

Pontifícia Universidade Católica de Goiás
Programa de Mestrado em Engenharia da Produção e Sistemas

**ANÁLISE COMPARATIVA DE MÉTODOS DE PREVISÃO DE SÉRIES
TEMPORAIS ATRAVÉS DE MODELOS ESTATÍSTICOS E REDE
NEURAL ARTIFICIAL**

Ana Paula de Sousa

2012

**ANÁLISE COMPARATIVA DE MÉTODOS DE PREVISÃO DE SÉRIES
TEMPORAIS ATRAVÉS DE MODELOS ESTATÍSTICOS E REDE
NEURAL ARTIFICIAL**

Ana Paula de Sousa

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador: José Elmo de Menezes, *Dr.*

Goiânia
Março de 2012

S725a Sousa, Ana Paula de.

Análise comparativa de métodos de previsão de séries temporais através de modelos estatísticos e rede neural artificial [manuscrito] / Ana Paula de Sousa. – 2012.
101 f. : il. graf.

Bibliografia: 89-93

Dissertação (mestrado) – Pontifícia Universidade Católica de Goiás, Departamento de Engenharia da Produção e Sistemas, 2012.

Orientador: Prof. Dr. José Elmo de Menezes.

Inclui Apêndice.

1. Séries temporais. 2. Modelos estatísticos. 3. Redes neurais artificiais. 4. Rede Elman. 5. Etanol. I. Título.

CDU: 519:658.5(043.3)

004.032.26

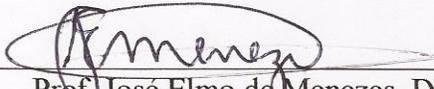
**ANÁLISE COMPARATIVA DE MÉTODOS DE PREVISÃO DE SÉRIES
TEMPORAIS ATRAVÉS DE MODELOS ESTATÍSTICOS E REDE
NEURAL ARTIFICIAL**

Ana Paula de Sousa

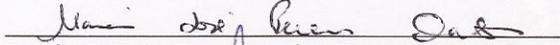
Esta Dissertação foi julgada adequada para obtenção do título de Mestre em Engenharia de Produção e Sistemas, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás em março de 2012.

Prof. Ricardo Luiz Machado, Dr. Eng.
Coordenador do Programa de Pós-Graduação em
Engenharia de Produção e Sistemas

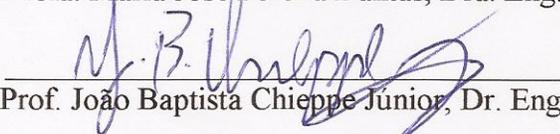
Banca Examinadora:



Prof. José Elmo de Menezes, Dr. Est.
Orientador



Prof. Maria José Pereira Dantas, Dra. Eng.



Prof. João Baptista Chieppe Junior, Dr. Eng.

Suplentes:

Prof. Iran Martins do Carmo, Dr. Est.

Prof. Ricardo Luiz Machado, Dr. Eng.

Goiânia - Goiás
Março de 2012

À minha família.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por todas as oportunidades.

Agradeço a meus pais, Paulo e Ana, meu eterno amor Emivaldo pelo apoio e paciência, minha irmã Pollyana e minha sobrinha Lídia que sempre estão presentes.

Ao meu orientador, professor Dr. José Elmo de Menezes, por todos os ensinamentos compartilhados e dedicação na orientação, sem os quais seria inviável desenvolver esse trabalho.

A professora Maria José Pereira Dantas, ao colega de PUC/GO Jeuel Alves e o professor João Baptista Chieppe Júnior pelas diversas contribuições para a melhoria deste trabalho.

Aos professores do Mestrado em Engenharia da Produção e Sistemas da PUC/GO pelos conhecimentos compartilhados nos cursos ministrados e aos colegas de curso.

Resumo da Dissertação apresentada ao MEPROS/ PUC Goiás como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia de Produção e Sistemas (M.Sc.)

ANÁLISE COMPARATIVA DE MÉTODOS DE PREVISÃO DE SÉRIES
TEMPORAIS ATRAVÉS DE MODELOS ESTATÍSTICOS E REDE NEURAL
ARTIFICIAL

Ana Paula de Sousa

Março/2012

Orientador: José Elmo de Menezes, Dr.

