurais dinâmicas com foco, tanto no desempenho preditivo desses modelos, quanto em técnicas de pré-processamento visando a melhoria desse desempenho. Assim, este trabalho se justifica não só pela importância do problema abordado, como também pela sua originalidade nos aspectos referentes a transformações e reversões de dados, possibilitando sua contribuição para o avanço das discussões relacionadas

tanto ao tema geral relativo ao estudo de séries temporais, quanto aos critérios de avaliação de performance preditiva desses modelos e de seu pré-processamento.

1.2. Objetivos

1.2.1. Objetivo Geral

Este trabalho tem como objetivo geral o desenvolvimento e a aplicação comparativa de dois modelos de redes neurais artificiais dinâmicas, um não recorrente, o da rede neural focada atrasada no tempo - FTDNN (*focused time delay neural network*) e o outro, recorrente, a rede neural auto-regressiva com entradas exógenas -NARX (*nonlinear autoregressive network with exogenous inputs*) para previsão de séries temporais financeiras fora-da-amostra (preços de ações), tendo como padrão de referência de desempenho mínimo um modelo estatístico tradicional do tipo ARMA-GARCH.

1.2.2 Objetivos Específicos

A partir do desenvolvimento e aplicação dos modelos neurais citados a séries temporais de preços de ações:

- Avaliar e comparar o impacto no desempenho preditivo, em cada uma das redes de três tipos de transformações de dados — diferenças de primeira ordem, retornos logarítmicos e transformação Box-Cox (ver capítulo 4).
- Propor, testar e comparar um método neural para a reversão dos dados previstos, em cada um dos modelos neurais, bem como para cada modalidade de transformação.
- Desenvolver uma métrica de erro capaz de verificar o desempenho preditivo das redes neurais quanto a sua capacidade de captar tendências de curto prazo e eficiência negocial.

1.3. Organização do Trabalho

Este trabalho está dividido basicamente em quatro partes, primeiramente, a parte introdutória, correspondente ao capítulo 1, onde se delineou a idéia principal da pesquisa relacionada à previsão de séries temporais com a utilização de redes neurais dinâmicas, submetidas a diferentes formas de pré-processamento de dados, bem como a relevância da pesquisa, seu objetivo geral e seus objetivos específicos.

Na segunda parte, abrangendo os capítulos de 2 a 6, realizou-se a revisão da literatura e a fundamentação teórica relativa aos tópicos necessários ao entendimento da metodologia utilizada para a realização desta pesquisa. Nessa parte abordou-se a previsão de séries temporais, com suas principais dificuldades; os aspectos básicos da construção de redes neurais dinâmicas dos tipos FTDNN (*focused time delay neural network*) e NARX (*Nonlinear Autoregressive with Exogenous inputs*); os processos de transformação e reversão de dados utilizados neste estudo (diferenças de primeira ordem, retornos logarítmicos e transformação Box-Cox); a definição do modelo estatístico de referência do tipo ARMA-GARCH e alguns conhecimentos essenciais sobre o mercado de capitais e operações com ações. Nesse último item também se apresentou a estratégia negocial adotada neste trabalho.

Na terceira parte, que corresponde aos capítulos 7 e 8, detalhou-se a metodologia utilizada, suas etapas e particularidades, contemplando aspectos referentes aos dados utilizados, desde a definição da carteira de ações utilizada e a delimitação do período (2009 a 27/08/2012) até as transformações de dados no pré-processamento e a reversão dos dados previstos no pós-processamento; ao treinamento dos modelos com a utilização do algoritmo de Levenberg-Marquardt, às métricas de erro utilizadas para avaliação do desempenho preditivo — MAPE (*Mean absolute percentage error*), U-Theil (o coeficiente de desigualdade de Theil), TEN (taxa de eficiência negocial, métrica desenvolvida neste trabalho baseada na estratégia negocial adotada) e o EDP (escore de desempenho de previsão, que é uma medida que unifica as três métricas anteriores por meio de uma média ponderada) — e as ferramentas estatísticas utilizadas para realizar a comparação entre os modelos: testes de hipótese do tipo pareado para médias e análise de variância ANOVA com fator

único. Nessa parte também se realizou a análise dos resultados obtidos, tanto quanto aos desempenho dos modelos, como quanto ao impacto produzido pelos diferentes tipos de transformação e reversão de dados.

Na quarta e última parte, que corresponde ao capítulo 9, são delineadas as conclusões desta pesquisa conforme os objetivos apresentado no capítulo 1, mostrando os pontos fortes e fracos de cada uma das redes neurais utilizadas, a influência das transformações de dados em seu desempenho preditivo, bem como a performance da abordagem neural para reversão dos dados previstos; nessa parte também são sugeridas algumas linhas de pesquisa para trabalhos futuros.

2. PREVISÃO DE SÉRIES TEMPORAIS

Prever o futuro: um grande sonho da humanidade desde tempos imemoriais. Um tema exótico, normalmente associado ao sobrenatural e ao metafísico, contudo, mesmo na grande saga "Fundação", uma das mais consagradas obras de ficção científica de todos os tempos, do inexpugnável escritor Isaac Asimov, toda a trama gira em torno de uma nova ciência chamada psichistória, desenvolvida por um matemático desconhecido habitante de um dos planetas mais longínquos da galáxia, mas que quando apresentou seu artigo em um congresso científico no planeta central do império galáctico, tornou-se o homem mais procurado de todo o império e acabou por determinar todo o futuro da humanidade, tudo em razão de um simples fato: suas técnicas matemáticas e estatísticas podiam, de fato, prever o futuro.

No mundo real, o estudo de técnicas para previsão do futuro recaem sobre a construção de modelos matemáticos que procuram capturar as características do comportamento histórico de sistemas dinâmicos e, a partir daí, estimar seu comportamento futuro, ou seja, seu valor futuro. Trata-se aqui da análise e previsão de séries temporais, ou um conjunto de observações ordenadas no tempo (MORETTIN; TOLOI, 2006), ou conceitualmente, conforme Cordeiro (2002, p.17): " Uma série temporal é um conjunto de observações aleatórias $X_1, X_2, X_3, \dots, X_t$, observadas em instantes discretos de tempo igualmente espaçados 1, 2, 3, ..., t."

Pode-se citar como exemplos de séries temporais: preços de fechamento diários de uma ação na bolsa de valores de São Paulo durante o ano de 2010; valores *intra-day* da taxa de câmbio R\$/US\$ durante uma determinada semana; índices pluviométricos semanais registrados na região de Bauru; vazões registradas diariamente em uma usina hidroelétrica etc.

Os objetivos para o estudo de séries temporais podem ser divididos em (MORETTIN; TOLOI, 2006; CORDEIRO, 2002; MAIA, 2008):

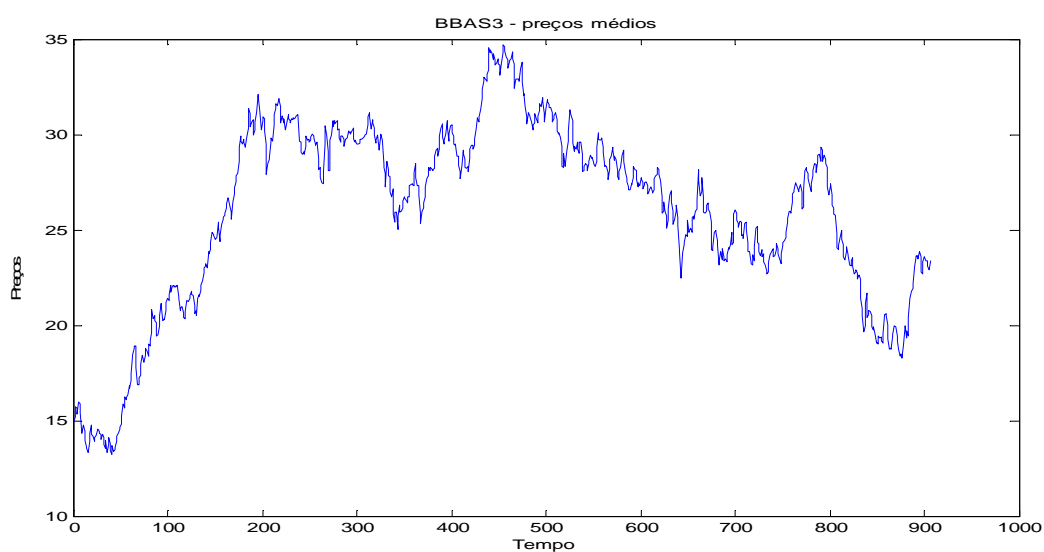
- Descrição: analisar o comportamento da série, construir seu gráfico, verificar suas tendências, suas características de sazonalidade e obter suas estatísticas básicas;
- Explicação: investigar o mecanismo gerador da série ou compará-la a uma outra série cuja variação pode explicar a variação da primeira;
- Controle: controlar processos e qualidade de dados;
- Predição: fazer previsões de valores futuros de uma série, seja no curto, seja no longo prazo.

Para fins deste trabalho o objetivo da abordagem das séries temporais será o de predição, tema que tem sido intensamente estudado, nas mais diversas áreas, tais como: carga elétrica (LEONE FILHO, 2006; VELASQUES, 2006); chuvas (NOURANI et al., 2011); clima espacial (REIKARD, 2011); consumo de água (FIRAT et al., 2010); demanda de energia elétrica (STEINMETZ, 2009; ANDRADE, 2010); demanda de equipamentos de infra-estrutura de telecomunicações (BRANCO; SAMPAIO, 2008); dilatação de solo argiloso (YILMAZ; KAYNAR, 2011); fluxo de tráfego urbano (CHEN; ZHANG, 2005) e, em especial, por ser também o foco deste trabalho, as séries financeiras, que, em geral, referem-se a variações de preços e índices do mercado financeiro (KIM; SHIN, 2007; BRAGA, 2006; ARAÚJO, 2010; ARAÚJO, 2011; CASTRO, 2001; CAVALHEIRO et al., 2010; CHAN, 2009; CHEN; LEUNG, 2004; CHEN, 2005; CICHINI et al., 2007; CICHINI, 2009; CLEMENTS et al., 2004; EBRAHIMPOUR, 2010; FENG; CHOU, 2011; KARA et al. 2011; LIMA, 2004; OLIVEIRA, 2007; SAAD; PROKHOROV, 1998; SANTOS FILHO, 2008; SHEN, 2011; SILVA, 2010; SOARES, 2008; SOARES et al., 2008; VICENTE et al. , 2010). De fato, segundo Zou; Yang (2004, apud LIMA, 2004), a área de finanças, nos últimos vinte e cinco anos, vem se tornando cada vez mais uma ciência essencialmente quantitativa ou matemática, com o uso extensivo de ferramentas estatísticas e computacionais, visando, no processamento de informações, resultados cada vez mais precisos, sobretudo para os modelos de previsão (HERAVI et al., 2004).

As principais dificuldades para previsão de séries temporais financeiras residem em suas próprias características: tendências; sazonalidade; pontos influentes (atípicos);

heteroscedasticidade condicional e não-linearidade (MORETTIN; TOLOI, 2006), ou seja, são características que revelam, volatilidade e não estacionariedade, como pode ser observado, a título de exemplo, na Figura 1. A presença de comportamento não-linear em séries financeiras tem gerado críticas quanto ao uso de modelos lineares na explicação e previsão de sua dinâmica (COELHO et al., 2008).

Figura 1 - Preços médios diários das ações ordinárias do Banco do Brasil no período de 01/01/2009 a 27/08/2012



Fonte: o autor

As teorias estatísticas e econométricas tradicionais, baseada em modelos de características lineares, não têm se mostrado tão eficazes em replicar a dinâmica inerente às séries econômicas e financeiras, aspecto que tem levado os pesquisadores de diversas formações acadêmicas a buscarem as técnicas de inteligência computacional (COELHO et al., 2008). Essas técnicas, também conhecidas como AI (*artificial intelligence*), oferecem uma abordagem mais efetiva para a manipulação de grandes quantidades de dados dinâmicos, não-lineares e ruidosos (NOURANI et al., 2011), que as tornam adequadas para a modelagem de séries temporais econômico financeiras, daí a grande quantidade de estudos na área, tais como Araújo (2011); Cheong; Kim; Yoon (2011); Oliveira (2010); Cichini (2009) e Panda; Narasimhan (2007).

Ainda que se trate de modelos não lineares, estatísticos ou de inteligência computacional, a presença de estacionariedade em uma série, dadas suas características estatísticas, pode ser um fator importante para melhorar a acurácia da previsão.

2.1. Estacionariedade das Séries

As séries temporais financeiras relativas aos preços diários de ações negociadas na Bovespa, invariavelmente, apresentam características como tendências, sazonalidade cíclicas e acíclicas, que as classificam como séries não estacionárias. Em termos mais formais existem duas formas de estacionariedade, a fraca ou ampla ou de segunda ordem, e a estrita ou forte (MORETTIN; TOLOI, 2006; HAMILTON, 1994):

- **Estrita:** uma série temporal X_t , $t \in T$ é considerada estritamente estacionária se todas as distribuições finito-dimensionais não variam quando t (tempo) varia, isto é,

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = F(x_1, \dots, x_n; t_1 + \Delta t, \dots, t_n + \Delta t) \quad (01)$$

para quaisquer $t_1, \dots, t_n, \Delta t \in T$.

- **Fraca:** uma série temporal X_t , $t \in T$ é considerada fraca ou amplamente estacionária se e somente se

$$E[X_t] = \mu_t = \mu, \forall t \in T \quad (02)$$

$$\text{Var}[X_t] = \sigma^2 < \infty, \forall t \in T \quad (03)$$

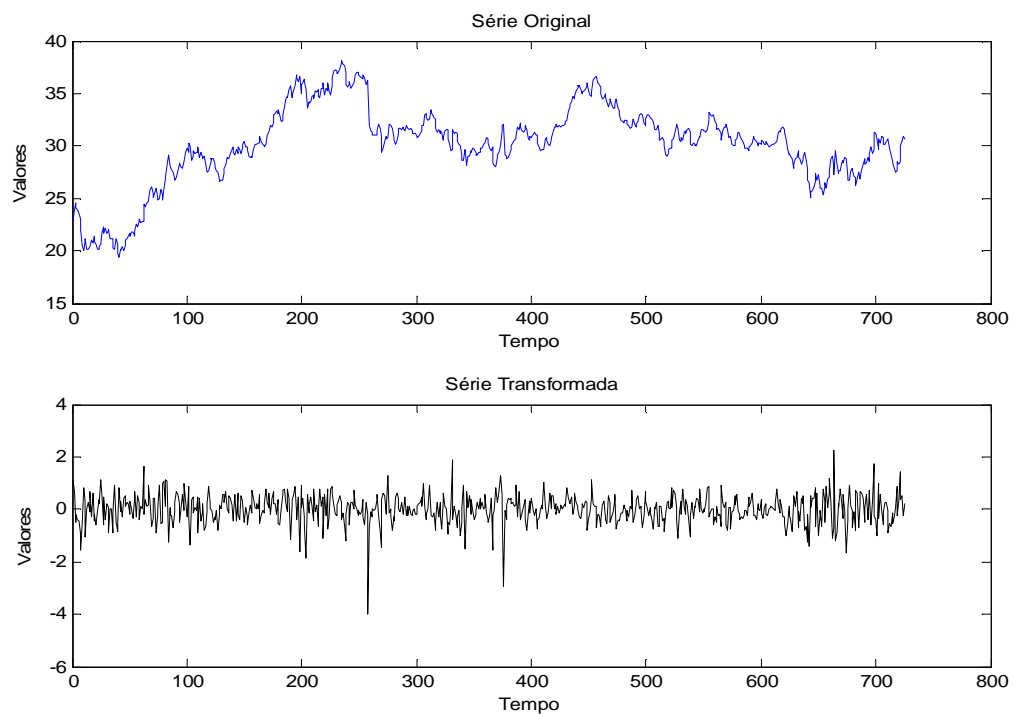
$$\text{Cov}_x(t_1; t_2) = \text{Cov}(|t_1 - t_2|) \quad (04)$$

o que significa dizer que a média é constante, a variância finita e a autocovariância só depende da defasagem entre os tempos t_1 e t_2 , ou seja, ela não depende de t , mas apenas do espaço de tempo entre as duas observações (HAMILTON, 1994).

Para este trabalho, toda menção à estacionariedade de uma série refere-se ao conceito de estacionariedade fraca.

Pode-se observar na Figura 02 abaixo, a série temporal original de preços mínimos da ação preferencial do BRADESCO - BBDC4 com característica não estacionária e a série transformada por diferenças de primeira ordem (ver item 4.1 - equação 13) , mostrando características de estacionariedade.

Figura 2 - Série de preços mínimos da ação BBDC4



Fonte: o autor

3. REDES NEURAIS ARTIFICIAIS

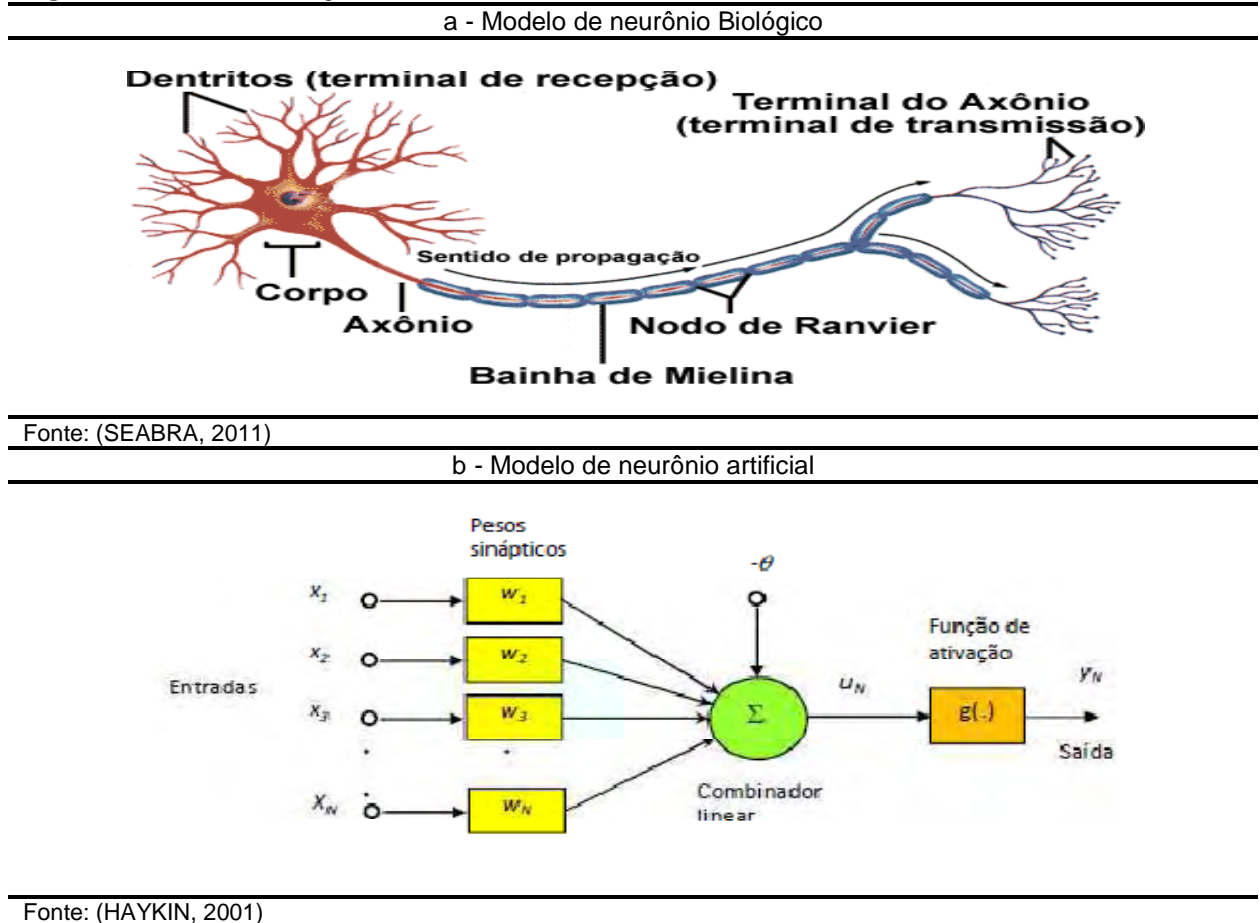
A partir das primeiras investigações que tentavam definir um modelo computacional para a forma de processamento de informações do cérebro humano nasceu o estudo de redes neurais artificiais. A primeira referência para a teoria de redes neurais artificiais data de 1943, quando Warren McCulloch e Walter Pitts publicaram o artigo *A Logical Calculus of the Ideas Immanent in Nervous Activity* (GOMES, 2005). Desde então o estudo das RNA se desenvolveu intensamente, tornando-se uma das técnicas de Inteligência Computacional mais estudadas e utilizadas, conforme Oliveira (2007), até 2007 existiam cerca de 10130 patentes que utilizam RNA registradas no escritório de patentes e marca registrada nos Estados Unidos (*USPTO - United States Patent and Trademark Office*).

Conceitualmente, redes neurais artificiais são modelos computacionais inspirados no cérebro biológico e que possuem a capacidade de aquisição, manutenção e generalização do conhecimento. Estruturalmente, as RNA podem ser definidas como conjuntos de unidades processadoras (neurônios) que são interligados por um grande número de interconexões (sinapses) (SILVA et al., 2010). Conforme Silva et al. (2010) as principais características na aplicação de RNA são:

- Aprender por meio de exemplos, ou seja, aprender a partir da interação com o ambiente. Utilizando um algoritmo de treinamento a rede consegue absorver o relacionamento entre as variáveis do modelo.
- Capacidade de generalização. Após o treinamento a rede consegue generalizar o conhecimento absorvido.
- Capacidade de se adaptar, isto é, após o treinamento a rede ajusta seus pesos sinápticos internos, para produzir a melhor resposta;
- Agrupar ou organizar dados e, dessa forma, criar grupamentos com características semelhantes e classificar padrões;

- Tolerância à falhas. Devido ao grande número de conexões internas, uma rede pode continuar funcionando mesmo se parte de sua estrutura interna for corrompida.

Figura 3 - O modelo biológico e o modelo artificial



Como se pode verificar pela comparação entre os modelos a e b da Figura 03, o corpo celular, por meio de seus dendritos, recebe entradas (sinais) de outros neurônios através das conexões sinápticas, o que, no neurônio artificial é realizado pelos pesos sinápticos (w). O núcleo celular soma esses sinais, o que no neurônio artificial é realizado pelo combinador linear, que, matematicamente e computacionalmente é a soma matricial ponderada dos valores de entrada. Quando a soma atinge um certo limiar ($-\theta$), o núcleo dispara pulsos que são transmitidos pelo axônio. No neurônio artificial isso é realizado pela função de ativação, que, a partir dos valores obtidos no combinador

linear, os transforma nas saídas do neurônio (SILVA et al., 2010). Cabe ressaltar que para as redes de múltiplas camadas, como as que serão utilizadas neste trabalho, frequentemente se utilizam as funções logística, Equação 05, e a tangente hiperbólica, Equação 06, para os neurônios das camadas intermediárias e uma função linear de transferência, Equação 07, para a camada de saída (KASHEI; BIJARI, 2010).

$$g_1(x) = \frac{1}{1 + e^{-x}} \quad (05)$$

$$g_2(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (06)$$

$$g_3(x) = x \quad (07)$$

As redes neurais artificiais podem se configurar em diversas arquiteturas. A forma mais simples é a alimentada adiante – *feedforward*, que se constitui de múltiplas camadas, em que a entrada de cada camada é a saída da camada predecessora e a interconexão dos neurônios é acíclica (CERVILHA et al., 2008).

Muitas redes, com diferentes topologias e formas de treinamento, foram desenvolvidas, ao longo do tempo, para se adaptar ao problema a ser resolvido nas mais diversas áreas das ciências. As aplicações potenciais das RNA podem ser classificadas da seguinte forma, segundo Silva et al. (2010):

- Aproximador universal de funções;
- Controle de processos;
- Reconhecimento/classificação de padrões;
- Agrupamento de dados (clusterização);
- Sistemas de previsão;
- Otimização de sistemas;
- Memórias associativas;

Para os fins deste trabalho, a aplicação abordada será a de sistemas de previsão. Com essa finalidade específica, as redes PMC - Perceptron multi-camadas com treinamento *backpropagation* são as mais estudadas e aplicadas. Pode-se

observar pela literatura uma intensa utilização dessa concepção de rede e de treinamento para previsão de séries temporais, muitas com modificações sutis na topologia ou no algoritmo de treinamento, mas mantendo as características essenciais do modelo, como em Flores; Werner (2007); Guimarães (2006); Melo (2003); Mileski Junior; Vieira (2008); Moshiri (1997); Sandmann (2006); Adeodato et al. (2009); Castillo et al. (2006); Corrêa et al. (2007); Flores (2009); Giorgi et al (2011); Gonzales (2009); Jiang (2002); Khosravi et al. (2009); Luna et al. (2011); Machado (2003); Marin (2009); Martínez-Rego et al. (2011); Maus; Sprott (2011); Menezes Junior; A Barreto (2008); Oliveira et al. (2010); Reuter; Möller (2010); Schopf (2007); Spaeth (2008); Thomaidis; Dounias (2010); Wang (2009).

3.1. Rede FTDNN

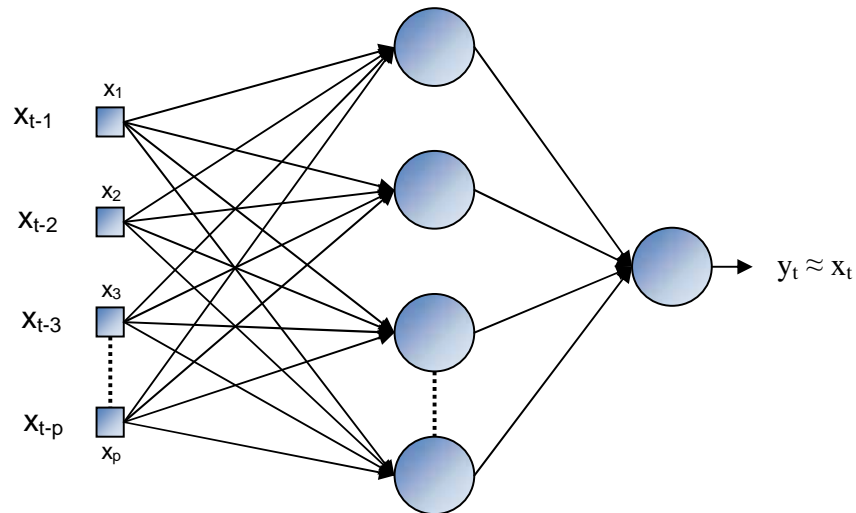
As redes FTDNN (*focused time delay neural network*) nada mais são do que Perceptrons Multi-camadas, em que se introduz o aspecto dinâmico e temporal por meio da camada de entrada, ou seja, proporciona-se à rede uma memória de curto prazo (HAYKIN, 2001), aplicando ao vetor de entrada da rede um atraso no tempo. Fica claro então que, para cada valor x_t desejado, a rede treinará com p valores passados, sendo p , portanto, a ordem de predição ou o alcance da memória de curto prazo da rede, ou seja,

$$y_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p}) \quad (08)$$

onde y_t é o valor previsto pela rede para o momento t .

Essas entradas atrasadas no tempo permitem que a rede saiba de parte da história da variável no tempo t , possibilitando a solução de problemas de decisão mais complexos, especialmente aqueles dependentes do tempo (KIM; SHIN, 2007). Esse mecanismo pode ser ilustrado pela Figura 04.

Figura 4 - Topologia de uma rede FTDNN com ordem de predição p



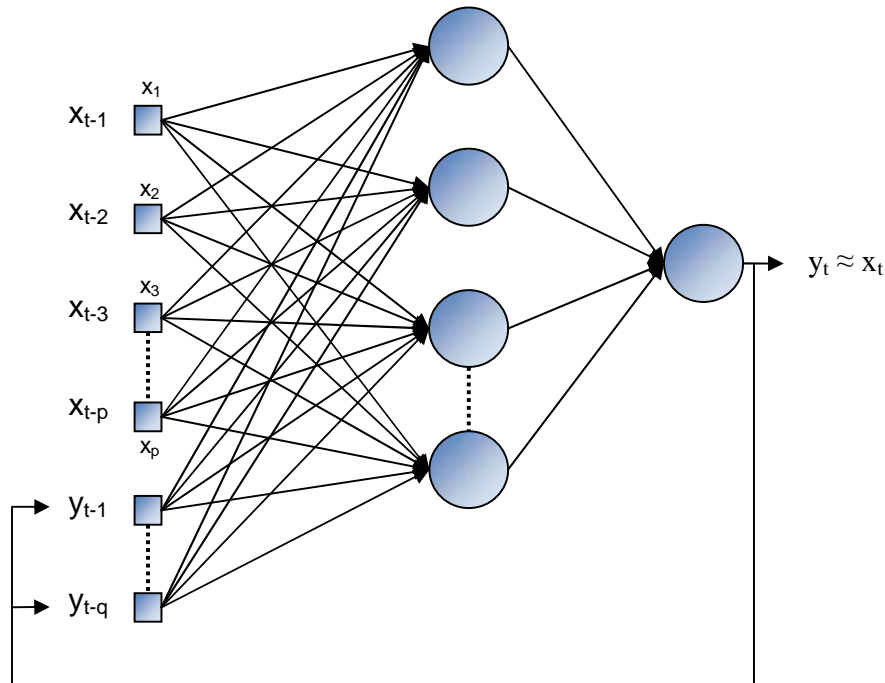
Fonte: adaptado de Silva et al.(2010)

No contexto geral de previsão de séries temporais as redes FTDNN são largamente utilizadas, como se pode verificar em Adeodato et al. (2009), Braga (2006), Branco; Sampaio (2008), Castillo et al. (2006), Chan (2009), Chen (2005), Jiang et al. (2002), Kara et al. (2011) entre outros.

3.2. Rede NARX

Os modelos não lineares auto-regressivos com entradas exógenas NARX (*Nonlinear Autoregressive with Exogenous inputs*), também têm sua arquitetura baseada na rede Perceptron Multi-camadas, bem como, também recebem entradas atrasadas no tempo, como a arquitetura FTDNN. Porém recebem reentradas do próprio sistema, aumentando a capacidade de mapeamento não-linear da rede (HAYKIN, 2001). Por meio da Figura 05 pode-se observar esse mecanismo.

Figura 5 - Topologia básica da rede NARX com ordem de predição p e q reentradas



Fonte: adaptado de Haykin, 2001 e Silva et al., 2010

Os valores previstos pela rede NARX, portanto, não apenas se baseiam em p valores passados, mas também em q valores previstos anteriormente, ou seja:

$$y_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p}, y_{t-1}, y_{t-2}, \dots, y_{t-q}) \quad (09)$$

onde p é a ordem de predição, indicando até quantos valores passados da série a previsão vai se basear, e q a quantidade de saídas passadas que também contribuirão para a previsão de y_t . Essas saídas passadas são denominadas sinais ou unidades de contexto e q, a ordem de contexto (SILVA et al., 2010).

Cabe observar também que, usualmente, as redes NARX aparecem na literatura, ora com a arquitetura paralela, ora com a arquitetura série-paralela, ou ambas. A diferença entre as duas reside na origem das unidades de contexto, isto é, na arquitetura paralela, a origem são as próprias saídas estimadas do sistema, enquanto na série-paralela, são os valores reais correspondentes às saídas estimadas (ARAUJO,

2010; HAYKIN, 2001). Neste trabalho adotou-se a arquitetura paralela, pois, dessa forma, as unidades de contexto seriam a única diferença em relação ao modelo FTDNN. A arquitetura NARX também é bastante utilizada no âmbito da previsão de séries temporais, como pode ser visto em Araújo (2010), Coelho et al. (2008), Flores (2009), Gomes (2005) entre outros.

3.3. Treinamento das redes neurais

O algoritmo de treinamento utilizado pelas redes NARX e FTDNN neste trabalho foi o Levenberg-Marquardt, considerado na literatura como um dos mais utilizados para minimizar o tempo de convergência e o esforço computacional (HAGAN & MENHAJ apud SILVA et al., 2010, p.117). Há, porém, trabalhos, em que, embora reconheçam e utilizem esse algoritmo, argumentam que ele exige um grande número de operações e, portanto, alto esforço computacional (COELHO et al., 2008). Neste estudo ele apresentou, em termos de tempo para convergência, os resultados médios conforme Tabela 02.

Tabela 2 - Tempo médio de convergência até o número máximo de iterações (épocas)

Rede	Quantidade de dados	Quantidade de iterações	Tempo médio
FTDNN	725	200	08 segundos
NARX	725	50	39 segundos

Fonte: o autor

Esse algoritmo é supervisionado e realizado por retro-propagação do erro ou, como é mais conhecido, *backpropagation*, o que significa dizer, que o dado de saída y_n é confrontado com seu valor desejado x_n e apurado um erro ε_n , que é retro-propagado pela rede buscando a minimização do erro quadrático por intermédio do reajuste dos pesos w_j , e fará tantas iterações quantas forem necessárias para que o erro médio quadrático (MSE) atinja um valor menor que o estipulado, ou que realize um número máximo de iterações, também chamadas épocas; ou seja, quando a rede atinge um dos dois critérios de parada, o que ocorrer primeiro, a rede pára e é considerada treinada (SILVA et al., 2010; HAYKIN, 2001).

O algoritmo Levenberg-Marquardt especificamente consiste em um método de gradiente descendente de segunda ordem, porém, com a vantagem computacional de não processar a matriz Hessiana dos erros em relação aos pesos, mas sim uma aproximação dada por:

$$H(n) \cong J^T(n).J(n) \quad (10)$$

onde J é a matriz Jacobiana de ordem n composta das derivadas parciais dos erros da rede em relação aos pesos, o que leva à equação geral do método para ajuste de pesos sinápticos:

$$w(n+1) = w(n) - (J^T(n).J(n) + \mu I)^{-1} J^T(n) \varepsilon(n) \quad (11)$$

onde μI é uma matriz diagonal adicionada para assegurar, tanto que $J^T(n).J(n)$ seja singular, como também a eficiência computacional, uma vez que μ , na implementação computacional, é variável, e pode ser entendido como uma taxa de aprendizagem inteligente, pois varia de forma inversa ao erro. Portanto, o i -ésimo valor de saída da rede será definido, de forma geral, por:

$$y_i = f([x], [d], w(n+h), g(u_m), \varepsilon_p) \quad (12)$$

onde $[x]$ é vetor de entradas, $[d]$ é vetor de valores desejados, $w(n+h)$ é o conjunto de pesos após h épocas, $g(u_m)$ é a função de ativação e ε_p é erro estipulado como critério de parada (SILVA et al, 2010; HAYKIN, 2001).

4 - TRANSFORMAÇÃO E REVERSÃO DE DADOS

As transformações dos dados das séries temporais se justificam pela busca de características próprias de estacionariedade e gaussianidade aproximada, o que, para modelos estatísticos tradicionais, é essencial. Quando se trata, no entanto, da utilização de redes neurais, parece não haver consenso na literatura. Zhang; Qi (2005) mostram que entre 1991 e 1997, vários trabalhos mostraram que as redes neurais eram capazes de mapear tendências e sazonalidades diretamente. Ainda assim Nelson et al. (1999) mostrou que séries transformadas (desazonalizadas) ajudavam as redes a melhorar seu desempenho de previsão. Heravi et al. (2004) usa retornos logarítmicos, Coelho et al.(2008) usa retornos e diferenças. Em finanças a utilização da transformação por retornos logarítmicos é bastante consensual (DENÍELSSON, 2011).

Considerando os objetivos deste trabalho, adotou-se três métodos de transformação de dados:

- diferenças de primeira ordem;
- retornos logarítmicos;
- Box-cox.

São, provavelmente, as transformações mais simples e conhecidas na literatura relativa a análise de séries temporais, presentes na maioria dos livros-texto que abordam esse tema (MORETTIN;TOLOI, 2006; HAMILTON,1994).

4.1. Diferenças de primeira ordem

Dada uma série temporal X_t a diferença de primeira ordem é definida por:

$$\Delta x_t = x_t - x_{t-1} \quad (13)$$

As diferenças de segunda ordem e superiores são dadas pela diferença da diferença e assim por diante, sendo a n-ésima diferença definida por:

$$\Delta^n x_t = \Delta[\Delta^{n-1} x_t] \quad (14)$$

Conforme Morettin; Toloi (2006), em condições normais a primeira ou segunda diferenças já bastam para tornar a série estacionária, o que pode ser constatado na Tabela 10 que mostra que, para todas as séries transformadas por esse método, o teste de estacionariedade aplicado (Dickey Fuller aumentado) indicou a presença dessa característica.

4.2. Retornos logarítmicos

Na literatura relativa a finanças os retornos logarítmicos são amplamente utilizados, sobretudo em função da necessidade de avaliações de risco (MORETTIN; TOLOI, 2006), assim como pelo fato de que os retornos têm propriedades estatísticas que facilitam bastante sua análise (DANIELSSON, 2011).

Dada uma série temporal de preços X_t , definem-se dois tipos de retornos:

- Simples:

$$X_{r(t)} = \frac{x_t - x_{t-1}}{x_{t-1}} \quad (15)$$

- Composto continuamente ou log-retorno:

$$X_{r(t)} = \ln \frac{x_t}{x_{t-1}} \quad (16)$$

Para este trabalho adotou-se a definição de retorno composto ou logarítmico, que será referido apenas como retorno, tanto pela sua larga utilização, como pelo fato de que o retorno simples converge para o composto para flutuações pequenas.

A propriedade de estacionariedade promovida por essa transformação pode ser verificada na Tabela 10, que apresenta essa indicação a partir do teste Dickey-Fuller aumentado.

4.3. Transformação Box-Cox

A transformação de Box-Cox (1964) é frequentemente utilizada com o objetivo de estabilizar a variância da série, bem como para tornar o efeito sazonal aditivo (MORETTIN, 2006), além de tornar a distribuição da série aproximadamente gaussiana, que são premissas assumidas em muitos estudos e análises estatísticas (SAKIA, 1992). Essa técnica faz parte de uma família de transformações, chamada transformações de potências introduzidas por Tukey (1957, apud SAKIA, 1992, p.170).

Na transformação Box-Cox os valores transformados são uma função monótona dos valores originais da série em um intervalo admissível e pode ser definida por:

$$x_t^{(\lambda_1, \lambda_2)} = \begin{cases} \frac{(x_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1}; & \lambda_1 \neq 0 \\ \ln(x_t + \lambda_2); & \lambda_1 = 0 \end{cases} \quad (17)$$

onde λ_1 e λ_2 precisam ser estimados. O valor de λ_2 deve ser estimado para tornar a série positiva, caso haja a presença de valores negativos na série. Quanto a λ_1 há farta literatura com várias propostas de métodos para sua estimação, Box; Cox (1964) propuseram métodos de máxima verossimilhança e Bayesianos, a partir de então, trabalhos como os de Draper & Cox (1969, apud SAKIA, 1992, p.170), Hinkley (1975, 1985 apud SAKIA, 1992, p.170, 171), Andrew et al. (1971, 1973, apud SAKIA, 1992, p.170), Dunn & Tubbs (1980, apud SAKIA, 1992, p.170), Beauchamp & Robson (1986, apud SAKIA, 1992, p.170), Chang (1977a, apud SAKIA, 1992, p.171), Huang et al. (1978, apud SAKIA, 1992, p.171), Cressie (1978, apud SAKIA, 1992, p.170), Hernandez & Johnson (1980, apud SAKIA, 1992, p.171), Pericchi (1981, apud SAKIA, 1992, p.170), Sweeting (1984, apud SAKIA, 1992, p.170), Carroll (1980; 1982a, apud SAKIA, 1992, p.170), Bickel & Doksum (1981, apud SAKIA, 1992, p.170), Carroll & Ruppert (1985, 187 apud SAKIA, 1992, p.170), Taylor (1983, 1985a, b, 1987, apud SAKIA, 1992,

p.170), Han (1987, apud SAKIA, 1992, p.171), Solomon (1985, apud SAKIA, 1992, p.171) e o próprio Sakia (1992), propuseram, desde extensões do procedimento de Box e Cox, até métodos não paramétricos, gráficos e computacionais para a estimação do parâmetro de transformação λ_1 .

Para este trabalho, o valor de λ_2 foi fixado em 0, uma vez que todas as séries utilizadas são de preços de ações e, portanto, contêm apenas dados positivos. Como a estimação do parâmetro λ_1 não era objeto de estudo, ele foi, por sua simplicidade e acurácia, estimado para cada série temporal, por um algoritmo de otimização computacional que maximiza sua função de log-verossimilhança por meio do método de busca simplex de Lagarias et al. (1998), disponível no software MATLAB® R2010a.

4.4 Transformações, reversões e previsões

De modo geral, a literatura que mostra o estudo do impacto das transformações de dados no desempenho de previsão de séries temporais concentra-se em modelos estatísticos tradicionais e está restrita à década de 1970. Trabalhos como os de Makridakis e Hibon (1979), Granger e Newbold (1976) e Nelson (1976, apud MORETTIN; TOLOI, 2006, p.8) verificaram que as transformações não afetam de maneira significativa a acurácia da previsão, mas que os dados transformados incluem um viés na previsão que precisa ser ajustado, o que também é estudado por Sakia (1990).

O estudo do viés introduzido pelas transformações — quando de sua reversão — e formas de ajuste é relativamente escasso na literatura. Não foi possível encontrar trabalhos especificamente para as reversões das transformações por diferenças e por retornos logarítmicos. Para a transformação Box-Cox, particularmente quanto à reversão da transformação dos dados previstos, e o viés introduzido nas estimações, a literatura parece ter atingido seu ápice durante a década de 80 e início da década de 90 e, com a exceção de Duan (1983), os demais estudos relativos ao assunto abordam modelos paramétricos.

Segundo Sakia (1992), essa era uma das discussões mais controversas no contexto das transformações de dados e referiam-se, sobretudo, a se as inferências

deveriam ser feitas a partir da escala transformada ou a partir da escala original com ajustes para o viés gerado. Pesquisadores como Bickel e Doksum (1981), Hinkley e Runger (1984), Miller (1984), Smallwood e Blayblock (1986, apud SAKIA, 1992), Taylor (1986), Sakia (1990) propuseram várias formas de correção do viés gerado quando da reversão.

Para clarificar a questão — relativamente à transformação Box-Cox — deve-se considerar o problema da reversão para os valores que foram previstos a partir de uma série temporal transformada: adotando uma série X_t e y_t como valores previstos para essa série, a dificuldade se impõe quando é preciso encontrar a função f tal que $y_t = f(y_t^{(\lambda)})$ em que f deve suprir a dependência de y_t com relação a x_{t-k} , $k = 1, 2, \dots, p$ em que p é a ordem de predição. Retomando a equação (17), agora para os valores previstos y_t da série X_t e fazendo $\lambda_2 = 0$ e $\lambda_1 = \lambda$, algebricamente teríamos:

$$y_t = \begin{cases} [y_t^{(\lambda)} \cdot \lambda + 1]^{1/\lambda}; & \lambda \neq 0 \\ e^{y_t^{(\lambda)}}; & \lambda = 0 \end{cases} \quad (18)$$

no entanto, como $y_t^{(\lambda)} = E[x_t^{(\lambda)} | x_{t-1}^{(\lambda)}, x_{t-2}^{(\lambda)}, \dots, x_{t-p}^{(\lambda)}]$ é razoável considerar que essa reversão gera um viés na medida em que:

$$[y_t^{(\lambda)} \cdot \lambda + 1]^{1/\lambda} \neq [E(x_t^{(\lambda)} | x_{t-1}^{(\lambda)}, x_{t-2}^{(\lambda)}, \dots, x_{t-p}^{(\lambda)}) \cdot \lambda + 1]^{1/\lambda} \quad \text{para } \lambda \neq 0 \text{ e} \quad (19)$$

$$e^{y_t^{(\lambda)}} \neq e^{E(x_t^{(\lambda)} | x_{t-1}^{(\lambda)}, x_{t-2}^{(\lambda)}, \dots, x_{t-p}^{(\lambda)})} \quad \text{para } \lambda = 0 \quad (20)$$

sobretudo, pelo fato evidente de que x_t tem distribuição diferente de $x_t^{(\lambda)}$.

Segundo Nelson e Granger (1979), uma reversão sem viés para $\lambda = 0$, desde que $x_t^{(\lambda)}$ seja normalmente distribuído, é dada por:

$$y_t = e^{(y_t^{(\lambda)} + \sigma_h^2/2)} \quad (21)$$

onde σ_h^2 é a variância dos erros $x_t^{(\lambda)} - y_t^{(\lambda)}$ dos h passos a frente previstos. Já para $\lambda \neq 0$, a reversão sem viés não tem uma forma fechada e tem que ser obtida pela integral:

$$y_t = \frac{1}{\sigma_h \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2 \left(\frac{w - y_t^{(\lambda)}}{\sigma_h} \right)^2} (\lambda w - 1)^{1/\lambda} dw \quad (22)$$

onde w é uma variável aleatória com distribuição normal tal que $y_t^{(\lambda)} = E[w]$.

A maior dificuldade está no fato de que essas equações somente são válidas para $x_t^{(\lambda)}$ gaussianamente distribuído, o que pode ser um evento bastante raro no universo financeiro, como se pode observar na Tabela 10 no capítulo 8, o teste Jarque-Bera (JARQUE; BERA, 1987) para a série transformada rejeitou a hipótese nula (de normalidade) em todas elas, ou seja, nenhuma das séries era normalmente distribuída e nenhuma das transformações, sobretudo a Box-Cox, que tem esse como um de seus objetivos, obteve sucesso em alcançá-lo.

No que tange especificamente ao estudo das transformações de dados de séries temporais e seu impacto em modelos neurais, a literatura também é relativamente escassa, pode-se mencionar o trabalho de Nelson et al.(1999), que testa uma rede neural, de arquitetura e topologia apenas indicadas como as mesmas utilizadas em Hill et al. (1996, apud NELSON et al. 1999, p.362), com séries que apresentam sazonalidade e com os mesmos dados desazonalizados, mostrando que essa transformação gerou melhores desempenhos de previsão, o processo utilizado na desazonalização também não é descrito.

Mais recentemente Zhang; Qi (2005) realizaram um estudo mais extensivo testando uma rede neural de arquitetura FTDNN com uma camada escondida, um neurônio na saída e várias topologias diferenciadas, para realizar previsões de séries temporais que apresentavam tendências e sazonalidades, primeiramente, com os dados originais, em seguida, com os dados desazonalizados, depois com os dados transformados para eliminar as tendências e, por fim, com os dados desazonalizados e

sem tendências, mostrando empiricamente que a dupla transformação gerou melhores desempenhos de predição.

Já quanto às transformações específicas utilizadas neste trabalho e seu impacto no desempenho de redes neurais, embora elas sejam largamente utilizadas, não foi possível identificar estudos nessa área. No entanto, envolvendo diretamente a transformação Box-Cox e redes neurais pode-se mencionar o trabalho de Hong (2006) que propõe a utilização de uma rede (RBF) ou rede de função de base radial para realizar essa transformação sem a necessidade de treinamento supervisionado e da estimação do parâmetro de transformação λ .

Não foi possível identificar trabalhos a respeito do processo de reversão dos dados previstos para a escala dos dados originais com a utilização de abordagens neurais.

5. MODELO ESTATÍSTICO DE REFERÊNCIA

Com o objetivo de tornar o padrão de desempenho das redes, simultaneamente, desafiador e consistente com os trabalhos de Coelho et al. (2008), Flores (2009), Heravi et al. (2004), Khashei; Bijari (2010), Oliveira et al. (2010), Oliveira (2007), Panda; Narasimhan (2007), Reikard (2011), que compararam o desempenho preditivo de redes neurais e modelos estatísticos tradicionais, concluindo, em sentido amplo, pela superioridade dos modelos neurais, adotou-se um modelo estatístico de referência para estabelecer o padrão mínimo de desempenho das redes utilizadas.

Porém, como as redes neurais são notoriamente mapeadoras eficientes da não linearidade inerente das séries temporais financeiras, mesmo considerando as séries de retornos financeiros, que, embora estacionários apresentam heterocedasticidade condicional, o modelo estatístico escolhido deveria também ser capaz de lidar com essas características, e, ao mesmo tempo, ser parcimonioso em parâmetros, cumprindo seu papel de referência *naïve* para as comparações executadas. Observando-se essas premissas o modelo selecionado foi o conhecido ARMA-GARCH, detalhado em seguida.

Segundo a metodologia de Box e Jenkins (1976), um processo auto-regressivo de médias móveis ARMA (p, q) é definido por:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (23)$$

onde ε_t é um ruído branco com média zero e variância σ^2 . Os termos $\phi_0, \phi_1, \dots, \phi_p$ e $\theta_1, \dots, \theta_q$ são parâmetros que devem ser estimados. Esse é um modelo linear na média.