O objetivo deste trabalho foi comparar os métodos de estatística e de inteligência artificial para o problema da previsão de séries temporais através de Holt-Winters, Box-Jenkins e a rede neural de Elman. Os modelos foram utilizados para previsão um passo a frente dos preços do etanol no estado de Goiás e comparados através medidas de erros específicas. Ao final, os resultados indicaram que todos os métodos se mostraram competitivos em termos de predição um passo à frente, destacando-se os modelos estatísticos como os mais adequados em termos de parcimônia entre desempenho e complexidade.

Summary of Dissertation submitted to MEPROS / Goias PUC as part of the requirements for the degree of Master of Production Engineering and Systems (M.Sc.)

COMPARATIVE ANALYSIS OF TIME SERIES FORECASTING METHODS
THROUGH STATISTICAL MODELS AND ARTIFICIAL NEURAL NETWORKS

Ana Paula de Sousa

March/2012

Leader: José Elmo de Menezes, Dr.

The objective of this study was to compare statistical methods and artificial intelligence to the problem of time series forecasting using Holt-Winters, Box-Jenkins and the Elman neural network. The models were used to predict one step ahead of the price of ethanol in the state of Goias and compared using measures of specific errors. At the end, the results indicated that all three techniques were competitive in terms of predicting one step ahead especially the statistical models appeared to be the most suitable methods in terms of balance between performance and complexity.

SUMÁRIO

1 INTRODUÇÃO	1
1.1 Justificativa da Pesquisa	2
1.2 Definição do Problema de Pesquisa	10
1.3 Objetivos	11
1.4 Contribuições da Pesquisa	11
1.5 Descrição dos Capítulos	12
2 REVISÃO DE LITERATURA	13
2.1 Modelos Estatísticos	13
2.1.1 Métodos de previsão	16
2.1.2 Alisamento exponencial	17
2.1.3 Modelos estacionários	20
2.1.3.1 Modelo auto-regressivo (AR)	20
2.1.3.2 Modelos de média móvel (MA)	22
2.1.3.3 Modelo auto-regressivo de médias móveis (ARMA)	25
2.1.4 Modelo auto-regressivo integrado de médias móvel (ARIMA)	26
2.1.4.1 Fase de identificação	27
2.1.4.2 Estimação	33
2.1.4.3 Validação dos parâmetros estimados	33
2.1.5 Modelos sazonais	36
2.1.6 Raiz unitária	37
2.2 Redes Neurais Artificiais	40
2.2.1 Funções de ativação	43
2.2.2 Arquitetura	45
2.2.3 Processos de aprendizado	47

2.2.3.1 Situações de aprendizado	50
2.2.4 Redes neurais recorrentes	51
2.2.4.1 Rede recorrente de Elman	53
3 METODOLOGIA	56
3.1 Caracterização da Pesquisa	56
3.2 Dados	57
3.3 Programa R e Modelos de Previsões	59
3.3.1 Programa R	59
3.3.2 Suavização exponencial	60
3.3.3 Modelo ARIMA	62
3.3.4 Redes neurais	64
3.3.5 Medidas de erros	66
4 RESULTADOS E DISCUSSÃO	68
4.1 Método ARIMA	70
4.1.1 Análise de Dados	70
4.1.2 Identificação	72
4.1.3 Testes	74
4.1.4 Modelo ARIMA (2,1,3)(0,0,1)[12]	75
4.2 Holt-Winters	76
4.3 Rede Neural de Elman	82
4.4 Análise Comparativa dos Modelos	84
5 CONCLUSÃO	87
6 BIBLIOGRAFIA	89
APÊNDICE A – Códigos no R	