Sabendo-se, no entanto, que as séries financeiras possuem variância residual não-constante, foi também utilizado um modelo com heterocedasticidade condicional auto-regressiva ARMA (p, q) – GARCH (p, q) (COELHO et al., 2008; MEADE, 2002), que é um modelo linear na média e não-linear na variância. Assim, considerando a equação 23 como uma das equações gerais do modelo, acrescentam-se:

$$\varepsilon_t = v_t \sqrt{h_t}, v_t \sim \text{i.i.d} (0, 1) \quad (24)$$

$$h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}^2 \quad (25)$$

onde h_t^2 é a variância condicional no período t , e α e β são parâmetros estimáveis. Como no caso deste estudo, esse modelo foi adotado apenas como referência para o desempenho mínimo admitido para os modelos neurais, a estrutura utilizada também foi mínima, ou seja, ARMA (1,1)-GARCH(1,1), reduzindo as equações gerais para:

$$x_t = \phi x_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (26)$$

$$\varepsilon_t = v_t \sqrt{h_t}, v_t \sim \text{i.i.d} (0, 1) \quad (27)$$

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}^2 \quad (28)$$

Ressalta-se por fim, que, devido à necessidade de estacionariedade das séries para a aplicação desse modelo, ele foi implementado somente com as séries de retornos das ações da carteira elencada para essa pesquisa.

6. ESSENCIALIDADES SOBRE MERCADO DE CAPITALIS

Não há aqui o objetivo de realizar uma longa fundamentação teórica sobre o mercado de capitais e as operações da bolsa de valores, uma vez que esse não é foco desse trabalho. Contudo, faz-se necessário abordar alguns tópicos essenciais para o entendimento, tanto dos critérios adotados para a eleição dos ativos que compõem a carteira utilizada neste trabalho, quanto da elaboração da estratégia negocial na qual se fundamenta a métrica desenvolvida neste estudo, a TEN (taxa de eficiência negocial).

As operações que acontecem no ambiente (virtual) da Bovespa estão inseridas em um dos mercados disponíveis, que podem ser classificados como:

- a vista;
- a termo;
- futuro;
- de opções.

As séries temporais dos preços das ações utilizadas neste trabalho referem-se exclusivamente a cotações do mercado a vista. Por essa razão, o detalhamento ficará restrito a esse mercado, no qual se realizam as operações de compra e venda de valores mobiliários emitidos por empresas de capital aberto, com prazo de liquidação física e financeira regulamentado pela CBLC (Companhia Brasileira de Liquidação e Custódia).

Conforme essa regulamentação, os negócios são realizados a partir das cotações em tempo real, a liquidação física (entrega dos títulos) ocorre em D+2 (dois dias úteis) e a liquidação financeira (pagamento ou recebimento dos valores) em D+3 (três dias úteis) (PINHEIRO, 2009).

Outra característica desse mercado, quanto às ações de empresas de capital aberto, é que esse papéis são negociados em lotes, denominados lotes padrão, que normalmente contêm 100 ações, e cuja cotação é específica para esse tipo de lote. Isso acontece porque o investidor pode também adquirir os chamados lotes fracionários,

cuja cotação é diferente da cotação do lote padrão, em geral esses lotes são ligeiramente mais caros para a compra e o contrário para a venda. A codificação dos lotes fracionários segue a codificação do mercado a vista acrescido de um F, tal como BBAS3F (lote fracionário de ação ordinária do Banco do Brasil). Neste trabalho, todas as cotações são de lotes padrão.

As ações, conforme os direitos que outorgam, podem ser ordinárias (ações com direito a voto) ou preferenciais (ações com preferência na distribuição de resultados e proventos). De acordo com a Lei das Sociedades Anônimas (Lei 10.303/2001 que alterou a lei 6.404/76 com relação a emissão de ações), uma empresa terá que emitir obrigatoriamente, no mínimo, 50% do seu capital social em ações ordinárias, e no máximo, 50% em ações do tipo preferencial. As ações preferenciais recebem participações nos resultados (dividendos) 10% maiores que as ações ordinárias e podem passar a ter direito a voto se houver referência no estatuto da empresa, ou se a empresa deixar de distribuir dividendos por mais de três anos.

Além dessa classificação básica, existe uma série de outros papéis relacionados às ações que também são cotados e negociados no mercado a vista, como pode ser verificado na Tabela 03.

Tabela 3 - Códigos e especificações de ações negociadas no mercado a vista além das ON e PN

Código	Especificação do Papel
PNA	Ações Preferenciais Nominativas Classe A
PNB	Ações Preferenciais Nominativas Classe B
PNC	Ações Preferenciais Nominativas Classe C
PND	Ações Preferenciais Nominativas Classe D
PNE	Ações Preferenciais Nominativas Classe
PNF	Ações Preferenciais Nominativas Classe F
PNG	Ações Preferenciais Nominativas Classe G
PNH	Ações Preferenciais Nominativas Classe H
PNV	Ações Preferenciais Nominativas Com Direito A Voto
OR	Ações Ordinárias Nominativas Resgatáveis
PRA	Ações Preferenciais Nominativas Classe A Resgatáveis
PRB	Ações Preferenciais Nominativas Classe B Resgatáveis
PRC	Ações Preferenciais Nominativas Classe C Resgatáveis
PRD	Ações Preferenciais Nominativas Classe D Resgatáveis
PRE	Ações Preferenciais Nominativas Classe E Resgatáveis
PRF	Ações Preferenciais Nominativas Classe F Resgatáveis
PRG	Ações Preferenciais Nominativas Classe G Resgatáveis
PRH	Ações Preferenciais Nominativas Classe H Resgatáveis
PNR	Ações Preferenciais Nominativas Resgatáveis
PRV	Ações Preferenciais Nominativas Com Direito A Voto Resg
ON P	Ações Ordinárias Nominativas Com Direitos Diferenciados

PNA P	Ações Preferenciais Nominativas Classe A C/ Direitos Difer
PNB P	Ações Preferenciais Nominativas Classe B C/ Direitos Difer
PNC P	Ações Preferenciais Nominativas Classe C C/ Direitos Difer
PND P	Ações Preferenciais Nominativas Classe D C/ Direitos Difer
PNE P	Ações Preferenciais Nominativas Classe E C/ Direitos Difer
PNF P	Ações Preferenciais Nominativas Classe F C/ Direitos Difer
PNG P	Ações Preferenciais Nominativas Classe G C/ Direitos Difer
PNH P	Ações Preferenciais Nominativas Classe H C/ Direitos Difer
PN P	Ações Preferenciais Nominativas Com Direitos Diferenciados
PNV P	Ações Preferenciais Nominativas Classe V C/ Direitos Difer
ON P	Ações Ordinárias Nominativas Com Direitos Diferenciados
OR P	Ações Ordinárias Nominativas Resgatáveis C/ Direitos Dif
PRA P	Ações Preferenciais Nominativas Classe "A" Resg. C/ Dir.Dif
PRB P	Ações Preferenciais Nominativas Classe "B" Resg. C/ Dir.Dif
PRC P	Ações Preferenciais Nominativas Classe "C" Resg. C/ Dir.Dif
PRD P	Ações Preferenciais Nominativas Classe "D" Resg. C/ Dir.Dif.
PRE P	Ações Preferenciais Nominativas Classe "E" Resg. C/ Dir.Dif
PRF P	Ações Preferenciais Nominativas Classe "F" Resg. C/ Dir.Dif.
PRG P	Ações Preferenciais Nominativas Classe "G" Resg. C/ Dir.Dif
PRH P	Ações Preferenciais Nominativas Classe "H" Resg. C/ Dir.Dif
PR P	Ações Preferenciais Nominativas Resgatáveis C/ Direitos Dif
PRV P	Ações Preferenciais Nominativas Resg. C/ Dir.Dif. E Dir.Voto

Fonte: BM&FBovespa (2012)

As cotações de ações utilizadas neste estudo referem-se apenas às ações ON e PN puras, ou seja, sem outros direitos agregados e não resgatáveis. Essa escolha se deu pelo fato de que são esses os papéis com maior volume de negócios por pregão.

Ainda quanto à cotação das ações no mercado a vista, cabe ressaltar que o recebimento de proventos, sejam eles, dividendos, juros sobre o capital próprio ou bonificações, geram ajustes nas cotações de mercado das ações no dia útil seguinte à realização do pagamento, o que se convencionou denominar de ações "Com" e ações "Ex", isto é, as ações "Com" ou cheias são aquelas que dão ao investidor que as adquirir o direito de receber os proventos, enquanto as ações "Ex" (vazias), são aquelas que acabaram de pagar proventos e tiveram, portanto, seus preços ajustados (PINHEIRO, 2009).

Esse ajuste de preço também ocorre quando as empresas realizam a operação de *split* (desdobramento), que, conforme Pinheiro (2009), tem se tornado muito comum nos últimos anos entre as empresas. Nessa operação as empresas emitem novas ações que podem variar de uma a centenas para cada ação antiga, levando, evidentemente, a uma abrupta queda da cotação da ação, sem que, com isso, nem o investidor, nem a empresa tenham alterado seu capital. E é essa a principal diferença

em relação à bonificação, na qual, a empresa entrega novas ações aos acionistas, causando também uma queda da cotação da ação. Porém, isso ocorre devido a uma incorporação de reservas que aumenta o capital da empresa e uma atualização da cota de propriedade de capital do acionista.

A dimensão dessa queda na cotação pode ser entendida pela fórmula geral de ajuste de preços ex-teóricos dada por:

$$P_{ex} = \frac{P_c + (S \cdot Z) - D - J - VET}{1 + B + S} \quad (29)$$

onde P_{ex} é o preço ex-teórico, P_c é o preço "Com" exatamente antes do pagamento do provento, S é o percentual de subscrição (em caso de bônus de subscrição), Z é o valor da ação a ser subscrita, D é o valor do dividendo por ação, J é o valor do juros recebidos por ação, VET é o valor econômico teórico, resultante do recebimento do provento na forma de outro ativo, e.g. uma ação de uma empresa coligada, e B é o percentual de bonificação ou *split*.

Pela análise da fórmula de ajuste, assim como do banco de dados de operações realizadas no período de 2009 a 27/08/2012, foi possível concluir que as empresas que realizaram nesse período operações de *split* ou bonificações causaram forte queda na cotação de um dia para o outro. Por essa razão foi adotado o critério de excluir esses papéis daqueles candidatos à carteira teórica dessa dissertação. Quanto aos demais ajustes, sobretudo aqueles referentes à distribuição de dividendos e juros de capital próprio, adotou-se o critério de privilegiar aquelas ações com pagamentos com frequência mensal, o que faz com que esses ajustes sejam mais distribuídos e menores, ou seja, com dimensões suficientemente pequenas para não interferir significativamente na oscilação das cotações. O que é o caso dos bancos, que são notoriamente conhecidos no mercado por distribuir dividendos ou juros sobre o capital próprio mensalmente. Com esses cuidados procurou-se minimizar as interferências não oriundas das forças do próprio mercado nas cotações das ações analisadas neste estudo.

Outra questão importante para o entendimento deste trabalho está relacionada com o posicionamento dos agentes do mercado. Um investidor pode se posicionar no mercado de duas formas:

- posição comprada;
- posição vendida.

A primeira delas é mais conhecida do investidor comum, ou seja, é aquela em que o investidor compra a ação e espera que ela se valorize para vendê-la e realizar seu lucro, posicionamento consistente com o mercado em alta. Já a segunda posição é mais conhecida do investidor profissional, e consiste na venda da ação com preço em queda e a posterior compra por um preço mais baixo, realizando assim o lucro da operação, esse posicionamento é, portanto, consistente com o mercado em baixa. Existem basicamente duas formas de se operar vendido, a primeira delas é em operações de *day-trade*, quando o investidor vende a descoberto e recompra a ação no mesmo dia obrigatoriamente; a segunda acontece por meio do aluguel de ações, que é calculado por meio de uma taxa anual de juros. Nessa modalidade o investidor realiza a venda alugada e depois recompra a ação por um preço mais baixo, realizando seu lucro (ABE, 2009).

6.1 Estratégia negocial

A estratégia negocial adotada neste trabalho, aqui denominada de "voando por instrumentos", e que fundamenta o cálculo da TEN - taxa de eficiência negocial - é consistente com ambos os posicionamentos de mercado. Essa estratégia se baseia na idéia de que se o investidor tiver conhecimento (por meio da previsão) das cotações de uma determinada ação nos próximos 10 pregões, os posicionamentos a serem tomados para que a rentabilidade no período seja máxima, desprezando-se os custos de transação, serão o de comprado no preço mínimo e vendido no preço máximo, independentemente da ordem cronológica dos eventos.

Portanto, a estratégia negocial "voando por instrumentos" obedece aos seguintes movimentos:

- a compra e a venda teóricas devem acontecer dentro do horizonte de previsão, ou seja, 10 passos a frente;
- a compra se dará no dia em que a previsão indicar a cotação mínima, mas será realizada pela cotação real do dia;
- a venda se dará no dia em que a previsão indicar o preço máximo, mas também será fechada pelo preço real.

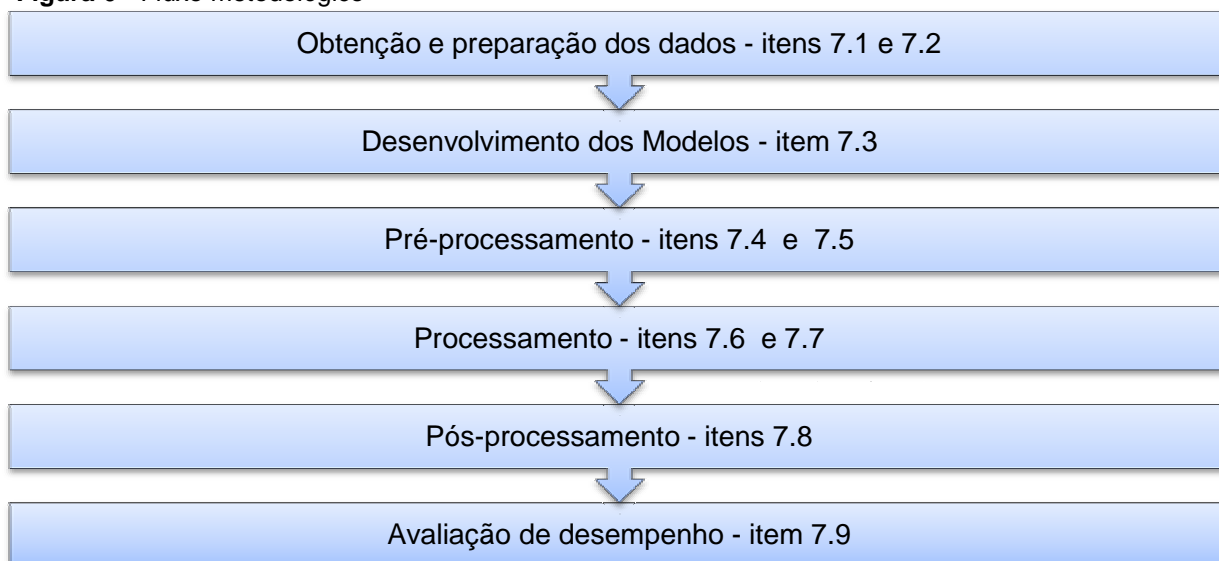
A taxa de eficiência negocial, portanto, consiste no percentual da rentabilidade máxima possível que foi conseguida com a estratégia adotada. Desta forma, fica fácil ver que a TEN de 100% indica que a previsão acertou exatamente os dias de preço máximo e mínimo reais, enquanto uma TEN de -100% indica o exato oposto. Fica claro também que essa métrica indica a capacidade do modelo previsor em captar a tendência de movimentação dos preços. Para maiores detalhes quanto ao algoritmo de cálculo da TEN, verificar o item 7.9.2.

7. METODOLOGIA

Para a consecução desta pesquisa estabeleceu-se um método em 6 etapas, conforme descrito a partir do item 7.1. Quanto aos aspectos computacionais, para a manipulação da base de dados, preparação das séries temporais e realização de testes para a análise dos resultado, foi utilizado o aplicativo Microsoft Office Excel® 2007; para implementação computacional das redes neurais, do modelo estatístico de referência, das transformações de dados, da apuração das características estatísticas de cada série e dos testes estatísticos de estacionariedade e normalidade das séries, utilizou-se o software MATLAB® R2010a.

O fluxo da metodologia adotada pode ser verificado na Figura 06, que traz o esquema relacionado aos itens abaixo com seu detalhamento.

Figura 6 - Fluxo metodológico



Fonte: o autor

7.1. Base de dados

A base de dados contendo as cotações históricas diárias de todos os papéis-mercado negociados nos anos de 2009, 2010, 2011 e 2012 (até 27/08/2012) foi obtida (*downloaded*) diretamente do site da BM&FBovespa em arquivos tipo texto (extensão

txt). É também disponibilizado o *layout* do arquivo para que se possa transferi-lo para um software com recursos para sua manipulação. Para este trabalho os arquivos foram transferidos para o Excel® para filtragem e preparação das séries temporais, que, posteriormente, foram transformadas em arquivos individuais com extensão ".dat", por ação e por categoria de preços (abertura, máximo, mínimo, médio e fechamento), ou seja, cada ação escolhida gerou 5 séries temporais com 906 elementos para submissão aos modelos de previsão.

7.2. Carteira de ações

Para montagem da carteira de ações que faz parte deste trabalho, estabeleceram-se duas categorias de critérios:

- Critério temporal:
 - período inicial delimitado ao ano de 2009 para se evitar os movimentos atípicos referentes à crise sub-prime do ano de 2008.
- Critérios para as ações:
 - negociadas no mercado à vista;
 - cotações referentes a lotes padrão;
 - ordinárias nominativas e preferenciais nominativas puras sem reclassificações por classes ou resgatáveis ou com direito a voto;
 - negociadas em todos os pregões no período determinado;
 - com volume de negociações acima de R\$ 100 milhões por pregão;
 - sem a ocorrência, no período, de operações de desdobramento (split) ou bonificações em ações.
 - com política de distribuição de dividendos ou juros de capital próprio com frequência mensal.

A partir dos critérios acima, foi montada uma carteira com 03 ações (conforme Tabela 04) que gerou 15 séries temporais para análise.

Tabela 4 - Carteira de ações

Ação	Empresa	Setor	Listagem
BBAS3	Banco do Brasil SA	Intermediários Financeiros / Bancos	NM
BBDC3	Banco Bradesco	Intermediários Financeiros / Bancos	N1
BBDC4	Banco Bradesco	Intermediários Financeiros / Bancos	N1

Legenda: NM=empresas listadas no Novo Mercado, N1=empresas listadas no nível 1 de governança corporativa.
Fonte: o autor.

Cabe ressaltar que os papéis escolhidos, além de estarem em pleno acordo com os critérios estabelecidos, são representativos dos dois maiores em bancos de capital nacional do país, além de serem também os dois maiores em volume de depósitos a vista e número de agências, conforme informações do Banco Central do Brasil. Esses fatos garantem a imagem de solidez das duas empresas. Adicionalmente, são empresas listadas na Bovespa nos dois mais altos níveis de governança corporativa, a saber: Novo Mercado e Nível 1 de governança; essa situação garante um alto nível de transparência de suas informações ao mercado.

Por fim, essas características, aliadas aos critérios estabelecidos, garantem em grande medida, que o movimento dos preços dessas ações sejam pouco influenciados por fatores não oriundos das próprias forças de mercado.

7.3. Desenvolvimento dos modelos de previsão

7.3.1. Redes Neurais

Para este trabalho foram desenvolvidas duas arquiteturas de redes dinâmicas visando-se o reconhecimento temporal de padrões, isto é, a previsão de valores de séries temporais. A primeira arquitetura selecionada foi a rede focada atrasada no tempo (FTDNN - *Focused Time-Delay Neural Network*), cuja característica dinâmica é introduzida apenas pela camada de entrada; e uma rede não-linear auto-regressiva

com entradas exógenas (NARX - *Non linear Autoregressive Network with Exogenous Inputs*), cujo aspecto dinâmico é introduzido tanto pela camada de entrada, como por valores de saída que realimentam a rede, daí essa arquitetura ser considerada de caráter recorrente.

O processo empírico de definição das topologias e parâmetros internos de cada modelo teve como base as seguintes premissas:

- as topologias e parâmetros internos em ambas as arquiteturas devem ser semelhantes, uma vez que o objetivo é a comparação entre os desempenhos dos modelos;
- as topologias devem ser robustas e capazes de apresentar bons desempenhos independentemente das características das séries apresentadas, uma vez que o objetivo essencial do trabalho é a comparação entre modelos e não o atingimento da melhor performance possível em cada série individualmente.

Para definição do número de entradas ou ordem de predição, o procedimento empírico adotado foi o de medir o desempenho de previsão 10 passos a frente de cada rede para a série temporal estacionária de referência, Mackey-Glass, detalhada adiante, por meio da métrica MAPE (também explicada adiante), sucessivamente para ordens de predição a partir de 2 até que o desempenho atingisse o erro mínimo e depois se deteriorasse. Lembrando que para cada ordem, 20 treinos foram realizados, adotando-se o menor MAPE; as médias dos MAPEs, entre as redes FTDNN e NARX, obtidas para cada ordem de predição podem ser verificados na Figura 07, que traz os valores obtidos, bem como seu gráfico. Os resultados indicaram que a ordem de predição que gera o menor erro foi 8. Cabe destacar que esse procedimento foi realizado com todas as topologias candidatas e o resultado mostrado na Figura 07 refere-se à topologia eleita.

Figura 7 - Definição da ordem de Predição

Fonte: o autor

Considerando-se as premissas adotadas, estabeleceram-se topologias candidatas com uma e duas camadas escondidas com o número de neurônios em cada uma delas variando entre 7 e 17 neurônios conforme o método de Fletcher-Gloss (SILVA et al., 2010), que segue o seguinte critério:

$$2 \cdot \sqrt{n} + n_2 \leq n_1 \leq 2 \cdot n + 1 \quad (30)$$

onde n é o número de entradas da rede, n_1 é a quantidade de neurônios na camada escondida e n_2 é o número de neurônios na camada de saída.

Para os testes das topologias candidatas foram utilizadas duas séries temporais de referência:

- série caótica gerada artificialmente por meio da equação diferencial de atraso temporal Mackey-Glass (SAMANTA, 2011) com a condição inicial de $y(0) = 0,9$ e $\tau = 30$:

$$\frac{dy(t)}{dt} = \frac{0,2y(t-\tau)}{1 + y^{10}(t-\tau)} - 0,1y(t) \quad (31)$$

que, além de ser uma série largamente utilizada na literatura relativa à previsão de séries temporais (GOMES-GIL et al., 2011; GONZALES, 2009; KATAGIRI et

al.,2011; SANDMANN, 2006; SCHOPF, 2007), é uma série que apresenta forte estacionariedade, tornando-a uma série de fácil previsão para os modelos neurais;

- série de preços de uma das ações pertencentes à carteira de referência deste trabalho, escolhida aleatoriamente. No caso, a série utilizada foi a de preços mínimos da ação ordinária do Banco do Brasil (BBAS3).

A tarefa apresentada para o teste das topologias candidatas foi a mesma que será solicitada aos modelos durante todo o trabalho, ou seja, realizar a previsão das séries apresentadas para 10 passos a frente. Ressalta-se aqui que não são 10 passos a frente com iterações da série real a cada passo, o que se configuraria na metodologia *one step ahead*, mas sim na previsão real de 10 passos a frente, ou seja, a cada previsão a reentrada no sistema é do próprio passo previsto anteriormente.

A topologia e parâmetros internos que foram adotados após os testes de performance, tanto para a rede FTDNN quanto para a rede NARX, foram:

- duas camadas escondidas com 10 neurônios cada uma;
- função tangente hiperbólica para ativação dos neurônios da camada escondida definida por:

$$g_1(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (32)$$

- função linear de transferência para a camada de saída definida por:

$$g_2(x) = x \quad (33)$$

- pesos sinápticos iniciados aleatoriamente;
- no caso da arquitetura NARX acrescenta-se uma reentrada com dois períodos de atraso;
- critérios de parada:

- 200 épocas para a rede FTDNN e 50 épocas para a rede NARX, essa diferença se justifica, tanto pela maior capacidade de armazenamento de memória de curto prazo da rede NARX, como pela sua maior lentidão no processamento computacional;
- erro quadrático inferior a 10^{-5} ou $\varepsilon_p \leq 10^{-5}$;
- taxa de aprendizagem variável $10^{-3} \leq \mu \leq 10^{10}$;

Os escores (EDP - escore de desempenho de previsão entre 0 e 10 que será detalhado adiante) obtidos pela topologia eleita, portanto, podem ser observados na Tabela 05:

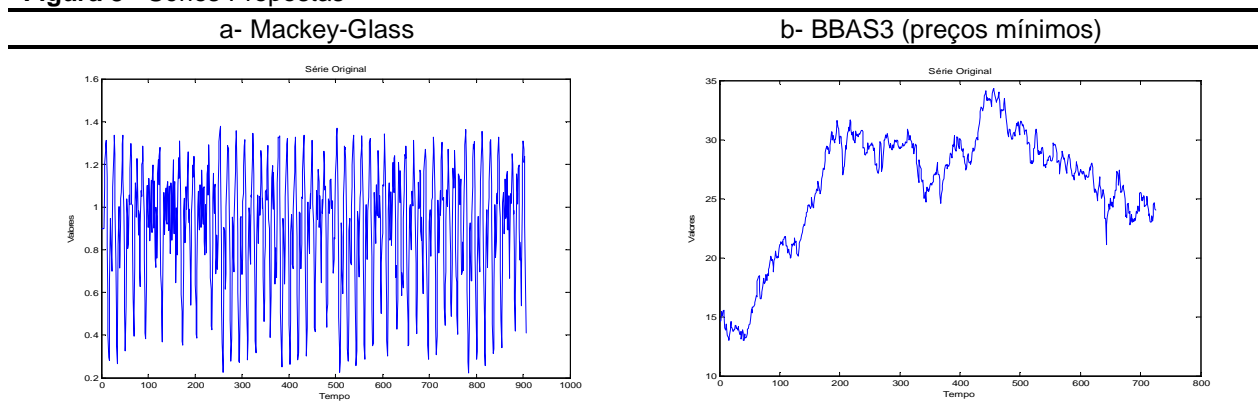
Tabela 5 - Performance das topologias eleitas

Série	Rede	EDP
Mackey-Glass	FTDNN	9,820
BBAS3 - preços mínimos	FTDNN	8,472
Mackey-Glass	NARX	9,913
BBAS3 - preços mínimos	NARX	9,573

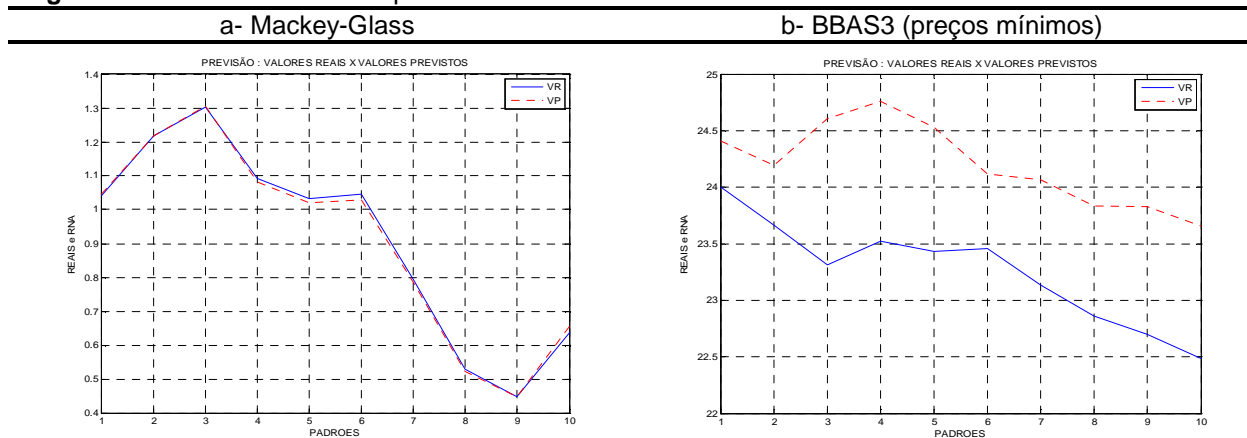
Fonte: o autor

As performances preditivas das duas topologias também podem ser verificadas nas Figuras de 08 a 10, "a" e "b", que trazem os gráficos das séries originais e da previsão de 10 passos a frente.

Figura 8 - Séries Propostas

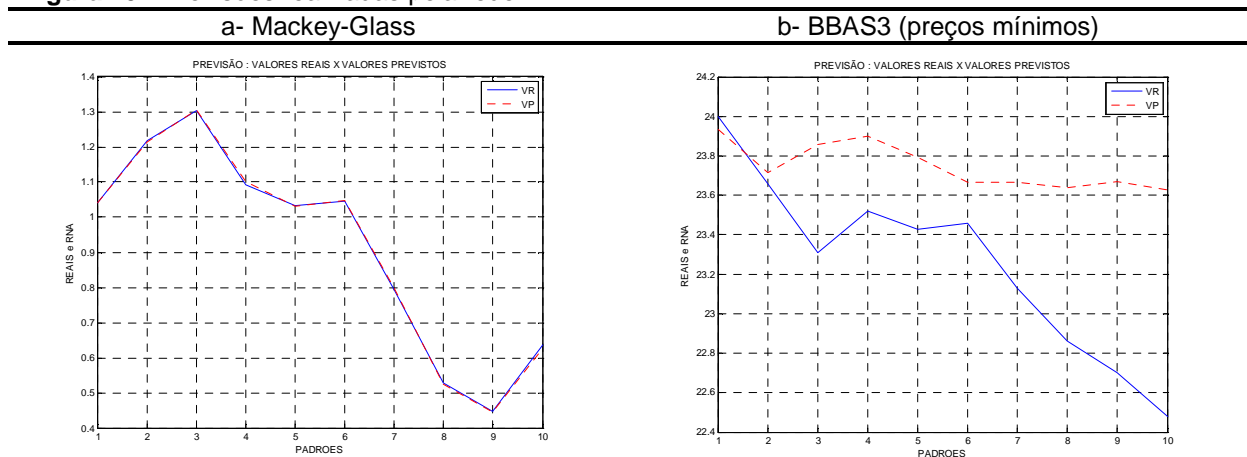


Fonte: o autor

Figura 9 - Previsões realizadas pela rede FTDNN

Legenda: VR = valores reais, VP = valores previstos

Fonte: o autor

Figura 10 - Previsões realizadas pela rede NARX

Legenda: VR = valores reais, VP = valores previstos

Fonte: o autor

7.3.2. Modelo ARMA-GARCH

Como modelo estatístico de referência utilizou-se o ARMA-GARCH, ou seja, um modelo largamente utilizado para análise e previsão de séries temporais financeiras, que sempre apresentam heterocedasticidade condicional. Como esse modelo foi utilizado apenas como padrão de desempenho mínimo para as redes neurais, estabeleceu-se uma estrutura paramétrica também mínima, ou seja, ARMA (1,1) -

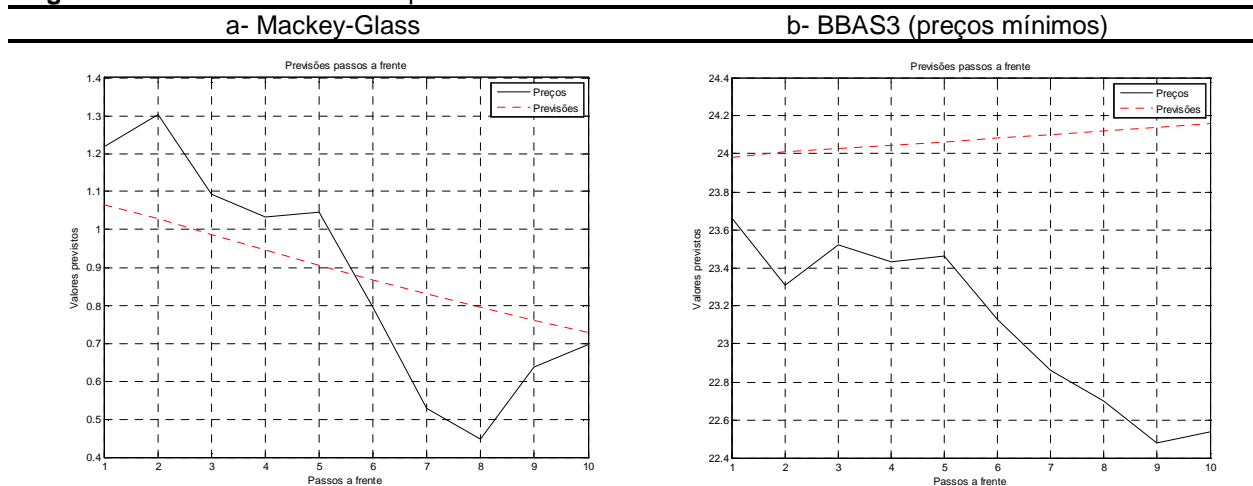
GARCH (1,1), que obteve um desempenho, relativamente às séries de teste (Mackey-Glass e BBAS3 - preços mínimos) conforme a Tabela 06 e a Figura 11.

Tabela 6 - Performance do modelo ARMA-GARCH

Série	EDP
Mackey-Glass	4,472
BBAS3 - preços mínimos	4,467

Fonte: o autor

Figura 11 - Previsões realizadas pelo modelo ARMA-GARCH



Fonte: o autor

7.4. Análise estatística pré e pós transformação de dados

Para verificação dos efeitos de cada modalidade de transformação de dados, foi realizada uma descrição estatística básica de cada série, apurando sua média, assimetria e coeficiente de variação, antes e depois de cada transformação. Optou-se pela utilização do coeficiente de variação, na análise de resultados, em função da modificação brusca da escala de valores das séries de preços, causada pela transformação Box-Cox, inviabilizando a comparação entre as variâncias. Uma vez que as transformações por diferenças de primeira ordem e por retornos logarítmicos buscam tornar as séries estacionárias ou mais próximas dessa condição, enquanto a transformação Box-Cox busca tornar a distribuição da série mais próxima da normalidade, para verificação da estacionariedade das séries, utilizou-se o teste Dickey-

Fuller aumentado (HAMILTON, 1994), que denota com 1 a rejeição da hipótese nula de presença de raízes unitárias e, com 0, a falha em se rejeitar tal hipótese, ou seja, 1 para presença de estacionariedade e 0 para ausência. Quanto à normalidade da distribuição da série utilizou-se o teste Jarque-Bera (JARQUE; BERA, 1987) como discriminador, que denota com 1 a rejeição da hipótese nula de normalidade e, com 0, a falha em rejeitá-la, ou seja, 0 para presença de normalidade e 1 a ausência dessa condição.

Os valores obtidos na análise estatística podem ser verificados na Tabela 10 do capítulo 8, análise dos resultados.

7.5. Transformação de dados

Como um dos objetivos deste trabalho é a avaliação do desempenho das redes quando realizam previsões com dados transformados, foram selecionados os três métodos de transformação de dados mais largamente conhecidos e utilizados no contexto de previsão de séries temporais:

- diferenças de primeira ordem;
- retornos logarítmicos;
- transformação Box-Cox.

Nesta etapa os dados originais de cada uma das séries de preços selecionadas são transformados e separados em suas porções de treinamento, correspondendo a 80% da série e, para teste, os restantes 20%. A parcela de treinamento é submetida aos modelos neurais, ou seja, cada série original gera três séries de dados transformados que são submetidas a cada um dos modelos neurais, FTDNN e NARX. Portanto, de forma geral, relativamente a cada série, são realizadas nove previsões, uma para cada tipo de transformação em cada modelo neural, duas para dados não transformados e uma realizada pelo modelo estatístico que usa, por definição, os retornos logarítmicos.

A normalização dos dados, neste trabalho, não foi considerada uma transformação de dados, e somente foi utilizada para evitar a saturação dos neurônios de modo que os valores apresentados à rede sejam compatíveis com as faixas de variação dinâmica

das funções de ativação de cada neurônio (SILVA et al., 2010), no caso, a função tangente hiperbólica que tem o domínio no intervalo $[-1, 1]$. Portanto, a normalização só foi utilizada quando, ou os dados não seriam transformados, ou a própria transformação já não colocava os dados no intervalo em questão, que é o caso da transformação Box-Cox.

Quanto aos parâmetros λ_1 e λ_2 da transformação Box-Cox, para este trabalho λ_2 foi fixado em 0, pois, por se tratar de preços de ações não contém dados negativos. A estimação do parâmetro λ_1 , por sua simplicidade e acurácia, foi realizada por um algoritmo de otimização computacional que maximiza sua função de log-verossimilhança por meio do método de busca simplex de Lagarias et al. (1998), já automatizado no software MATLAB®.

7.6. Treinamento das redes de previsão e reversão

Como o aspecto dinâmico e temporal é introduzido, em ambas as arquiteturas, por meio da camada de entrada, ou seja, proporciona-se à rede uma memória de curto prazo (HAYKIN, 2001) transformando o vetor de entrada $[x]_{1 \times N}$, onde N é o número de elementos da série temporal destinados ao treinamento da rede, em uma matriz $\{x\}_{N-p-1 \times p+1}$, em que, a cada linha da matriz estarão dispostos os valores da série de x_t a x_{t+p+1} , separa-se então, a coluna $p+1$, que será então designada como $[d]$, ou seja, vetor de valores desejados. Fica claro então que, para cada valor x_t desejado, a rede treinará com p valores passados, sendo p , portanto, a ordem de predição ou o alcance da memória de curto prazo da rede. Essas entradas atrasadas no tempo permitem que a rede saiba de parte da história da variável no tempo t , possibilitando a solução de problemas de decisão mais complexos, especialmente aqueles dependentes do tempo (KIM; SHIN, 2007). Para este trabalho estabeleceu-se $N = 725$, que corresponde a 80% do total de elementos da série e $p = 8$.

No caso da arquitetura NARX, para este trabalho, acrescenta-se à entrada da rede mais dois valores de saída que realimentam a rede, ou seja, cada valor y_t previsto, também depende de y_{t-1} e y_{t-2} , além dos x_{t-1} , x_{t-2} , ..., x_{t-8} , o que reforça a capacidade de manutenção da memória de curto prazo da rede.

Conforme os objetivos deste trabalho, foram utilizadas duas redes idênticas do ponto de vista arquitetônico e topológico por modelo, uma que realiza a previsão com os dados transformados e outra que realiza a reversão da transformação dos dados previstos. Essas redes vão apenas diferir quanto ao vetor de valores desejados, na rede de previsão o vetor se constitui dos valores transformados da série, enquanto, na rede treinada para reverter a transformação dos valores previstos, o vetor de valores desejados se constitui dos valores originais da série. Para as duas redes as matrizes de entrada são idênticas e se constituem de valores transformados.

O algoritmo de treinamento utilizado para ambas as arquiteturas foi o Levenberg-Marquardt, pela sua rapidez, eficiência e economia de esforço computacional.

Ressalta-se também que, devido à inicialização aleatória de pesos sinápticos e características do processo de convergência, cada treino resulta em desempenhos diferentes, adotou-se então o critério de, para cada série, realizar 30 treinos e optar por aquele que resultou em melhor performance.

7.7. Previsão de dados transformados e não transformados

Nesta etapa, a tarefa designada para cada rede já treinada foi realizar a previsão dos 10 próximos valores fora da amostra de treinamento. Esse horizonte de previsão foi adotado, basicamente, por duas razões: primeiro, observou-se, na fase preliminar de desenvolvimento dos modelos, que a qualidade das previsões se deteriorava mais rapidamente após esse horizonte; segundo, esse horizonte se adapta melhor à estratégia negocial adotada, representando um período correspondente a duas semanas de pregão, o que dá um tempo razoável ao investidor para tomar suas decisões, além de evidenciar melhor a rentabilidade do negócio no curto prazo.

Evidentemente, para cada uma das arquiteturas utilizadas, FTDNN e NARX, uma rede idêntica foi treinada com os dados transformados em cada uma das modalidades de transformação, além daquelas treinadas com os dados não transformados.

Para este trabalho, adotou-se uma rotina em que as redes deveriam realizar as previsões, simulando uma situação real em que o investidor, de fato, não possui os

valores futuros para comparar e iterar com o sistema a cada dia, fazendo com que a rede dependa de suas próprias previsões anteriores para realizar a predição do próximo valor. Estabelecendo x_t como valores da série original de treinamento (transformados ou não) e y_t como valores previstos, e lembrando que, como as séries apresentadas aos modelos tinham extensão de 725 períodos, a tarefa se torna prever os preços para os períodos 726º ao 735º. Esquemáticamente, essa rotina pode ser observada na Figura 12.

Figura 12 - Rotina geral de previsão com $N=725$ e $p=8$, para 10 passos à frente

Valores submetidos à rede								Valores Previstos
x_{718}	x_{719}	x_{720}	x_{721}	x_{722}	x_{723}	x_{724}	x_{725}	y_{726}
x_{719}	x_{720}	x_{721}	x_{722}	x_{723}	x_{724}	x_{725}	y_{726}	y_{727}
x_{720}	x_{721}	x_{722}	x_{723}	x_{724}	x_{725}	y_{726}	y_{727}	y_{728}
x_{721}	x_{722}	x_{723}	x_{724}	x_{725}	y_{726}	y_{727}	y_{728}	y_{729}
x_{722}	x_{723}	x_{724}	x_{725}	y_{726}	y_{727}	y_{728}	y_{729}	y_{730}
x_{723}	x_{724}	x_{725}	y_{726}	y_{727}	y_{728}	y_{729}	y_{730}	y_{731}
x_{724}	x_{725}	y_{726}	y_{727}	y_{728}	y_{729}	y_{730}	y_{731}	y_{732}
x_{725}	y_{726}	y_{727}	y_{728}	y_{729}	y_{730}	y_{731}	y_{732}	y_{733}
y_{726}	y_{727}	y_{728}	y_{729}	y_{730}	y_{731}	y_{732}	y_{733}	y_{734}
y_{727}	y_{728}	y_{729}	y_{730}	y_{731}	y_{732}	y_{733}	y_{734}	y_{735}

Fonte: o autor

7.8. Reversão dos dados Transformados

7.8.1. Reversão algébrica dos dados transformados por diferenças de primeira ordem

Como, nesse caso, as redes previram as diferenças e não os preços propriamente ditos, para reverter as diferenças previstas para preços previstos, adotou-se o procedimento de apresentar o último valor não transformado da série de treinamento, e, a partir daí, somar recursivamente as diferenças e desprezar o primeiro valor da série, ou seja, considerando $y_{d(t)}$ como diferença prevista, y_t como preço revertido e x_t como

valor da série original, aplicou-se recursivamente o algoritmo descrito nas equações 34 e 35:

$$y_t = x_t \quad \text{para } t = 725 \quad (34)$$

$$y_t = y_{t-1} + y_{d(t-1)} \quad \text{para } 726 \leq t \leq 735 \quad (35)$$

7.8.2. Reversão algébrica dos dados transformados por retornos logarítmicos

Considerando $y_{r(t)}$ como retorno previsto, y_t como preço revertido e x_t como valor da série original, e, de forma similar à transformação por diferenças, adotou-se o procedimento de apresentar o último valor não transformado da série de treinamento, e, a partir daí, aplicar-se recursivamente o algoritmo descrito nas equações 36 e 37, e desprezar o primeiro valor:

$$y_t = x_t \quad \text{para } t = 725 \quad (36)$$

$$y_t = y_{t-1} \cdot e^{y_{r(t)}} \quad \text{para } 726 \leq t \leq 735 \quad (37)$$

7.8.3. Reversão algébrica dos dados transformados por Box-Cox

No caso da transformação Box-Cox as redes fazem a previsão de um preço transformado, baseado nos valores transformados da série de treinamento. Deste modo, considerando $y_t^{(\lambda)}$ como preço previsto transformado e y_t como preço previsto revertido e fazendo $\lambda_2 = 0$ e $\lambda_1 = \lambda$, aplicou-se para cada $y_t^{(\lambda)}$ para $726 \leq t \leq 735$ a equação 38.

$$y_t = \begin{cases} [y_t^{(\lambda)} \cdot \lambda + 1]^{1/\lambda}; & \lambda \neq 0 \\ e^{y_t^{(\lambda)}}; & \lambda = 0 \end{cases} \quad (38)$$

Cabe ressaltar que, com o algoritmo de estimação de λ utilizado, não houve para as séries deste trabalho, o caso de λ nulo.

7.8.4. Reversão neural dos dados transformados

Considerando y_t como preço revertido, $y_{t(t)}$ como diferença, retorno ou valor transformado pelo procedimento Box-Cox previstos e $x_{t(t)}$ como valor transformado da série de treinamento, o procedimento neural para reversão dos dados previstos se assemelha ao procedimento de previsão, como se pode verificar na Figura 13, com a diferença de que os valores que são reapresentados à rede são os valores transformados previstos e não os próprios valores revertidos, que são os valores desejados, como no caso da previsão.

Figura 13 - Rotina geral da reversão neural de dados

Valores submetidos à rede								Valores Revertidos
$x_{t(718)}$	$x_{t(719)}$	$x_{t(720)}$	$x_{t(721)}$	$x_{t(722)}$	$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	y_{726}
$x_{t(719)}$	$x_{t(720)}$	$x_{t(721)}$	$x_{t(722)}$	$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	y_{727}
$x_{t(720)}$	$x_{t(721)}$	$x_{t(722)}$	$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	y_{728}
$x_{t(721)}$	$x_{t(722)}$	$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	y_{729}
$x_{t(722)}$	$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	y_{730}
$x_{t(723)}$	$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	$y_{t(730)}$	y_{731}
$x_{t(724)}$	$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	$y_{t(730)}$	$y_{t(731)}$	y_{732}
$x_{t(725)}$	$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	$y_{t(730)}$	$y_{t(731)}$	$y_{t(732)}$	y_{733}
$y_{t(726)}$	$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	$y_{t(730)}$	$y_{t(731)}$	$y_{t(732)}$	$y_{t(733)}$	y_{734}
$y_{t(727)}$	$y_{t(728)}$	$y_{t(729)}$	$y_{t(730)}$	$y_{t(731)}$	$y_{t(732)}$	$y_{t(733)}$	$y_{t(734)}$	y_{735}

Fonte: o autor

7.9. Avaliação de desempenho

7.9.1 Métricas de erro

A avaliação do desempenho ou acurácia da previsão fora-da-amostra foi realizada pelas seguintes métricas:

Erro médio absoluto percentual, ou MAPE (*Mean Absolute Percentage Error*), definido por:

$$\text{MAPE} = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{y_t - x_t}{x_t} \right| \cdot 100 \quad (39)$$

onde n é o número de previsões, x_t é o valor original da amostra, y_t é o valor previsto já revertido ou não transformado, $n = 10$ e $726 \leq t \leq 735$.

Índice de Desigualdade de Theil (U-Theil) definido por:

$$\text{U - Theil} = \frac{\sqrt{\sum_{t=1}^n (y_t - x_t)^2 / n}}{\sqrt{\sum_{t=1}^n (x_t)^2 / n} + \sqrt{\sum_{t=1}^n (y_t)^2 / n}} \quad (40)$$

onde n é o número de previsões, x_t é o valor original da amostra, y_t é o valor previsto já revertido ou não transformado, $n = 10$ e $726 \leq t \leq 735$.

Enquanto o MAPE avalia a acurácia da previsão pela média das distâncias absolutas entre os valores previstos e reais, o índice U-Theil avalia o ajustamento da série prevista à série original, o que significa dizer que, quanto menores forem essas métricas mais acurada foi a previsão (COELHO et al., 2008).

O MAPE e o índice de desigualdade U-Theil, além de sua larga utilização na literatura relativa previsão de séries temporais, foram eleitos para este trabalho, tanto pela sua simplicidade, quanto pela sua capacidade de avaliação das distâncias entre valores previstos e reais, embora essas métricas falhem na avaliação da tendência tomada pelos valores previstos, sobretudo quanto se precisa tomar decisões de compra e venda de um ativo, ou seja, nem sempre, como se verá na próxima seção, análise de resultados, "muito próximo" significa "na mesma direção".

7.9.2 Taxa de eficiência negocial (TEN) - "voando por instrumentos"

Com o intuito de avaliar, tanto a captura da tendência assumida pelo conjunto dos preços previstos 10 dias à frente, quanto a eficiência da estratégia negocial estabelecida para este trabalho, foi desenvolvida uma métrica que estabelece o percentual da rentabilidade que o investidor teria obtido, em relação ao percentual máximo possível (desprezando-se os custos de transação) de rentabilidade, caso esse investidor tenha seguido cegamente os valores previstos pelas redes. Ou seja, fazendo uma analogia com um vôo por instrumentos, esse investidor somente poderia contar com as previsões para determinar, quando, no período dos próximos 10 dias, haveria o preço máximo e o mínimo, para, nesses dias, abrir posição vendida e comprada, respectivamente, ressaltando que, embora o investidor siga a previsão, as transações de compra e venda são efetivamente realizadas pelos preços reais daqueles dias.