LISTA DE FIGURAS

Figura 1 – Produção de etanol em Goiás, de 2004 até 2011	9
Figura 2 – Exemplo simulado de AR (1) com $\phi_1 = +0,9$ e $\phi_1 = -0,9$.	21
Figura 3 – Exemplo simulado de MA (1) com $\theta = +0,5$ e $\theta = -0,5$	23
Figura 4 – Neurônio de McCulloch e Pitts	40
Figura 5 – Generalização do modelo McCulloch e Pitts	41
Figura 6 – Rede acíclica com única camada de neurônios	45
Figura 7 – Rede com múltiplas camadas	46
Figura 8 – Redes recorrentes	51
Figura 9 - Rede genérica de Elman	53
Figura 10 – Arquitetura detalhada da rede Elman	54
Figura 11– Representação esquemática do algoritmo Box-Jenkins	61
Figura 12– Evolução do preço do etanol	67
Figura 13 – <i>Boxplot</i> do preço do etanol	68
Figura 14 - Decomposição da série	69
Figura 15 - Autocorrelação (ACF) e autocorrelação parcial (PACF)	70
Figura 16 – Dependência entre defasagens	71
Figura 17 – Histograma e <i>QQ-Plot</i> dos resíduos	73
Figura 18 – Histograma e <i>QQ-Plot</i> dos resíduos do ETS (M, Ad, M)	77
Figura 19 – Histograma e <i>QQ-Plot</i> dos resíduos do ETS (A, Ad, A)	78
Figura 20 – Histograma e <i>QQ-Plot</i> dos resíduos do ETS (Z, Z, Z)	79
Figura 21 – Processo de aprendizado da rede de Elman	83
Figura 22 – Análise comparativa entre valores observados e preditos pelos modelos propostos para o preço do etanol em Goiás, de janeiro de 2005 até dezembro de 2011	85

LISTA DE TABELAS

Tabela 1 – Medidas descritivas da série logarítmica de preços	68
Tabela 2 – Parâmetros estimados do ARIMA (2,1,3)(0,0,1)[12]	72
Tabela 3 – Critérios de informação	73
Tabela 4 – Testes de normalidade	74
Tabela 5 – Testes de autocorrelação	74
Tabela 6 – Valores iniciais para as equações de recorrência	76
Tabela 7 – Parâmetros de suavização	76
Tabela 8 – Critérios de informação	76
Tabela 9 – Testes de normalidade do modelo ETS (M, Ad, M)	77
Tabela 10 – Testes de normalidade do modelo ETS (A, Ad, A)	78
Tabela 11 – Testes de normalidade do modelo ETS (Z, Z, Z)	79
Tabela 12 – Erros dos modelos de suavização exponencial	80
Tabela 13 – Erros da rede de Elman	83
Tabela 14 – Resumo das medidas de erros dos modelos propostos	84

LISTA DE QUADROS

Quadro 1 – Equações recursivas para modelos de Holt-Winters	18
Quadro 2 – Classificação dos modelos de suavização exponencial	19
Quadro 3 – Coeficientes $\{a_{n-i+1}\}$ para o teste W de normalidade	34
Quadro 4 – Coleta de dados da ANP em Goiás	57
Quadro 5 – Fórmulas recursivas para previsão dos modelos de suavização exponencial	59

LISTA DE SIGLAS

ADF – Dickey-Fuller Aumentado

ANP – Agência Nacional do Petróleo

AR – *AutoRegressive*

ARCH – *AutoRegressive Conditional Heterocedasticity*

ARMA – *AutoRegressive Moving Averege*

ARIMA – *AutoRegressive Integrated Moving Average*

ARFIMA – *AutoRegressive Fractionally Integrated Moving Average*

AIC – *Akaike Information Criterion*

B – Operador de Defasagem

BIC – *Bayesian Information Criterion*

DF – Dickey-Fuller

FAC – Função de Autocorrelação Amostral

FACP – Função de Autocorrelação Parcial

MA – *Moving Averege*

MAE – *Mean Absolute Error*

ME – *Mean Error*

MAPE – *Mean Average Percentage Error*

RNA – Rede Neural Artificial

RMSE – *Root Mean Square Error*

RSNNS – *R Stuttgart Neural Network Simulator*

SNSS – *Stuttgart Neural Network Simulator*

1 INTRODUÇÃO

No presente trabalho faremos um estudo dos modelos estatísticos clássicos de previsões representados pela suavização exponencial e Box-Jenkins que serão comparados com o método de inteligência artificial através da rede neural de Elman. Esses modelos serão aplicados ao problema de previsão de séries temporais, usando como variável os preços do etanol no estado de Goiás, com objetivo de identificar o modelo com melhor capacidade preditiva, ou seja, menor erro.