Considerando x_t o valor original da amostra e y_t o valor previsto já revertido ou não transformado e $726 \leq t \leq 735$, temos o seguinte algoritmo descrito pelas equações 41 a 45:

$$\max [y_t] = y_{t(\max)} \rightarrow x_{t(a)} \quad (41)$$

$$\min [y_t] = y_{t(\min)} \rightarrow x_{t(i)} \quad (42)$$

$$\max [x_t] = x_{t(\max)} \quad (43)$$

$$\min [x_t] = x_{t(\min)} \quad (44)$$

$$TEN = \left[\frac{x_{t(a)} - x_{t(i)}}{x_{t(\max)} - x_{t(\min)}} \cdot \frac{x_{t(\min)}}{x_{t(i)}} \right] \cdot 100 \quad (45)$$

onde $x_{t(a)}$ é o preço da série original correspondente ao preço previsto máximo em t , $x_{t(i)}$ é o preço da série original correspondente ao preço previsto mínimo em t e $x_{t(\max)}$ e $x_{t(\min)}$ são respectivamente, os valores máximos e mínimos da série original correspondente à série prevista.

Fica claro que se houver perfeita correspondência de tendências, e os preços máximos e mínimos das séries previstas e reais coincidirem, a TEN será de 100% e, por oposição, se houver completa inversão de tendências entre as séries prevista e real, e o preço máximo da série prevista coincidir com o mínimo da série real e vice-versa a TEN será de -100%, estabelecendo assim o intervalo de variação dessa métrica.

7.9.3 Escore de desempenho de previsão (EDP)

Com o objetivo de obter um escore único de desempenho para cada previsão realizada com variação entre 0 e 10, considerando 0 o pior desempenho e 10 o melhor, agregou-se as métricas utilizadas (MAPE, U-Theil e TEN) por meio de uma média ponderada dos escores (de 0 a 10) relativos de cada métrica, considerando seus respectivos intervalos de variação; estabelecendo como pesos 1, 1 e 2, respectivamente.

O estabelecimento do peso 2 para a TEN visou dar equilíbrio ao escore, uma vez que as métricas MAPE e U-theil têm comportamentos semelhantes entre si e diferentes em relação à TEN, como pode ser observado na Tabela 07, que traz os coeficientes de correlação (ρ_{xyz}) entre essas métricas, apurados a partir de uma amostra de 120 previsões realizadas. Como pode ser verificado o coeficiente de correlação entre o MAPE e o U-Theil é bem próximo de 1, indicando uma correlação linear positiva e intensa enquanto os coeficientes de correlação entre MAPE e TEN e entre U-Theil e TEN situam-se em torno de -0,6, indicando uma correlação linear negativa e menos intensa.

Tabela 7 - Correlação entre as métricas adotadas

	MAPE	U-Theil	TEN
MAPE	1		
U-Theil	0,9885	1	
TEN	-0,6034	-0,6026	1

Fonte: o autor

Os intervalos de variação das métricas, adotados para este trabalho, tiveram como base principal a observação dos valores ocorridos durante o processamento das séries e podem ser verificados na Tabela 08:

Tabela 8 - Variação das métricas e escores relativos

Métricas	Intervalos de variação	Escores Relativos
MAPE	[0 25%]	0 a 10
U-Theil	[0 0,5]	0 a 10
TEN	[-100% 100%]	0 a 10

Fonte: o autor

Considerando, para cada uma das métricas, seu intervalo de variação relativamente à variação do escore (de 0 a 10), pode-se estabelecer uma versão normalizada de cada métrica para esse intervalo, ou seja, para o MAPE que terá um escore de 10 quando for nulo e 0 quando for 25%, sua versão normalizada z_1 será definida conforme a equação 46.

$$z_1 = 10 - 10\left(\frac{\text{MAPE}}{25}\right) \quad (46)$$

Já o coeficiente de desigualdade de Theil, ou U-Theil, terá um escore de 10 quando for nulo e 0 quando for 0,5, logo sua versão normalizada z_2 será definida conforme a equação 47.

$$z_2 = 10 - 10\left(\frac{\text{U-Theil}}{0,5}\right) \quad (47)$$

Finalmente, a TEN, que terá um escore de 10 quando for 100% e 0 quando for 100% negativo, terá sua versão normalizada z_3 definida segundo a equação 48.

$$z_3 = 10\left(\frac{100+\text{TEN}}{200}\right) \quad (48)$$

Fazendo então a média ponderada das versões normalizadas de cada métrica conforme a equação 49.

$$EDP = \frac{z_1 + z_2 + 2z_3}{4} \quad (49)$$

Substituindo-se as equações 46, 47 e 48 na equação 49, e após alguma álgebra, estabeleceu-se a seguinte expressão para a determinação do EDP.

$$EDP = 0,025.TEN - 0,1.MAPE - 5.U_{Theil} + 7,5 \quad (50)$$

Para realização das comparações de desempenho, além do EDP, também foram utilizados testes de hipótese (t - duas amostras em par para médias e ANOVA fator único), como poderá ser verificado no capítulo de análise dos resultados.

8. ANÁLISE DE RESULTADOS

Neste estudo, além de se propor o processo de reversão neural para as transformações Box-Cox, também se realizou a avaliação e comparação dos modelos utilizados, bem como o impacto, em termos de desempenho, das transformações de dados propostas em cada modelo neural. Avaliou-se também o desempenho das previsões por tipo de preço. O desempenho das previsões realizadas pelo modelo estatístico tradicional ARMA-GARCH foram tomadas como parâmetros de referência mínima de desempenho, já considerando os resultados de Coelho et al. (2008), que indicam sua inferioridade relativamente a modelos neurais.

As previsões realizadas utilizaram as séries de cinco modalidades de preços — abertura, fechamento, máximo, mínimo e médio para cada pregão — para cada ação selecionada para a carteira, como séries independentes e, portanto, comparáveis entre si; para garantir que não haveria qualquer viés nos resultados provocados por peculiaridades relacionadas a cada modalidade de preço, realizou-se uma comparação entre todas as previsões realizadas para cada modalidade de preço e para cada métrica utilizada neste trabalho, para tanto utilizou-se um teste de análise de variância de fator único (ANOVA). Como se pode verificar pela Tabela 09 os valores p obtidos indicam que, ao nível de 5% de significância, não há diferença significativa entre as previsões realizadas para cada modalidade de preços.

Tabela 9 - Desempenho de previsões para as modalidades de preços - Anova: fator único $\alpha = 5\%$

Métrica	MAPE	U-Theil	TEN	EDP
Valor p	0,5744	0,6507	0,2478	0,3275

Fonte: o autor

8.1. Análise estatística das séries

Os primeiros fatos estatísticos relevantes a serem ressaltados quanto às séries utilizadas neste trabalho são sua heterocedasticidade ou não estacionariedade e suas distribuições não apresentarem características de gaussianidade. Como pode ser

verificado na Tabela 10, o teste de estacionariedade Dickey-Fuller aumentado denotou com 0 todas as séries originais, indicando a falha em rejeitar a hipótese de presença de raízes unitárias, o que indica a não estacionariedade; já o teste Jarque-Bera denotou com 1 todas as séries originais indicando a falha em rejeitar a hipótese nula de normalidade.

Quanto às transformações, diferenças de primeira ordem, retornos logarítmicos e Box-Cox, todas falham em tornar a distribuição das séries normais, o que, de fato, era objetivo apenas da transformação Box-Cox. Cabe notar, no entanto, que a transformação Box-Cox logra êxito em reduzir a assimetria da distribuição das séries, que são todas, forte e negativamente assimétricas, como pode ser verificado na Tabela 10 e nos histogramas das séries pré e pós transformação, apresentados nas Figuras 13 a 15, gráficos b e c. Outro ponto que se percebe em relação à transformação Box-Cox é que ela não aumenta tanto a variabilidade da série como as demais transformações, o que também pode ser verificado pelo coeficiente de variação na Tabela 10.

Tabela 10 - Estatística descritiva das séries pré e pós transformação

Ação	Transformação	jb	jb1	df	df1	ass	ass1	CV	CV1	med	med1
BBAS3ab	Diferenças	1	1	0	1	-0,771	0,094	18,754	5328,640	25,923	0,009
BBAS3max	Diferenças	1	1	0	1	-0,772	0,322	18,512	4850,507	26,282	0,009
BBAS3min	Diferenças	1	1	0	1	-0,756	-0,267	18,958	5284,253	25,561	0,009
BBAS3med	Diferenças	1	1	0	1	-0,760	-0,055	18,738	4818,199	25,928	0,009
BBAS3fe	Diferenças	1	1	0	1	-0,763	0,118	18,776	5438,858	25,944	0,009
BBDC3ab	Diferenças	1	1	0	1	-0,641	-0,232	10,151	5079,133	25,067	0,008
BBDC3max	Diferenças	1	1	0	1	-0,609	-0,528	9,998	5547,621	25,347	0,007
BBDC3min	Diferenças	1	1	0	1	-0,668	-0,734	10,335	4942,168	24,724	0,008
BBDC3med	Diferenças	1	1	0	1	-0,643	-0,698	10,173	5323,383	25,029	0,007
BBDC3fe	Diferenças	1	1	0	1	-0,640	-0,741	10,141	6599,912	25,106	0,007
BBDC4ab	Diferenças	1	1	0	1	-0,757	-0,181	11,182	4482,393	30,641	0,012
BBDC4max	Diferenças	1	1	0	1	-0,723	-0,551	10,962	4790,351	31,017	0,010
BBDC4min	Diferenças	1	1	0	1	-0,781	-0,687	11,378	4388,403	30,249	0,012
BBDC4med	Diferenças	1	1	0	1	-0,758	-0,881	11,176	4686,078	30,637	0,011
BBDC4fe	Diferenças	1	1	0	1	-0,764	-0,551	11,170	5705,369	30,654	0,010
BBAS3ab	Retornos	1	1	0	1	-0,771	0,199	18,754	4080,059	25,923	0,0005
BBAS3max	Retornos	1	1	0	1	-0,772	0,310	18,512	3723,173	26,282	0,0005
BBAS3min	Retornos	1	1	0	1	-0,756	-0,224	18,958	4040,794	25,561	0,0005
BBAS3med	Retornos	1	1	0	1	-0,760	-0,034	18,738	3697,512	25,928	0,0005
BBAS3fe	Retornos	1	1	0	1	-0,763	0,145	18,776	4247,484	25,944	0,0005

BBDC3ab	Retornos	1	1	0	1	-0,641	-0,214	10,151	4899,678	25,067	0,0004
BBDC3max	Retornos	1	1	0	1	-0,609	-0,346	9,998	5383,645	25,347	0,0003
BBDC3min	Retornos	1	1	0	1	-0,668	-0,544	10,335	4753,276	24,724	0,0004
BBDC3med	Retornos	1	1	0	1	-0,643	-0,606	10,173	5128,047	25,029	0,0003
BBDC3fe	Retornos	1	1	0	1	-0,640	-0,653	10,141	6476,754	25,106	0,0003
BBDC4ab	Retornos	1	1	0	1	-0,757	-0,035	11,182	4283,846	30,641	0,0004
BBDC4max	Retornos	1	1	0	1	-0,723	-0,343	10,962	4649,245	31,017	0,0003
BBDC4min	Retornos	1	1	0	1	-0,781	-0,477	11,378	4189,324	30,249	0,0004
BBDC4med	Retornos	1	1	0	1	-0,758	-0,685	11,176	4521,444	30,637	0,0004
BBDC4fe	Retornos	1	1	0	1	-0,764	-0,348	11,170	5634,226	30,654	0,0003
BBAS3ab	Box-Cox	1	1	0	0	-0,771	-0,155	18,754	40,807	25,923	1553,681
BBAS3max	Box-Cox	1	1	0	0	-0,772	-0,150	18,512	40,550	26,282	1670,366
BBAS3min	Box-Cox	1	1	0	0	-0,756	-0,155	18,958	40,702	25,561	1326,943
BBAS3med	Box-Cox	1	1	0	0	-0,760	-0,151	18,738	40,517	25,928	1441,983
BBAS3fe	Box-Cox	1	1	0	0	-0,763	-0,149	18,776	40,528	25,944	1428,604
BBDC3ab	Box-Cox	1	1	0	0	-0,641	0,056	10,151	24,971	25,067	1725,543
BBDC3max	Box-Cox	1	1	0	0	-0,609	0,059	9,998	24,034	25,347	1438,585
BBDC3min	Box-Cox	1	1	0	0	-0,668	0,051	10,335	25,904	24,724	2001,175
BBDC3med	Box-Cox	1	1	0	0	-0,643	0,055	10,173	25,077	25,029	1753,767
BBDC3fe	Box-Cox	1	1	0	0	-0,640	0,057	10,141	24,889	25,106	1698,508
BBDC4ab	Box-Cox	1	1	0	0	-0,757	0,053	11,182	28,000	30,641	3984,656
BBDC4max	Box-Cox	1	1	0	0	-0,723	0,058	10,962	26,978	31,017	3360,791
BBDC4min	Box-Cox	1	1	0	0	-0,781	0,047	11,378	28,826	30,249	4460,211
BBDC4med	Box-Cox	1	1	0	0	-0,758	0,053	11,176	28,003	30,637	4008,014
BBDC4fe	Box-Cox	1	1	0	0	-0,764	0,053	11,170	28,052	30,654	4113,903

jb e jb1: testes de normalidade Jarque-Bera antes e após a transformação; df e df1: teste Dickey-Fuller aumentado de estacionariedade antes e após a transformação; ass e ass1: coeficiente de assimetria antes e após a transformação; CV e CV1: coeficiente de variação antes e após a transformação; med e med1: média antes e após a transformação.

Fonte: o autor

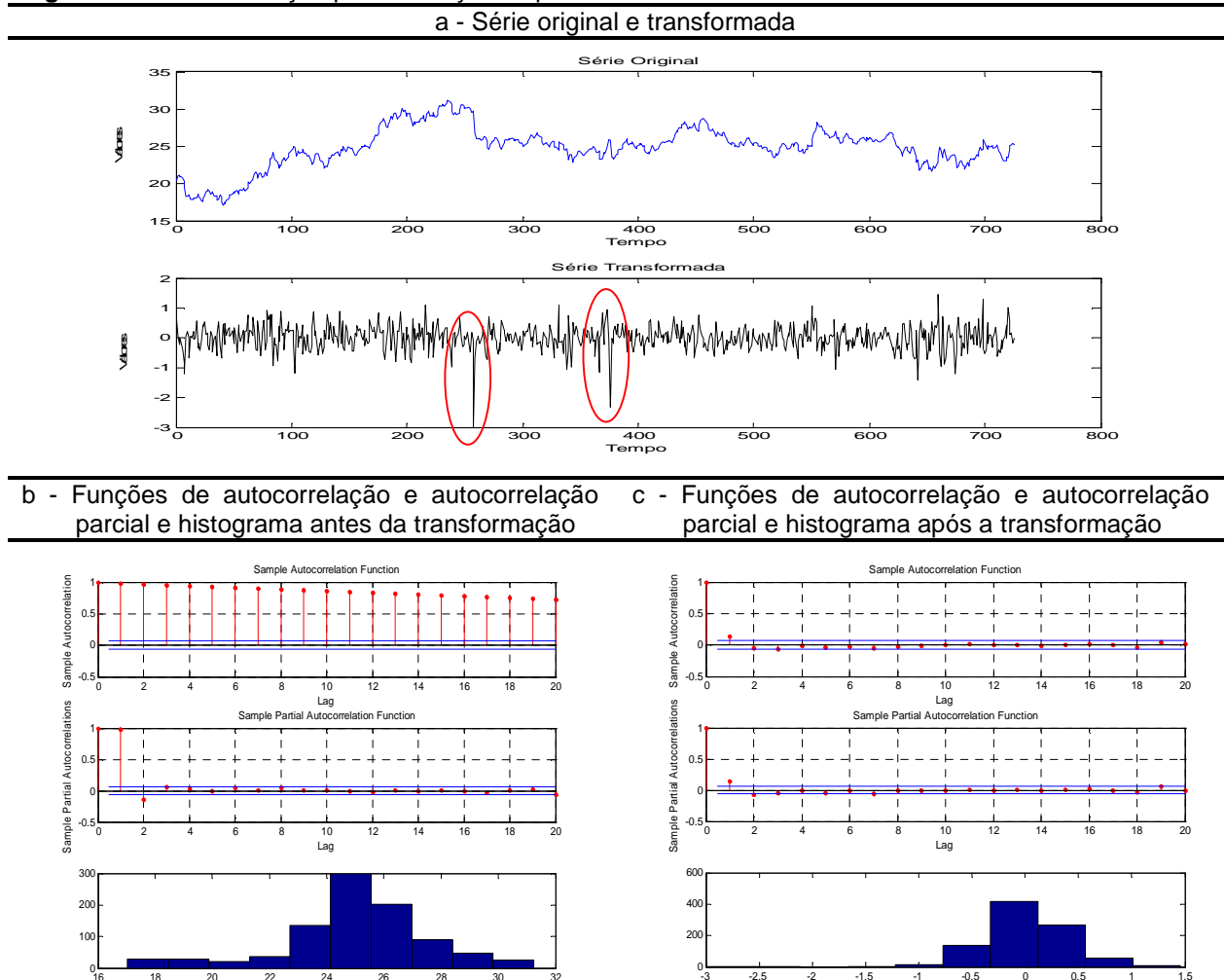
As questões acima também podem ser ilustradas pela observação das Figuras 14, 15 e 16 que trazem os gráficos da série BBDC3med, relativamente a cada uma das transformações.

Verifica-se também que as transformações por retornos e por diferenças logram êxito em tornar a série estacionária, como pode ser verificado, tanto pelo teste Dickey-Fuller aumentado, que alterou sua denotação de 0 para 1 para essas transformações, quanto pelos gráficos b e c das Figuras 14 e 15 que apresentam a modificação das funções de autocorrelação e autocorrelação parcial, que passam a convergir para zero, indicando a presença de estacionariedade.

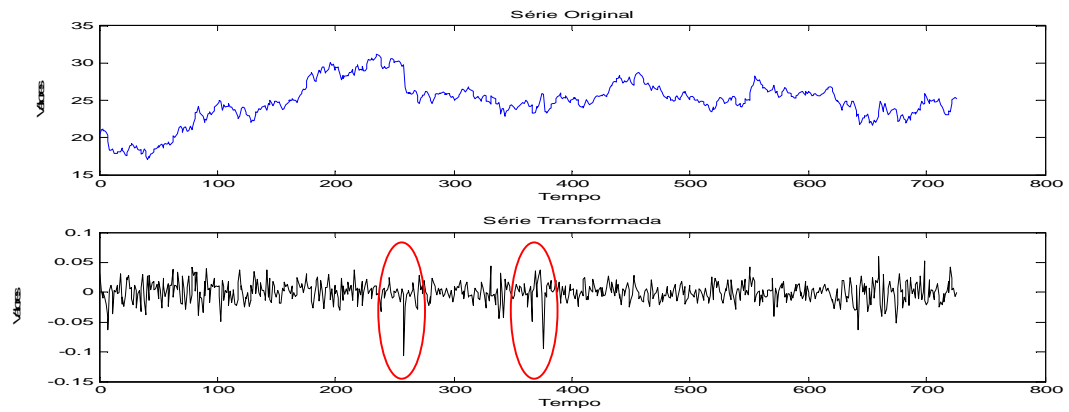
A aparente incoerência entre a presença de estacionariedade e o aumento da variabilidade das séries, quando transformadas por retornos ou diferenças, pode ser explicada pela presença de um dos fatos estilizados, extensivamente estudados pela literatura (DANIELSSON, 2011), relativos a retornos financeiros, sejam eles simples ou logarítmicos, ou seja, agrupamento de volatilidades, que podem ser observados nas Figuras 14 e 15, nos gráficos "a", pelas elipses em vermelho que sinalizam esse tipo de ocorrência.

Os efeitos de cada transformação no desempenho de previsão dos modelos neurais serão analisados adiante.

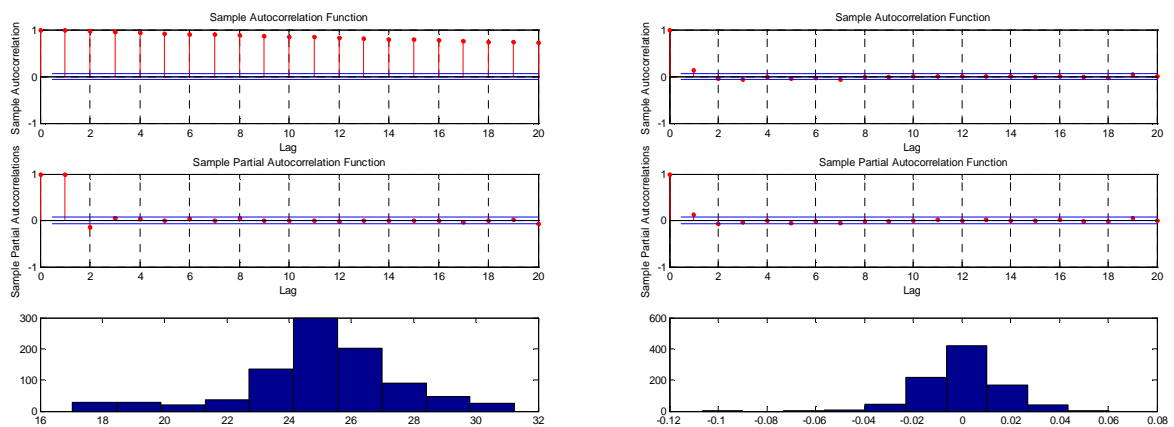
Figura 14 - Transformação por diferenças de primeira ordem da Série BBDC3med



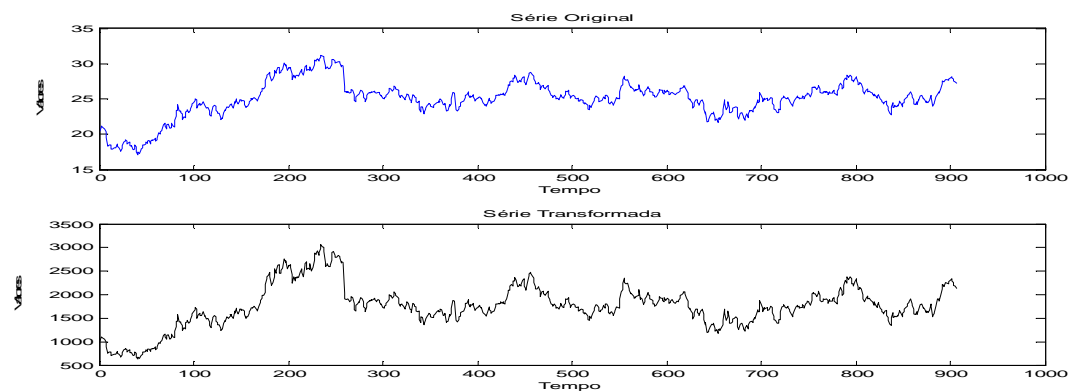
Fonte: o autor

Figura 15 - Transformação por retornos logarítmicos da Série BBDC3med**a - Série original e transformada**

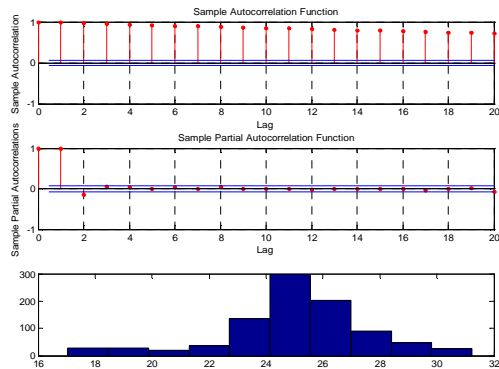
b - Funções de autocorrelação e autocorrelação parcial e histograma antes da transformação **c - Funções de autocorrelação e autocorrelação parcial e histograma após a transformação**



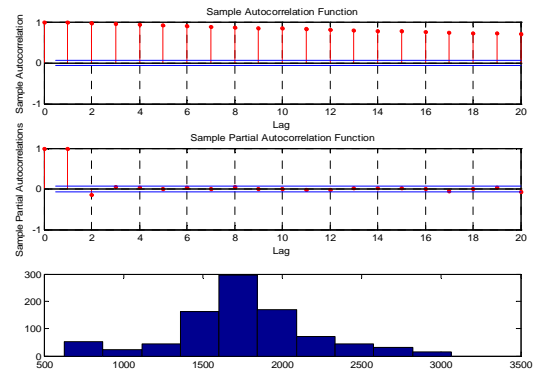
Fonte: o autor

Figura 16 - Transformação Box-Cox da Série BBDC3med**a - Série original e transformada**

b - Funções de autocorrelação e autocorrelação parcial e histograma antes da transformação



c - Funções de autocorrelação e autocorrelação parcial e histograma após a transformação



Fonte: o autor

8.2 Resultados por modelo para séries não transformadas

Para essa avaliação, as séries de preços das ações foram submetidas às redes neurais sem qualquer transformação de dados; no caso do modelo ARMA-GARCH, tanto para que houvesse um desempenho mínimo para referência, quanto para alcançar as condições mínimas de estacionariedade demandadas pelo modelo, as séries foram apresentadas na forma de retornos logarítmicos e posteriormente revertidas algebricamente.

A Tabela 11 apresenta as métricas obtidas pelas redes NARX e FTDNN, bem como a referência do modelo estatístico ARMA-GARCH.

Tabela 11 - Desempenho de previsão 10 passos a frente para séries não transformadas

Ação	NARX				FTDNN				ARMA-GARCH			
	MAPE	U-Theil	TEN	EDP	MAPE	U-Theil	TEN	EDP	MAPE	U-Theil	TEN	EDP
BBAS3ab	1,023	0,018	90,541	9,570	3,403	0,061	-72,135	5,052	3,065	0,056	-93,416	4,576
BBAS3max	3,036	0,050	65,306	8,581	4,324	0,083	-75,744	4,759	3,271	0,060	-94,251	4,517
BBAS3min	2,198	0,041	100,000	9,573	4,081	0,066	68,421	8,472	4,207	0,072	-90,182	4,467
BBAS3med	1,923	0,036	17,485	7,563	2,215	0,040	73,885	8,926	3,894	0,068	-80,271	4,764
BBAS3fe	3,063	0,055	54,922	8,292	3,704	0,067	9,377	7,030	4,863	0,082	-48,346	5,394
BBDC3ab	1,418	0,026	73,453	9,064	1,626	0,033	3,366	7,255	2,468	0,049	-57,102	5,579
BBDC3max	0,879	0,017	96,939	9,752	1,673	0,034	73,826	9,010	1,625	0,034	-41,222	6,139
BBDC3min	1,617	0,029	23,895	7,792	2,521	0,051	7,148	7,172	2,378	0,048	-49,124	5,795
BBDC3med	1,214	0,025	43,303	8,337	3,072	0,056	30,947	7,687	2,270	0,045	-32,697	6,231

BBDC3fe	1,531	0,027	41,794	8,254	2,210	0,038	41,794	8,132	2,660	0,049	-19,816	6,493
BBDC4ab	1,499	0,028	83,832	9,308	2,360	0,043	83,832	9,144	1,712	0,037	-40,513	6,132
BBDC4max	1,526	0,027	30,530	7,975	1,568	0,029	25,155	7,826	1,842	0,038	-35,922	6,227
BBDC4min	2,078	0,040	81,507	9,132	1,218	0,022	96,575	9,680	1,984	0,040	-47,063	5,924
BBDC4med	1,202	0,022	71,160	9,051	1,573	0,029	33,313	8,029	2,018	0,040	-31,384	6,312
BBDC4fe	1,757	0,031	51,094	8,445	1,534	0,030	28,165	7,898	2,649	0,051	-17,804	6,533

Sufixos acrescentados aos códigos das ações designam os tipos de preços, ab: preço de abertura; max: preço máximo; min: preço mínimo; med: preço médio; fe: preço de fechamento.

Fonte: o autor

Como se pode verificar pelas médias das métricas obtidas (Tabela 12), a rede NARX obteve o melhor desempenho em todas elas, sobretudo no que diz respeito à TEN, obtendo, em média, 61,7% da rentabilidade máxima possível, contra apenas 28,5% da rede FTDNN, o que se evidenciou nos EDPs médios de 8,7 e 7,7 respectivamente, contra a referência mínima de 5,7 do modelo GARCH.

Tabela 12 - Médias das métricas por modelo

Modelos	MAPE	U-Theil	TEN	EDP
NARX	1,731	0,031	61,717	8,713
FTDNN	2,472	0,046	28,528	7,738
GARCH	2,727	0,051	-51,941	5,672

Fonte: o autor

Como as médias do MAPE e U-Theil para a rede FTDNN e o modelo GARCH ficaram muito próximas, realizou-se também um teste de hipótese, com nível de significância de 5% (teste t para duas amostras em par), cuja hipótese nula adotada foi a igualdade entre as médias, e a hipótese alternativa a de que, entre os pares testados (ver Tabela 13), a média das métricas referentes ao primeiro modelo seriam superiores à do segundo; como se pode observar pelos valores p obtidos, a hipótese nula só não foi rejeitada no caso do MAPE e do U-Theil entre a rede FTDNN e o modelo GARCH, porém, a hipótese nula foi rejeitada para esse par nas métricas TEN e EDP, indicando claramente que a rede neural FTDNN, apesar de ter obtido desempenho semelhante ao modelo estatístico quanto à distância entre pontos previstos e reais, obteve melhor desempenho quanto à captura da tendência real dos preços. Fato que também pode ser observado em relação à rede NARX, que com 61,7% de TEN contra 28,5% da rede

FTDNN, mostra que consegue capturar a tendência dos preços com ainda mais eficiência.

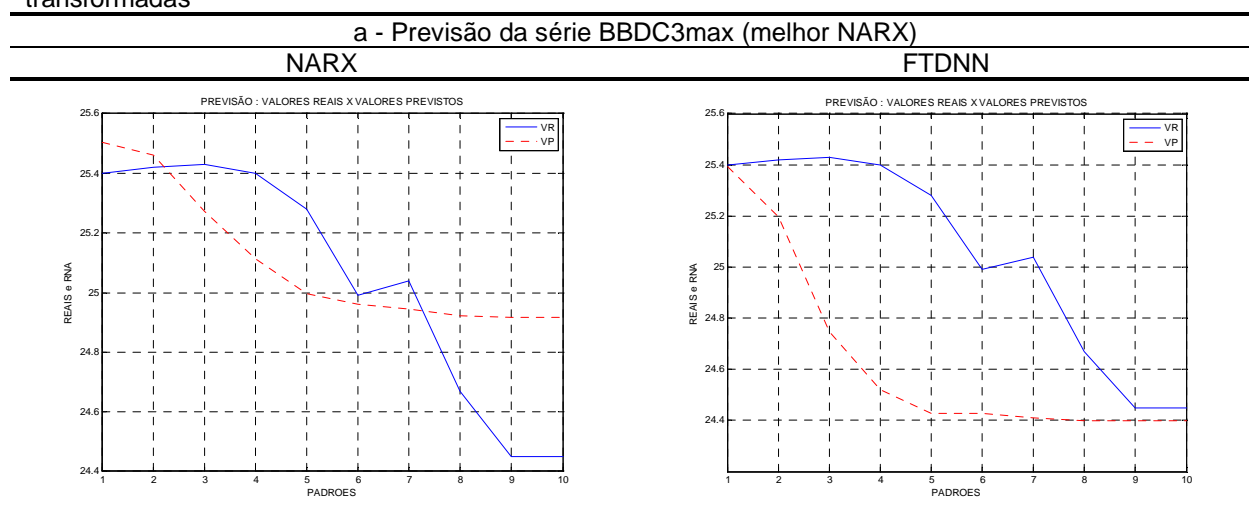
Tabela 13 - Teste-t: duas amostras em par para médias

Valor P bi-caudal ($\alpha = 5\%$)				
Modelos	MAPE	U-Theil	TEN	EDP
NARX/FTDNN	4,63E-03	2,83E-03	3,82E-02	2,52E-02
NARX/GARCH	7,05E-05	1,56E-06	5,11E-08	3,28E-08
FTDNN/GARCH	2,27E-01	1,27E-01	9,49E-06	1,20E-05

Fonte: o autor

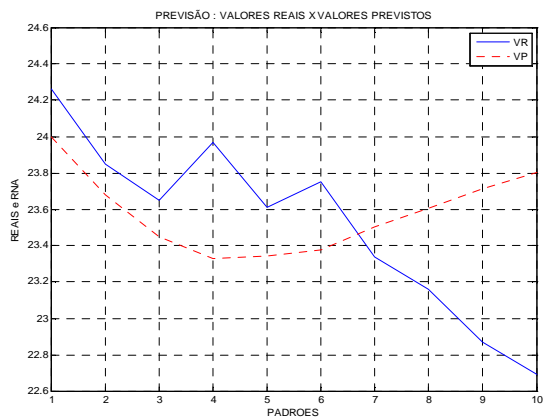
Para ilustrar as análises acima também se pode observar a Figura 17 que traz a comparação entre os gráficos obtidos com as previsões realizadas, para as séries BBDC3max (melhor performance da rede NARX), BBAS3med (pior performance da rede NARX), BBDC4min (melhor performance da rede FTDNN) e BBAS3max (pior performance da rede FTDNN), onde se pode verificar graficamente a superioridade da rede NARX no que diz respeito à assimilação da tendência dos preços.

Figura 17 - Melhores e piores performances de previsão das redes NARX e FTDNN para séries não transformadas

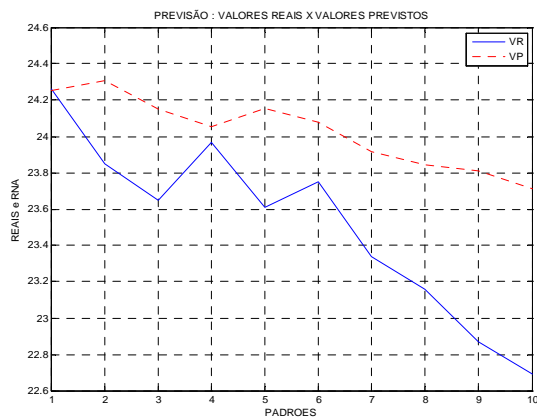


b - Previsão da série BBAS3med (Pior NARX)

NARX

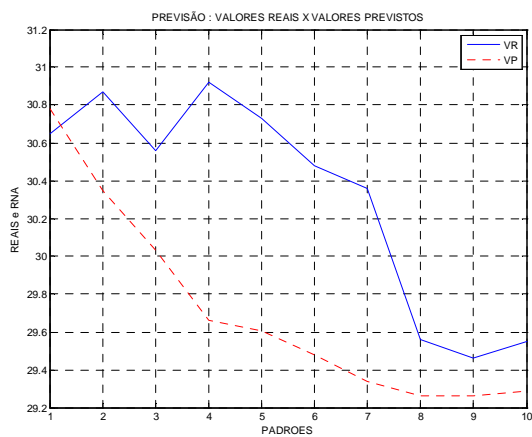


FTDNN

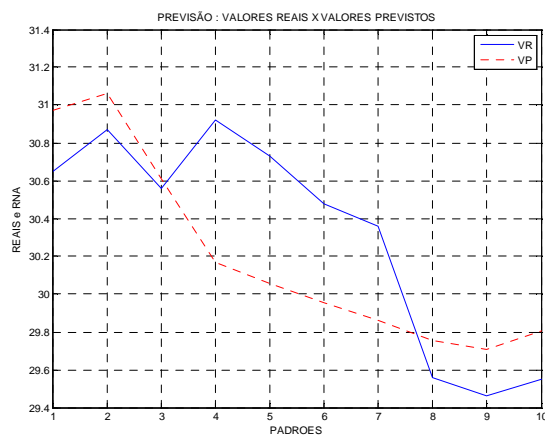


c - Previsão da série BBDC4min (Melhor FTDNN)

NARX

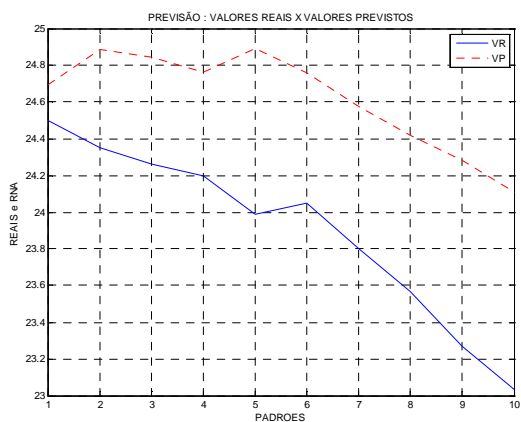


FTDNN

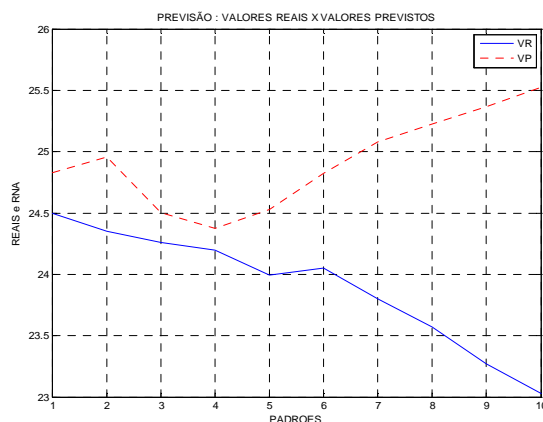


d - Previsão da série BBAS3max (Pior FTDNN)

NARX



FTDNN



VR = valores reais; VP = valores previstos
 Fonte: o autor

8.3. Comparação entre os processos de reversão em cada rede

Antes de se iniciar a análise dos desempenhos de previsão de cada modelo, faz-se necessário avaliar o desempenho dos processos de reversão das transformações aplicadas, pois os desempenhos de previsão para os modelos e transformações sempre foram mensurados pelos dados revertidos pelo melhor método de reversão para cada transformação.

Para cada transformação realizada, a reversão dos dados previstos foi executada por dois métodos: algébrico e neural, e para cada um desses métodos foram calculadas todas as métricas adotadas neste trabalho. A Tabela 14 traz as médias obtidas em cada métrica relativa a cada método de reversão, por modelo e por tipo de transformação, bem como os valores p para o teste de hipótese t, cuja hipótese nula era a igualdade das médias e a hipótese alternativa a superioridade de um dos métodos.

Pode-se observar pela Tabela 14 que, em sentido amplo, tanto para a rede NARX, quanto para a rede FTDNN as reversões algébricas têm melhor desempenho que as reversões neurais para as transformações por diferenças e retornos logarítmicos, e o contrário acontece para a transformação Box-Cox.

Cabe observar, no entanto, em sentido restrito, relativamente à transformação Box-Cox, que, para a rede FTDNN e para as métricas MAPE e U-Theil, o teste de hipótese aplicado não aponta diferenças significativas de desempenho, o que não acontece para as métricas TEN e EDP; já para o modelo NARX, acontece o oposto, o teste indica não haver diferenças significativas para a TEN e para o EDP, mas indica para o MAPE e U-Theil, embora as médias sejam melhores para a reversão neural, essa ocorrência pode, tanto ser explicada pela maior capacidade da rede NARX em captar tendências, colocando os dados previstos, mesmo que transformados algebricamente mais próximos dos dados transformados pela rede neural, ou seja com um viés menor.

Tabela 14 - Reversões - Teste t - duas amostras em par para médias $\alpha = 5\%$

Métricas	Parâmetros	NARX			FTDNN		
		Diferenças	Retornos	Box-Cox	Diferenças	Retornos	Box-Cox
MAPE	Média_ra	1,512	1,353	2,150	1,327	1,208	1,992
	Média_rn	5,175	7,675	1,900	6,088	8,083	1,857
	Valor p (bi-caudal)	0,009	0,001	0,010	0,004	0,001	0,070
U-Theil	Média_ra	0,030	0,027	0,040	0,027	0,024	0,037
	Média_rn	0,084	0,134	0,037	0,099	0,139	0,036
	Valor p (bi-caudal)	0,009	0,001	0,010	0,003	2,78E-04	0,263
TEN	Média_ra	51,515	74,971	35,775	62,766	84,146	30,296
	Média_rn	30,664	-15,718	46,381	14,559	-32,595	38,479
	Valor p (bi-caudal)	0,177	2,39E-07	0,295	0,002	1,85E-06	0,009
EDP	Média_ra	8,488	9,105	7,980	8,799	9,365	7,872
	Média_rn	7,331	5,671	8,286	6,760	5,179	8,098
	Valor p (bi-caudal)	0,023	2,52E-06	0,238	1,78E-04	3,12E-06	0,004

Os sufixos ra e rn acrescentados à palavra média, referem-se a reversão algébrica e reversão neural.

Fonte: o autor

As razões matemáticas e estatísticas pelas quais a reversão algébrica obteve melhor desempenho que a neural para as transformações por diferenças e por retornos logarítmicos, enquanto ocorreu o oposto para a transformação Box-Cox foge ao escopo deste trabalho, mas deve ser uma inquietação que motivará trabalhos futuros. Com relação especificamente à transformação Box-Cox, como já visto no item 4.4, a literatura identificou um viés introduzido na previsão, que, pelo que se verificou neste estudo, pode ser ajustado, independentemente da distribuição dos dados transformados, pela abordagem neural.

Para as demais avaliações comparativas, portanto, adotar-se-á a reversão algébrica para as transformações por diferenças e retornos, e a neural para as transformações Box-Cox.

8.4. Avaliação de desempenho de previsão entre as transformações em cada rede

Para as séries analisadas, como se pode verificar na Tabela 15, para a rede NARX, o teste de hipótese aplicado à média das métricas, cuja hipótese nula foi a inexistência de diferença significativa entre essas medidas de desempenho, somente a rejeitou para as comparações "Retornos contra Box-Cox" para as métricas MAPE e U-Theil, indicando

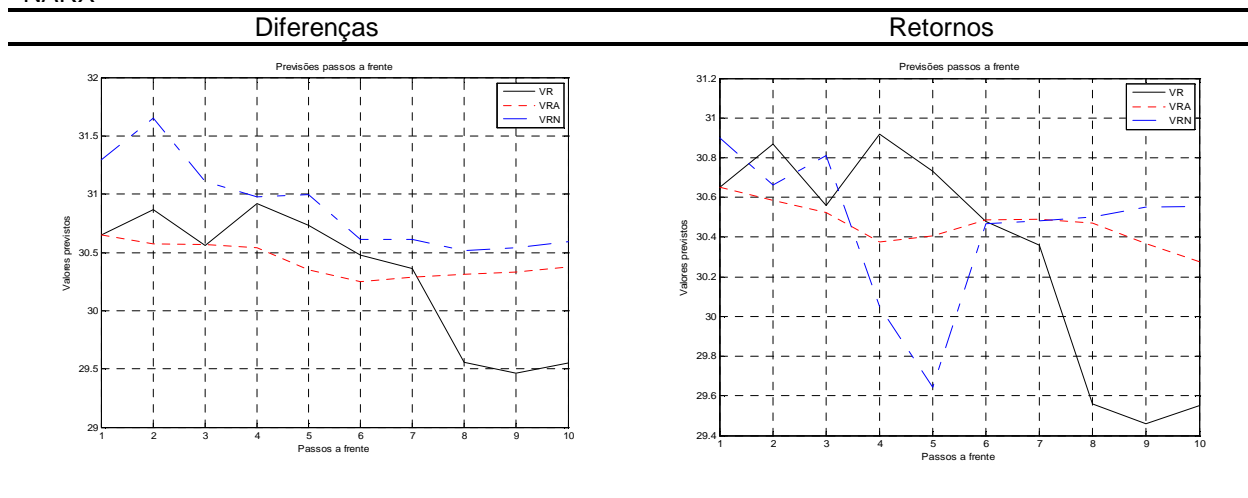
melhor desempenho dos retornos logarítmicos e “Retornos contra Originais” para as métricas MAPE, TEN e EDP, também indicando o melhor desempenho das transformações por retornos logarítmicos. Essas comparações também podem ser observadas na Figura 18, que traz os pares comparados para a série de preços mínimos da ação do Bradesco BBDC4.

Tabela 15 - Transformações entre si - NARX - Teste t - duas amostras em par para médias bi-caudal $\alpha=5\%$

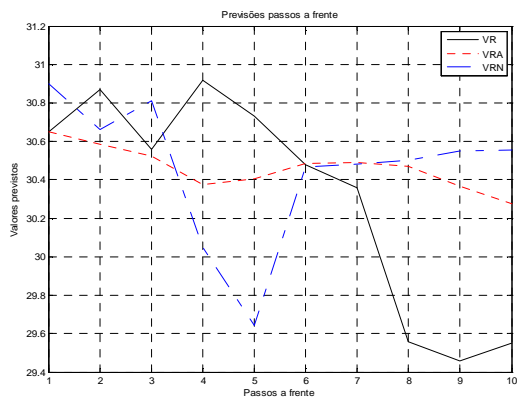
Métricas	Parâmetros	Diferenças(1)/ Retornos(2)	Retornos(1)/ Box-Cox(2)	Diferenças(1)/ Box-Cox(2)	Diferenças(1)/ Originais(2)	Retornos(1)/ Originais (2)	Box-Cox (1)/ Originais (2)
MAPE	Média 1	1,512	1,353	1,512	1,512	1,353	1,900
	Média 2	1,353	1,900	1,900	1,731	1,731	1,731
	Valor p	0,124	0,018	0,129	0,142	0,029	0,567
U-Theil	Média 1	0,030	0,027	0,030	0,030	0,027	0,037
	Média 2	0,027	0,037	0,037	0,031	0,031	0,031
	Valor p	0,138	0,015	0,121	0,428	0,119	0,293
TEN	Média 1	51,515	74,971	51,515	51,515	74,971	46,381
	Média 2	74,971	46,381	46,381	61,717	61,717	61,717
	Valor p	0,065	0,088	0,782	0,411	0,022	0,386
EDP	Média 1	8,488	9,105	8,488	8,488	9,105	8,286
	Média 2	9,105	8,286	8,286	8,713	8,713	8,713
	Valor p	0,057	0,071	0,683	0,465	0,014	0,371

Fonte: o autor

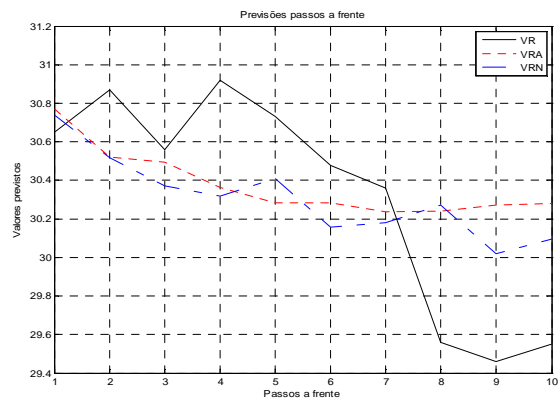
Figura 18 - Comparações Gráficas entre as transformações para a série BBDC4-preços mínimos - Rede NARX



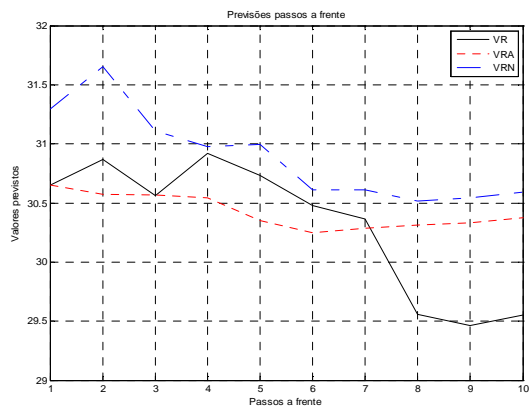
Retornos



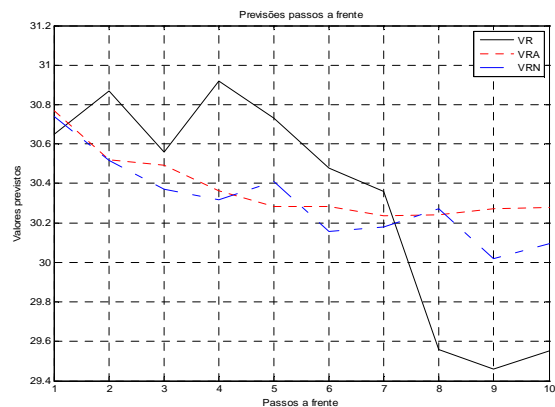
Box-Cox



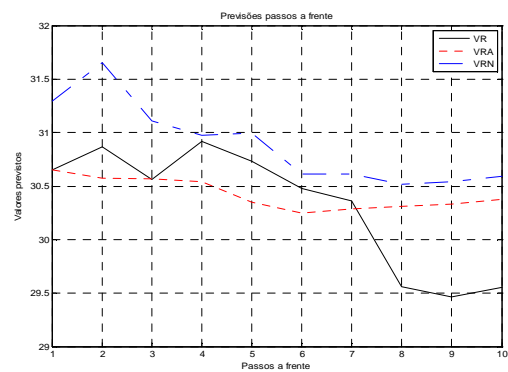
Diferenças



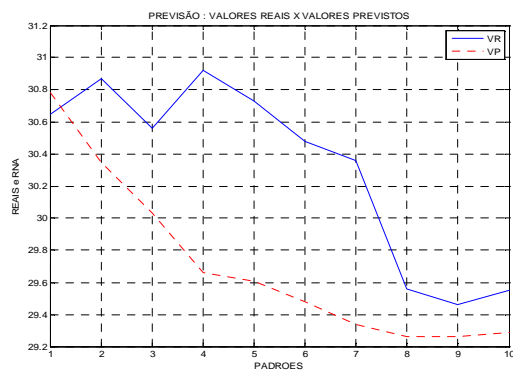
Box-Cox

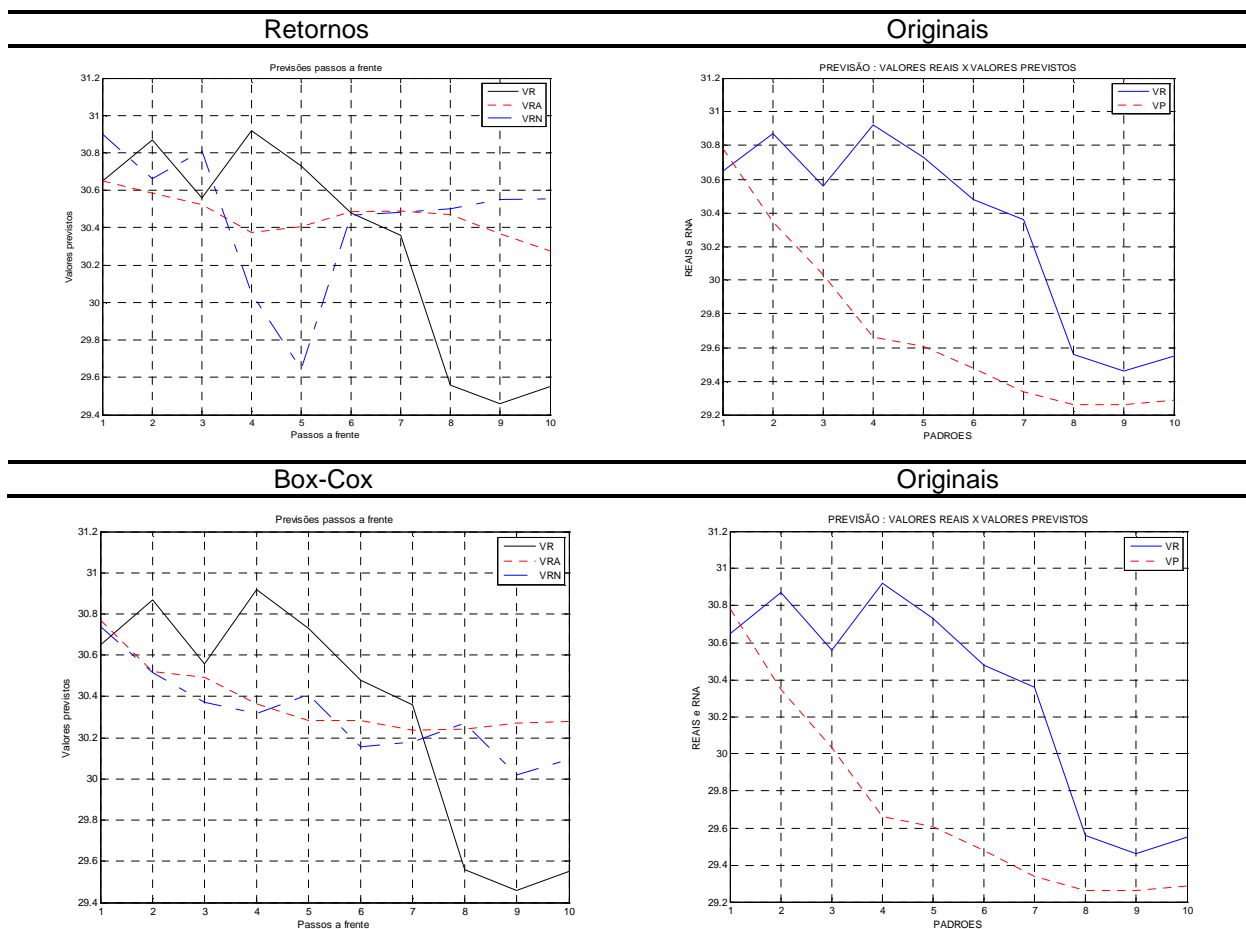


Diferenças



Originais





VR= valores reais, VRA=valores revertidos algebricamente; VRN=valores revertidos por rede neural; VP= valores previstos
 Fonte: o autor

Já para as redes FTDNN acentuam-se as disparidades de desempenho, uma vez que o teste aplicado indicou a inexistência de diferenças significativas apenas para as comparações entre as transformações por diferenças e por retornos logarítmicos e somente para as métricas MAPE e U-Theil, assim como para as comparações entre a transformação Box-Cox e Originais (dados não transformados) e somente para a métricas TEN e EDP. Em sentido amplo, no entanto, verifica-se, conforme Tabela 16, que o teste aplicado indica que os retornos logarítmicos também geraram os melhores desempenhos de previsão, seguidos pelas transformações por diferenças e, por último, a transformação Box-Cox. Ressalta-se também que todas as transformações geraram desempenho melhores que os dados originais submetidos diretamente à rede FTDNN.

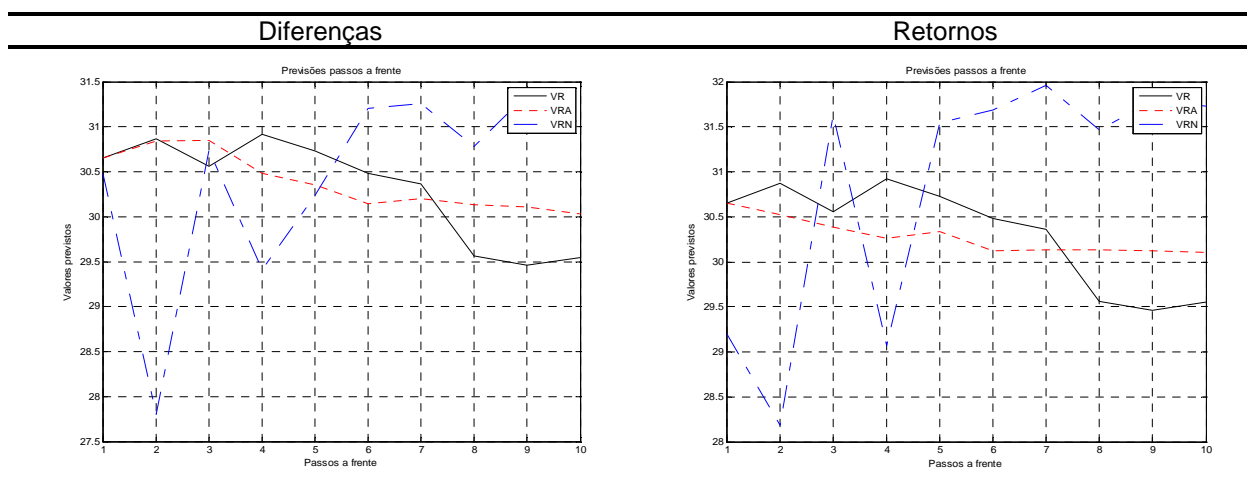
Essas comparações também podem ser observadas na Figura 19, que traz os pares comparados para a série de preços mínimos da ação do Bradesco BBDC4.

Tabela 16 - Transformações entre si - FTDNN - Teste t - duas amostras em par para médias bi-caudal ($\alpha=5\%$)

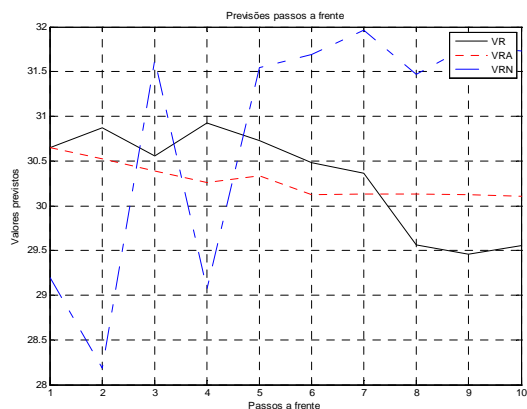
Métricas	Parâmetros	Diferenças(1)/ Retornos(2)	Retornos(1)/ Box-Cox(2)	Diferenças(1)/ Box-Cox(2)	Diferenças(1)/ Originais(2)	Retornos(1)/ Originais (2)	Box-Cox (1)/ Originais (2)
MAPE	Média 1	1,327	1,208	1,327	1,327	1,208	1,857
	Média 2	1,208	1,857	1,857	2,472	2,472	2,472
	Valor p	0,433	4,7E-04	0,003	4,6E-04	2,0E-04	0,043
U-Theil	Média 1	0,027	0,024	0,027	0,027	0,024	0,036
	Média 2	0,024	0,036	0,036	0,046	0,046	0,046
	Valor p	0,231	0,001	0,020	0,001	1,9E-04	0,044
TEN	Média 1	62,766	84,146	62,766	62,766	84,146	38,479
	Média 2	84,146	38,479	38,479	28,528	28,528	28,528
	Valor p	0,035	7,1E-05	0,026	0,029	0,001	0,464
EDP	Média 1	8,799	9,365	8,799	8,799	9,365	8,098
	Média 2	9,365	8,098	8,098	7,738	7,738	7,738
	Valor p	0,041	4,4E-05	0,017	0,015	0,001	0,337

Fonte: o autor

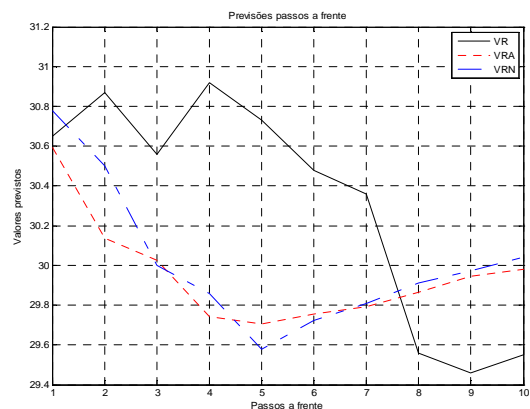
Figura 19 - Comparações Gráficas entre as transformações para a série BBDC4-preços mínimos - Rede FTDNN



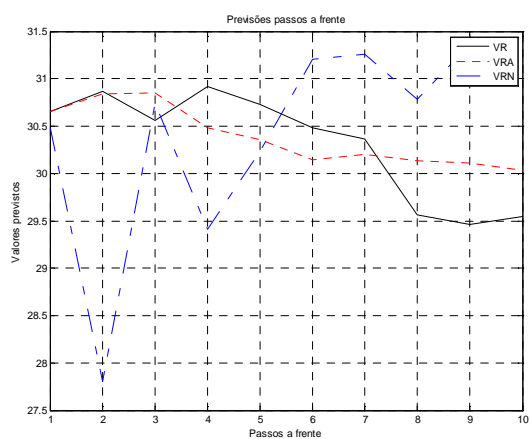
Retornos



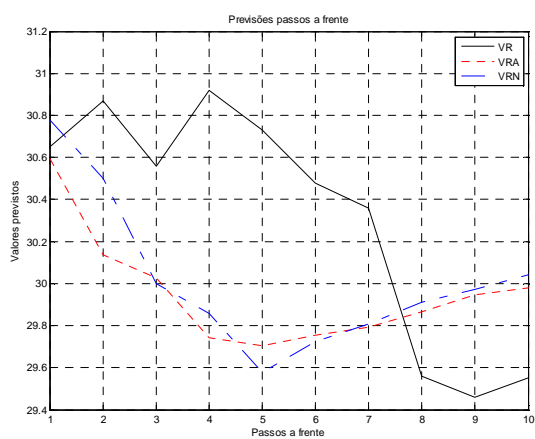
Box-Cox



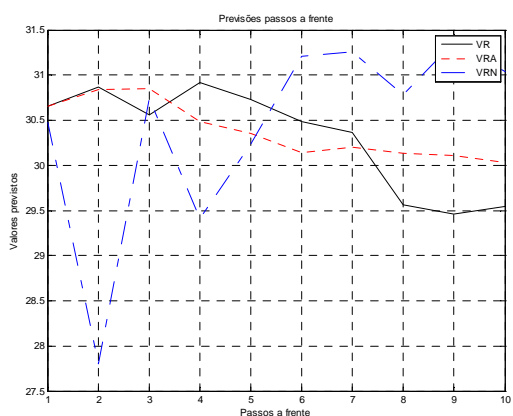
Diferenças



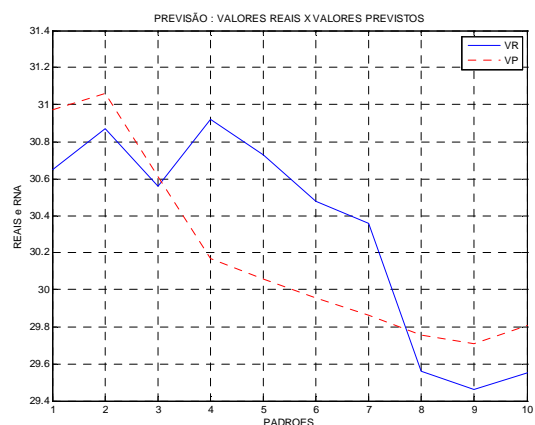
Box-Cox

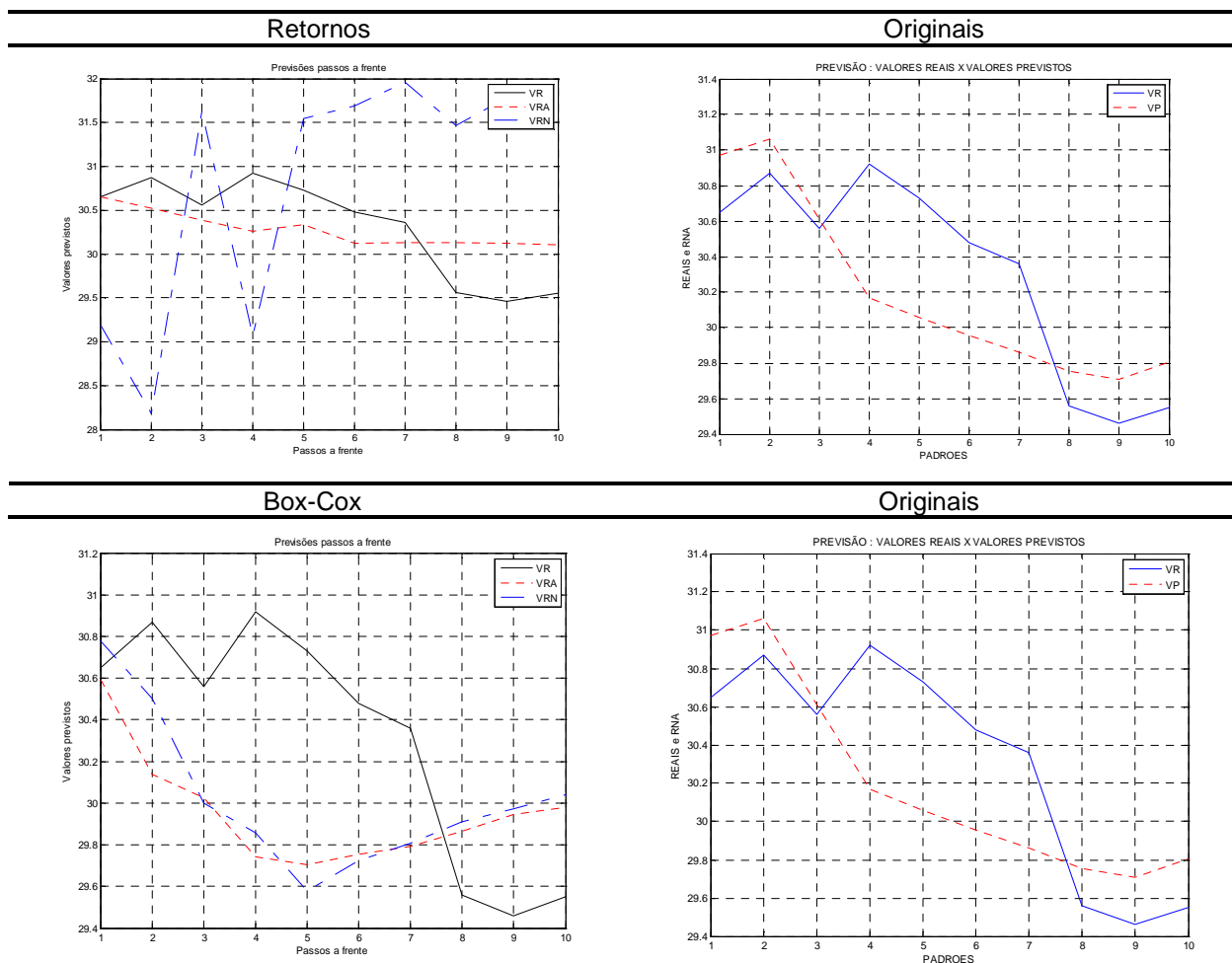


Diferenças



Originais





VR= valores reais, VRA=valores revertidos algebricamente; VRN=valores revertidos por rede neural; VP= valores previstos
 Fonte: o autor

Analisando ambos os resultados, verifica-se que as transformações, sobretudo os retornos logarítmicos, podem trazer contribuições significativas para o processo de previsão de séries temporais financeiras no contexto das redes neurais, em oposição ao que acontece com os modelos estatísticos, para os quais a literatura já afirmava que as transformações não melhoravam a qualidade da previsão (NELSON, 1976 apud MORETTIN; TOLOI, 2006; MAKRIDAKS; HIBON, 1979).

8.5. Comparação entre as redes para dados transformados

Enquanto no item 8.2, verificou-se que a Rede NARX apresenta melhor desempenho de previsão que a rede FTDNN para as séries não transformadas, para os dados transformados não foi detectada diferença significativa entre as redes, considerando-se a mesma transformação. Conforme se verifica na Tabela 17, o teste de hipótese aplicado não rejeita a hipótese nula de igualdade entre as médias das métricas para nenhuma delas e para nenhuma das transformações.

Tabela 17 - Transformações entre redes - Teste t - duas amostras em par para médias $\alpha = 5\%$

Métricas	Parâmetros	Diferenças	Retornos	Box-Cox
MAPE	Média 1 - NARX	1,512	1,353	1,900
	Média 2 - FTDNN	1,327	1,208	1,857
	Valor p (bi-caudal)	0,268	0,412	0,878
U-Theil	Média 1 - NARX	0,030	0,027	0,037
	Média 2 - FTDNN	0,027	0,024	0,036
	Valor p (bi-caudal)	0,394	0,348	0,814
TEN	Média 1 - NARX	51,515	74,971	46,381
	Média 2 - FTDNN	62,766	84,146	38,479
	Valor p (bi-caudal)	0,286	0,297	0,636
EDP	Média 1 - NARX	8,488	9,105	8,286
	Média 2 - FTDNN	8,799	9,365	8,098
	Valor p (bi-caudal)	0,255	0,258	0,678

Fonte: o autor

9. CONCLUSÕES

Esta dissertação teve como objetivo principal, desenvolver, aplicar e comparar o desempenho preditivo de séries temporais financeiras de duas arquiteturas diferenciadas de redes neurais dinâmicas, uma, não recorrente, a rede focada atrasada no tempo (FTDNN) e outra, recorrente, a rede não linear auto-regressiva com entradas exógenas (NARX). E cada uma delas aplicadas com quatro formas de pré-processamento dos dados:

- dados originais, sem transformação;
- dados transformados por diferenças de primeira ordem;
- dados transformados por retornos logarítmicos;
- dados submetidos à transformação Box-Cox.

De modo geral, como pode ser verificado mais detalhadamente no capítulo 8, a rede NARX apresenta melhor desempenho preditivo que a rede FTDNN, sobretudo, no que diz respeito à captura de tendências, ainda que, do ponto de vista somente das distâncias entre dados previstos e reais, essa diferença não seja muito significativa. Deve-se ressaltar, no entanto, que o custo computacional de processamento da rede NARX é significativamente maior que da rede FTDNN.

Verificou-se também que as transformações de dados, na fase de pré-processamento de ambas as redes, podem melhorar o nível de acurácia das previsões, tornando essa uma boa prática a ser adotada nos trabalhos dessa natureza. Ressalta-se também que as transformações que geram condições de estacionariedade nas séries temporais tem maior impacto na melhoria do desempenho preditivo que as que buscam a gaussianidade das séries, o que confirma os resultados de Zhang; Qi (2005), uma vez que as transformações por diferenças e por retornos logarítmicos eliminam tanto tendências, quanto sazonalidades.

Constatou-se também que, para ambas as arquiteturas adotadas, em sentido amplo, a transformação por retornos logarítmicos gera melhores desempenhos, seguidas pelas diferenças de primeira ordem, e, por último, a transformação Box-Cox.

Quanto ao processo de reversão dos dados previstos para a escala da série original, o método neural proposto, como um dos objetivos deste trabalho, logrou êxito no que diz respeito à transformação Box-Cox, já para as transformações por diferenças de primeira ordem e retornos, a reversão algébrica obteve melhores resultados.

Conforme os objetivos estipulados para este trabalho, também se desenvolveu uma métrica para avaliação do desempenho preditivo das redes — a TEN (taxa de eficiência negocial) — baseada em uma estratégia negocial mais elaborada e realista que estratégias ingênuas do tipo *buy-and-hold* (COELHO et al., 2008), comumente utilizadas na literatura. Cabe ressaltar que a métrica desenvolvida tem seu foco avaliativo na captura da tendência dos preços no período de previsão.

9.1. Trabalhos futuros

Como este trabalho teve um caráter empírico, sugere-se para trabalhos futuros, um estudo mais aprofundado sobre as razões matemáticas de seus achados. Cabem também, trabalhos que analisem o desempenho preditivo de modelos neurais diversos diante de outras formas de pré-processamento dos dados, inclusive, focando a combinação de transformações com diferentes objetivos.

Sugerem-se também trabalhos sobre processos de reversão de dados, uma área bastante carente de literatura, sobretudo no âmbito das previsões de séries temporais com redes neurais, bem como, novos desenhos de algoritmos de treinamento de redes neurais de reversão que possam trazer bons resultados para transformações como diferenças e retornos, uma vez que, neste trabalho, o processo de treinamento não levou em consideração a questão das dependências, isto é, na transformação Box-Cox, o dado transformado depende apenas do dado real e da função de transformação, ou seja, $y_t = f(x_t)$; já nas transformações por diferenças e retornos, o dado transformado depende da função e de dois dados reais, ou seja, $y_t = f(x_t, x_{t-1})$.

10. REFERÊNCIAS

ABE, Marcos. Manual de Análise Técnica: Essência e Estratégias Avançadas. São Paulo: Novatec, 2009.

ADEODATO, Paulo J L et al. MLP ensembles improve long term prediction accuracy over single networks. **International Journal Of Forecasting**, [s.l.], v. 27, n. 3, p.661-671, 6 Aug. 2009.

ANDRADE, Luciano Carli Moreira de. Abordagem neurofuzzy para previsão de demanda de energia elétrica no curtíssimo prazo. 2010. 109 f. Dissertação (Mestrado) - Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 2010.

ARAÚJO, Fernando Henrique Pimentel. Trading system aplicado à Bovespa utilizando redes neurais e computação evolutiva. 2010. 124 f. Dissertação (Mestrado) - Curso de Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica, São José Dos Campos, 2010.

ARAÚJO, Ricardo de A.. Translation Invariant Morphological Time-lag Added Evolutionary Forecasting method for stock market prediction. **Expert Systems with Applications**, [s.l.], v. 38, n. , p.2835-2848, 2011

BICKEL, P. J.; DOKSUM, K. A.. An analysis of transformations revisited: **Journal Of The American Statistical Association**, [S.L.], v. 76, n. 374, p.296-311, jun. 1981.

BM&FBOVESPA. Ibovespa. Disponível em: <<http://www.bmfbovespa.com.br>>. Acesso em: 10 nov. 2011 e 08 ago. 2012.

BOX, G. E. P.; COX, D. R.. An analysis of transformations. **Journal Of The Royal Statistical Society. Series B (methodological)**, [S.L.], v. 26, n. 2, p.211-252, 01 jan. 1964.

BRAGA, Pedro Constant de Almeida Dos Santos. Previsão do IBOVESPA utilizando modelos Híbridos. 2006. 212 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006."

BRANCO, Sacha Tadeu; SAMPAIO, Raimundo José Borges de. Aplicação De Redes Neurais Artificiais Em Modelos De Previsão De Demanda Para Equipamentos De Infraestrutura De Telecomunicações. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 28., 2008, Rio de Janeiro. **Anais...** . [s.l.]: Associação Brasileira de Engenharia de Produção, 2008. p. 01 - 14.

CARROLL, Raymond J.. A Robust Method for Testing Transformations to Achieve Approximate Normality. **Journal Of The Royal Statistical Society. Series B** (methodological), [s.l.], v. 42, n. 1, p.71-78, 1980.

CARROLL, Raymond J.. Tests for Regression Parameters in Power Transformation Models. **Scandinavian Journal Of Statistics**, [s.l.], v. 9, n. 4, p.217-222, 1982.

CARROLL, Raymond J.; RUPPERT, David. Diagnostics and Robust Estimation When Transforming the Regression Model and the Response. **Technometrics**, [s.l.], v. 29, n. 3, p.287-299, 1987.

CASTILLO, Enrique et al. A Very Fast Learning Method for Neural Networks Based on Sensitivity Analysis. **Journal Of Machine Learning Research**, [s.l.], v. 7, n. , p.1159-1182, July 2006.

CASTRO, Maria Cristina Felippetto de. Predição não-linear de séries temporais usando redes neurais RBF por decomposição de componentes principais. 2001. 193 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas, 2001.

CAVALHEIRO, Everton Anger et al. Previsibilidade de Mercados: Um Estudo Comparativo Entre Bovespa e S&P500. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 17., 2010, Bauru. **Anais...** . Bauru: Simpep, 2010. p. 1 - 10.

CERVILHA, Guilherme Batista; SILVA FILHO, Antonio Carlos da; LIMA, Fabiano Guasti. O uso de Wavelets para melhorar a previsão de vazão de rios com Redes Neurais. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 15., 2008, Bauru. **Anais...** . Bauru: Simpep, 2008. p. 1 - 12.

CHAN, Eric Glenn. Forecasting the S&P500 index using time series analysis and simulation methods. 2009. 102 f. Dissertação (Mestrado) - Curso de Engineering And Management, Massachusetts Institute Of Tecnology, Cambridge, 2009.

CHEN, An-sing; LEUNG, Mark T..Regression neural network for error correction in foreign exchange forecasting and trading.**Computers & Operations Research**, [s.l.], v. 31, n. , p.1049-1068, 2004.

CHEN, De-wang; ZHANG, Jun-ping. Time Series Prediction Based on Ensemble ANFIS. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS, 4., 2005, Guangzhou. **Proceedings...** . [s.l.]: IEEE, 2005. p. 3552 - 3556.

CHEN, Jianhua. Neural network applications in agricultural economics.2005. 212 f. Tese (Doutorado) - Curso de AgriculturalEconomics, UniversityOfKentuky, Lexington, 2005.

CHEONG, Chongcheul; KIM, Young-jae; YOON, Seong-min. Can We Predict Exchange Rate Movements at Short Horizons? **Journal Of Forecasting**, [s.l.], n. , p.[S.I]-[S.I], 3 June 2011.

CHERIF, Aymen; CARDOT, Hubert; BONE, Romuald. SOM time series clustering and prediction with recurrent neural networks. **Neurocomputing**, [s.l.], v. 74, n. , p.1936-1944, 3 Mar. 2011

CICHINI, Faabio Augusto Leandrin; CASTRO, Rosani de; FLAUZINO, Rogério Andrade. Uma abordagem baseada em sistemas de inferência fuzzy para estimação de cotação de ações. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 27., 2007, Foz do Iguaçu. **Anais...** . [s.l.]: Associação Brasileira de Engenharia de Produção, 2007. p. 01 - 10.

CICHINI, Fabio Augusto Leandrin. Aplicação de um sistema de inferência fuzzy de suporte à decisão para estimação de valores de ações cotadas na bolsa de valores de São Paulo. 2009. 104 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Faculdade de Engenharia da Universidade Estadual Paulista Julio de Mesquita Filho, Bauru, 2009.

CLEMENTS, Michael P.; FRANCES, Philip Hans; SWANSON, Norman R..Forecasting economic and financial time-series with non-linear models. **International Journal of Forecasting**, [s.l.], v. 20, n. , p.169-183, 2004.

COELHO, Leandro dos Santos; SANTOS, André Alves Portela; COSTA JR., Newton Carneiro Affonso da. Podemos prever a taxa de cambio brasileira? Evidência empírica utilizando inteligência computacional e modelos econométricos. **Gestão & Produção**, São Carlos, v. 15, n. 3, dez. 2008

COELHO, Leandro Dos Santos; SILVA, Wesley Vieira da; PROTIL, Roberto Max. Previsão não-linear dos preços de troncos de eucalipto baseada em uma abordagem neuroevolutiva. **Gestão & Produção**, São Carlos, v. 14, n. 1, p.139-154, abr. 2007.

CORDEIRO, Dirac Moutinho. Séries Temporais: análise quantitativa-teoria e aplicações. Recife: Edupe, 2002.

CORREIA, Suelene de Jesus do Carmo; SANTOS, Márcio Braga Dos; SERRA, Cláudio Mauro Vieira. Estudo comparativo de estruturas de redes neurais artificiais na previsão de séries temporais da produção brasileira de gás natural e petróleo (2000-2006). In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 14., 2007, Bauru. **Anais...** . Bauru: Simpep, 2007. p. 1 - 12.

CRONE, Sven F; HIBON, Michèle; NIKOLOPOULOS, Konstantinos. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time

series prediction. **International Journal Of Forecasting**, [s.l.], v. 27, n. 3, p.635-660, 12 May 2011.

D'ANGELO, Marcos Flávio Silveira Vasconcelos. Uma Nova Formulação Fuzzy/Bayesiana para Detecção de Ponto de Mudança em Séries Temporais: Aplicações na Detecção de Falhas Incipientes. 2010. 131 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Faculdade de Engenharia Elétrica da Universidade Federal de Minas Gerais, Belo Horizonte, 2010"

DENÍELSSON, Jón. Financial Risk forecasting. 1a. ed. Chippenham (UK): Wiley, 2011

DUAN, Naihua. Smearing Estimate: A Nonparametric Retransformation Method. **Journal Of The American Statistical Association**, [S.L.], v. 78, n. 383, p.605-610, set. 1983.

EBRAHIMPOUR, Reza et al. Mixture of MLP-experts for trend forecasting of time series: A case study of the Tehran stock exchange. **International Journal Of Forecasting**, [s.l.], v. 27, n. 3, p.804-816, 18 June 2010.

FENG, Hsuan-ming; CHOU, Hsiang-chai. Evolutional RBFNs prediction systems generation in the applications of financial time series data. **Expert Systems with Applications**, [s.l.], v. 38, n. , p.8285-8292, 2011.

FIGUEIREDO, Clodoaldo José; CHAVES NETO, Anselmo. Previsão de séries temporais utilizando a metodologia Box & Jenkins e redes neurais para inicialização de planejamento e controle de produção. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 15., 2008, Bauru. **Anais...** . Bauru: Simpep, 2008. p. 1 - 10.

FILDES, Robert et al. Forecasting and operational research: a review. **Journal of The Operational Research Society**, Londres, v. 59, n. 9, p.1150-1172, 14 maio 2008.

FIRAT, Mahmut; TURAN, Mustafa Erkan; YURDUSEV, Mehmet Ali. Comparative analysis of neural network techniques for predicting water consumption time series. **Journal of Hydrology**, [s.l.], v. 384, n. , p.46-51, 2010."

FLORES, João Henrique Ferreira. Comparação de modelos MLP/RNA e modelos Box-Jenkins em séries temporais não lineares. 2009. 133 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal do Rio Grande Sul, Porto Alegre, 2009.

FLORES, João Henrique Ferreira; WERNER, Liane. Aplicação de redes neurais artificiais á previsão de vendas de máquinas agrícolas. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 27., 2007, Foz do Iguaçu. **Anais...** . [s.l.]: Associação Brasileira de Engenharia de Produção, 2007. p. 01 - 9. "

GIORGI, Maria Grazia De; FICARELLA, Antonio; TARANTINO, Marco. Error analysis of short term wind power prediction models. **Applied Energy**, [s.l.], n. , p.1298-1311, 2011.

GOMES, Daniel Takata. Redes neurais recorrentes para previsão de séries temporais de memórias curta e longa. 2005. 153 f. Dissertação (Mestrado) - Curso de Estatística, Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas, Campinas, 2005.

GÓMEZ-GIL, Pilar et al. A Neural Network Scheme for Long-Term Forecasting of Chaotic Time Series. **Neural Processing Letters**, [s.l.], v. 33, n. 3, p.215-233, 2011.

GONZALES, Silvia Ortin. Reconstrucción de ladinámica no lineal de sistemas caóticos con retraso mediante redes neuronales. 2009. 249 f. Tese (Doutorado) - Curso de Ciencias Físicas, Universidad de Cantabria, Santander, 2009.

GRANGER, C. W. J.; NEWBOLD, P.. Forecasting Transformed Series: . **Journal of The Royal Statistical Society: Series B (Methodological)**, [S.L.], v. 38, n. 2, p.189-203, 1976.

GUIMARÃES, Lourenço da Rocha. Previsão de inadimplência e redes neurais artificiais. 2006. 85 f. Dissertação (Mestrado) - Curso de Modelagem Computacional, Instituto Politécnico da Universidade do Estado do Rio de Janeiro, Nova Friburgo, 2006.

HAMILTON, James Douglas. Time series analysis. Princeton: Princeton University Press, 1994.

HAN, Aaron K.. A non-parametric analysis of transformations. **Journal of Econometrics**, [s.l.], v. 35, n. 3, p.191-209, 1987

HAYKIN, Simon. Redes Neurais: Princípios e Práticas. 2. ed. Porto Alegre: Bookman, 2001.

HERAVI, Saeed; OSBORN, Denise R.; BIRCHENHALL, C.r.. Linear versus neural network forecasts for European industrial production series. **International Journal Of Forecasting**, [s.l.], v. 20, n. , p.435-446, 2004.

HILL, Tim; O'CONNOR, Marcus; REMUS, William. Neural Network Models for Time Series Forecasts. **Management Science**, [S.L.], v. 42, n. 7, p.1082-1092, jul. 1996.

HINKLEY, D. V.; RUNGER, G.. The Analysis of Transformed Data: **Journal Of The American Statistical Association**, [S.L.], v. 79, n. 386, p.302-309, jun. 1984.

HINKLEY, David. Transformation Diagnostics for Linear Models. **Biometrika**, [s.l.], v. 72, n. 3, p.487-496, dez. 1985.

HONG, Xia. A fast identification algorithm for box-cox transformation based radial basis function neural network. **Ieee Transactions On Neural Networks**, [S.L.], v. 17, n. 4, p.1064-1-69, 1 jul. 2006.

JARQUE, Carlos M.; BERA, Anil K.. A test for normality of observations and regression residuals: **International Statistical Review**, [S.L.], v. 55, n. 2, p.163-172, 1 ago. 1987.

JIANG, Minghu et al. A fast learning algorithm for time-delay neural networks. **Information Sciences**, [s.l.], v. 148, n. 1-4, p.27-39, Dec. 2002.

KARA, Yakup; BOYACIOGLU, MelekAcar; BAYKAN, ÖmerKaan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. **Expert Systems with Applications**, [s.l.], v. 38, n. , p.5311-5319, 2011.

KATAGIRI, Hideki et al. Multiobjective Evolutionary Optimization of Training and Topology of Recurrent Neural Networks for Time-Series Prediction. **The Computer Journal Advance Access**, [s.l.], n. , p.1-12, May 2011.

KHASHEI, Mehdi; BIJARI, Mehdi. An artificial neural network (p,d,q) model for time series forecasting. **Expert Systems With Applications**, [s.l.], n. , p.479-489, 2010.

KHOSRAVI, Abbas; NAHAVANDI, Saeid; CREIGHTON, Doug. Developing Optimal Neural Network Metamodels Based on Prediction Intervals. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 9., 2009, Atlanta. **Proceedings...** . [s.i.]: Ieee, 2009. p. 1583 - 1589.

KIM, Hyun-jung; SHIN, Kyung-shik. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. **Applied Soft Computing**, [S.L.], v. 7, n. 2, p.569-576, 2007.

LAGARIAS, Jeffrey C et al. Convergence properties of the nelder-mead simplex method in low dimensions. **Siam Journal Of Optimization**, [S.L.], v. 9, n. 1, p.112-147, 2 dez. 1998.

LEONE FILHO, Marcos de Almeida. Previsão de carga de curto prazo usando ensembles de previsores selecionados e evoluídos por Algoritmos Genéticos. 2006. 143 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas, 2006.

LIMA, Fabiano Guasti. Um método de análise e previsão de sucessões cronológicas unidimensionais lineares e não-lineares. 2004. 228 f. Tese (Doutorado) - Curso de Administração, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, 2004.

LUBECKE, Thomas H. et al. Combining Foreign Exchange Rate Forecasts Using Neural Networks. **Global Finance Journal**, [s.l.], v. 9, n. 1, p.5-27, 1998.

LUNA, Ivette; BALLINI, Rosangela; SOARES, Secundino. Técnica de identificação de modelos lineares e não-lineares de séries temporais. **Sba Controle & Automação**, Campinas, v. 17, n. 3, set. 2006.

MACHADO, Birajara Soares. Determinismo e estocasticidade em séries temporais empíricas. 2003. 97 f. Dissertação (Mestrado) - Curso de Física, Instituto de Física Teórica da Universidade Estadual Paulista, São Paulo, 2003.

MAIA, Jorcenilson Pereira. Aplicação de redes neurais na predição de demanda de crédito no sistema financeiro nacional. 2008. 71 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade de Brasília, Brasília, 2008.

MAKRIDAKIS, Spyros; HIBON, Michele; MOSER, Claus. Accuracy of Forecasting: An Empirical Investigation. **Journal Of The Royal Statistical Society: Series A (General)**, [S.L.], v. 142, n. 2, p.97-145, 1979.

MARIN, Luciano Heitor Gallegos. MECTIP: Um método computacional para análise de séries temporais utilizando suavização exponencial, narmax e redes neurais. 2009. 134 f. Dissertação (Mestrado) - Curso de Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica, São José Dos Campos, 2009

MARTÍNEZ-REGO, David; FONTENLA-ROMERO, Oscar; ALONSO-BETANZOS, Amparo. Efficiency of local models ensembles for time series prediction. **Expert Systems with Applications**, [s.l.], v. 38, n. , p.6884-6894, 2011.

MAUS, A.; SPROTT, J.c..Neural network method for determining embedding dimension of a time series. **Commun Nonlinear Sci Numer Simulat**, [s.l.], v. 16, n. , p.3294-3302, 2011.

MEADE, Nigel. A comparison of the accuracy of short term foreign exchange forecasting methods. **International Journal Of Forecasting**, [s.l.], v. 18, n. , p.67-83, 2002.

MELLADO, José Luis Aznarte. Modelling time series through fuzzy ruled-based models: A statistical approach.2008. 226 f. Tese (Doutorado) - Curso de Ciencias de La Computación e Inteligencia Artificial, Universidad de Granada, Granada, 2008

MELO, Brício de. Previsão de séries temporais usando modelos de composição de especialistas locais. 2003. 194 f. Dissertação (Mestrado) - Curso de Engenharia Aeronáutica e Mecânica, Instituto Tecnológico de Aeronáutica, São José Dos Campos, 2003. "

MELO, Brício de; MILIONI, Armando Zeferino; NASCIMENTO JUNIOR, Cairo Lucio. Daily and monthly sugar price forecasting using the mixture of local expert models. **Pesquisa Operacional**, Rio de Janeiro, v. 27, n. 2, ago. 2007.

MENEZES JÚNIOR, José Maria P; A BARRETO, Guilherme. Long-term time series prediction with the NARX network: An empirical evaluation. **Neurocomputing**, [s.l.], v. 71, n. , p.3335-3343, 2008.

MILESKI JUNIOR, Albino; VIEIRA, Guilherme Ernani. Análise de métodos de previsão de demanda baseados em séries temporais em uma empresa do setor de perfumes e cosméticos. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 15., 2008, Bauru. **Anais...** . Bauru: Simpep, 2008. p. 1 - 12."

MILLER, Don M.. Reducing Transformation Bias in Curve Fitting: . **The American Statistician**, [S.L.], v. 38, n. 2, p.124-126, maio 1984.

MORETTIN, Pedro A.; TOLOI, Clélia M. C.. Análise de Séries Temporais. 2. ed. São Paulo: Edgar Blücher, 2006.

MOSHIRI, Saeed. Forecasting inflation using econometric and artificial neural network models.1997. 119 f. Tese (Doutorado) - Curso de Economics, UniversityOfManitoba, Winnipeg, 1997.

NELSON Jr., Harold L.; GRANGER, C. W. J.. Experience with using the Box-Cox transformation when forecasting economic time series: . **Journal Of Econometrics**, [S.L.], v. 10, n. 1, p.57-69, 1 mar. 1979.

NELSON, Michael; HILL, Tim; REMUS, William; O'CONNOR. Time Series Forecasting Using Neural Networks: Should the Data be Deseasonalized First? **Journal of Forecasting**, [s.l.], v.18, n.1,p.359-367, 1999

NOURANI, Vahid; KISI, Özgür; KOMASI, Mehdi. Two hybrid Artificial Intelligence approaches for modeling rainfall-runoff process. **Journal of Hydrology**, [s.l.], v. 402, p.41-59, 2011.

OLIVEIRA, Anderson Castro Soares de et al . Aplicação de redes neurais artificiais na previsão da produção de álcool. **Ciênc. agrotec.**, Lavras, v. 34, n. 2, abr. 2010.

OLIVEIRA, André Barbosa. Usando redes neurais para estimação da volatilidade: Redes neurais e modelo híbrido GARCH aumentado por redes neurais. 2010. 133 f. Dissertação (Mestrado) - Curso de Economia, Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande Sul, Porto Alegre, 2010.

OLIVEIRA, Mauri Aparecido de. Aplicação de redes neurais artificiais na análise de séries temporais econômico-financeiras. 2007. 317 f. Tese (Doutorado) - Curso de Administração, Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, 2007.

PANDA, Chakradhara; NARASIMHAN, V..Forecasting exchange rate better with artificial neural network.**Journal of Policy Modeling**, [s.l.], v. 29, n. , p.227-236, 2007.

PERICCHI, L. R.. TrustA Bayesian Approach to Transformations to Normality. **Biometrika**, [s.l.], v. 68, n. 1, p.35-43, abr. 1981.

PINHEIRO, Juliano L..Mercado de Capitais: Fundamentos e Técnicas. 5ª ed. São Paulo: Editora Atlas, 2009

RAPACH, David E.; WOCHAR, Mark E..The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. **International Journal of Forecasting**, [s.l.], v. 22, n. , p.341-361, 2006.

REIKARD, Gordon. Forecasting space weather: Can new econometric methods improve accuracy? **Advances In Space Research**, [s.l.], v. 47, n. , p.2073-2080, 2011.

REUTER, U.; MÖLLER, B..Artificial Neural Networks for Forecasting of Fuzzy Time Series.**Computer-aided Civil And Infrastructure Engineering**, [s.l.], v. 25, n. , p.363-374, 2010.

SAAD, Emad W.; PROKHOROV, Danil V.; II, Donald C. Wunsch.Comparative Study of Stock Trend Prediction Using Time Delay, Recurrent and Probabilistic Neural Networks. **Transactions On Neural Networks**, [s.l.], v. 9, n. 6, p.1456-1470, 6 Nov. 1998.

SAKIA, R. M.. Retransformation Bias: A look at the Box-Cox transformation to linear balanced mixed ANOVA models. **Metrika**, [S.L.], v. 37, n.d , p.345-351, 1990.

SAKIA, R.M.. The Box-Cox Transformation technic: a review. **Journal of The Royal Statistical Society. Series D (the Statistician)**, [S.L.], v. 41, n. 2, p.169-178, 1992.

SAMANTA, B..Prediction of chaotic time series using computational intelligence. **Expert Systems With Applications**, [s.l.], v. 38, n. , p.11406-11411, 2011.

SANDMANN, Humberto Rodrigo. Predição não-linear de séries temporais usando sistemas de arquitetura neuro-fuzzy. 2006. 73 f. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2006. "

SANTOS FILHO, Edgar Leite Dos. Previsão dos retornos do índice Bovespa usando redes neurais artificiais. 2008. 109 f. Dissertação (Mestrado) - Curso de Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, 2008.

SCHOPF, Eliseu Celestino. Método neuro-estatístico para predição de séries temporais ruidosas. 2007. 105 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Instituto de Informática da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

SEABRA, Tiago. O último grito em tecnologia. Disponível em: <<http://www.tecnologiasdeultimogrito.com/individualmente-os-neurnios-tm-capacidades-computacionais/>>. Acesso em: 15 jun. 2011.

SHEN, Wei. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm. **Knowledge-based Systems**, [s.l.], v. 24, n. , p.378-385, 2011.

SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. Redes Neurais Artificiais para engenharia e ciências aplicadas: Curso prático. São Paulo: Artliber, 2010.

SILVA, Rafael Schardosin. Previsão de venda de produtos em uma indústria de telecomunicação utilizando redes neurais artificiais. 2010. 78 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Rio Dos Sinos, São Leopoldo, 2010.

SOARES, Anderson da Silva. Predição de séries temporais econômicas por meio de redes neurais artificiais e transformada wavelet: Combinando modelo técnico e fundamentalista. 2008. 78 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos, 2008

SOARES, Fabrício; FROZZA, Rejane; PAZOS, Rubén EdgardoPanta. Predição de séries temporais financeiras utilizando wavelets e a rede neural TLFN distribuída. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 28., 2008, Rio de Janeiro. **Anais...** . [s.i.]: Associação Brasileira de Engenharia de Produção, 2008. p. 01 - 9.

SOLOMON, P. J.. TrustTransformations for Components of Variance and Covariance. **Biometrika**, [s.l.], v. 72, n. 2, p.233-239, ago. 1985.

SPAETH, Francisco. Modelo para previsão de vazão fluvial baseado em wavelets e redes neurais. 2008. 92 f. Dissertação (Mestrado) - Curso de Ciências da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2008.

STEINMETZ, Tarcísio Roberto. Extração de regras de conhecimento a partir de redes neurais artificiais aplicadas para previsão de demanda por energia elétrica. 2009. 78 f. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Rio Dos Sinos, São Leopoldo, 2009.

STEKLER, H.o.. The future of macroeconomic forecasting: Understanding the forecasting process. **International Journal of Forecasting**, [s.l.], v. 23, n. , p.237-248, 2007.

SWEETING, Trevor J.. On The Choice of Prior Distribution for the Box-Cox Transformed Linear Model. **Biometrika**, [s.l.], v. 71, n. 1, p.127-134, abr. 1984.

TAYLOR, Jeremy M. G.. Measures of Location of Skew Distributions Obtained Through Box-Cox Transformations. **Journal of The American Statistical Association**, [s.l.], v. 80, n. 390, p.427-432, jun. 1985a.

TAYLOR, Jeremy M. G.. Power Transformations to Symmetry. **Biometrika**, [s.l.], v. 72, n. 1, p.145-152, abr. 1985.

TAYLOR, Jeremy M. G.. The Retransformed Mean After a Fitted Power Transformation: **Journal of The American Statistical Association**, [S.L.], v. 81, n. 393, p.114-118, mar. 1986.

THOMAIDIS, Nikos S; DOUNIAS, George D.On detecting the optimal structure of a neural network under strong statistical features in errors. **Journal of Time Series Analysis**, [s.i.], n. , p.204-222, 10 Oct. 2010.

VELÁSQUEZ, Roberto Miguel Gutiérrez. Seleção de variáveis de entrada para previsores neurais de carga. 2006. 182 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006

VICENTE, Leonardo Cardoso; MATIAS, Ítalo de Oliceira; GOMES, Geórgia Regina Rodrigues. Utilização De Redes Neurais Artificiais Como Ferramenta De Investimento No Mercado De Ações. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 17., 2010, Bauru. **Anais...** . Bauru: Simpep, 2010. p. 1 - 12.

VILLANUEVA, Wilfredo Jaime Puma. Comitê de máquinas em predição de séries temporais. 2006. 168 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas, 2006.

WAN, Eric A.. Time Series. Disponível em: <<http://www.bme.ogi.edu/~ericwan>>. Acesso em: 10 nov. 2011.

WANG, Wen-chuan et al. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. **Journal of Hydrology**, [s.l.], v. 9, n. , p.294-306, June 2009

WUERGES, Artur Filipe Ewald; BORBA, José Alonso. Redes Neurais, Lógica Nebulosa e Algoritmos Genéticos: Aplicações E Possibilidades Em Finanças e Contabilidade. **Revista de Gestão da Tecnologia e Sistemas de Informação**, [s.l.], v. 7, n. 1, p.163-182, 2010

XAVIER, Priscila Branquinho et al. Estudos comparativos entre modelos de previsão de vazões utilizando séries temporais. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO, 16., 2009, Bauru. **Anais...** . Bauru: Simpep, 2009. p. 1 - 10.

YEN,J.; LANGARI.R. Fuzzy Logic: Intelligence, Control, and Information. New Jersey: Prentice Hall, 1999.

YILMAZ, Isik; KAYNAR, Oguz. Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. **Expert Systems With Applications**, [s.l.], v. 38, n. , p.5958-5966, 2011.

ZHANG, Peter G.; QI, Min. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, [s.l.], v.160, n., p.501-514, 2003