Para Bressan e Lima (2002), as previsões econômicas buscam reduzir os cenários de incertezas, principalmente no setor agropecuário que está constantemente sujeito a distúrbios devido a fatores de natureza sazonal, cíclica e irregular e que influenciam o preço final dos produtos. Logo, as projeções de preços representam instrumentos de grande relevância para a tomada de decisões, pois podem reduzir o risco na tomada de decisão dos agentes econômicos.

Os métodos estatísticos para previsão de séries temporais são uma alternativa para reduzir a incerteza no processo de tomada de decisão, pois são estimados através de processos estocásticos, onde existe um padrão persistente ou sistemático no comportamento da variável e que pode ser captado através de uma representação paramétrica. As estimações baseiam-se nas suposições de que a média e variância são constantes e independentes do tempo, porém testes de raízes unitárias mostram que esses pressupostos nem sempre são satisfeitos em séries reais, sendo necessário diferenciar a série até que ela se torne estacionária.

Para Sharda (1994), as formas funcionais dos modelos de previsão das redes neurais são mais flexíveis que os métodos estatísticos, pois assumem que existe uma relação (conhecida ou desconhecida) subjacente entre as entradas, representadas por valores passados da série história e/ou outras variáveis relevantes, e as saídas, que são

os valores futuros. Os modelos estatísticos de previsão têm limitações na estimativa desta função subjacente devido a complexidade do sistema real, enquanto que as redes neurais podem indicar um método alternativo para identificar esta função.

Conforme Zhang *et al.* (1998), uma das principais críticas aos modelos de redes neurais é que elas são uma “caixa-preta”, pois não existe forma explícita para analisar e explicar a relação entre as entradas e saídas, dificultando a interpretação dos resultados. Essa rede pode produzir previsões satisfatórias, mas geralmente oferecem um pequeno conhecimento sobre a estrutura dos dados. Para Faraway e Chatfield (1995), um bom ajuste da rede pode não refletir os resultados esperados, pois é necessário escolher adequadamente o conjunto de variáveis de entrada, arquitetura, funções de ativação, escolha dos pesos iniciais e um algoritmo de treinamento. Esse problema reflete na incerteza quanto a determinação da rede neural.

1.1 Justificativa da Pesquisa

De acordo com Sharda (1994), uma área de grande aplicação para as redes neurais é a previsão, pois ao contrário de modelos tradicionais que são baseados em métodos, as redes neurais são orientadas para dados e métodos adaptativos onde elas aprendem a captar relações funcionais entre os dados. Portanto, são mais adequadas para problemas em que as soluções requerem um conhecimento de difícil especificação, desde que existam dados para que a rede tenha um processo de aprendizado e possa inferir corretamente sobre parte da população, mesmo se os dados apresentarem informações ruidosas.

A abordagem estatística tradicional para previsão de séries temporais, como Box-Jenkins, assume que a série em estudo é gerada a partir de processos lineares. Porém, essa análise pode ser inadequada se a relação subjacente é não linear e não é

razoável supor que uma realização particular de uma série temporal seja gerada por um processo linear, pois os sistemas do mundo real geralmente não são lineares.

Para contornar o problema de não linearidade, Engle (1982) desenvolveu o modelo denominado ARCH (*Autoregressive Conditional Heterocedasticity*) que considera a variância heterocedástica, ou seja, não é constante ao longo do tempo, porém esse modelo ainda é limitado, pois em geral não captura todas as características importantes dos dados. Já as redes neurais, baseadas em métodos não lineares são capazes de realizar essa modelagem, sem o conhecimento das relações entre as variáveis de entrada e saída, representando uma ferramenta mais geral e flexível para as previsões.

A idéia de utilizar redes neurais para previsão não é recente. Um dos primeiros trabalhos data de Hu (1964) que aplicou a rede linear adaptativa de Windrow¹ para a previsão do tempo, porém devido à falta de um algoritmo de treinamento para as redes multicamadas na época, a pesquisa foi bastante limitada.

Werbos (1988) identificou que as redes treinadas com *backpropagation* superaram os métodos estatísticos como as abordagens de regressão e Box-Jenkins. Lapedes e Farber (1988) treinaram uma rede neural não-linear e simularam a relação entre pontos sucessivos de séries temporais geradas artificialmente. Os resultados indicaram que as redes podem ser usadas para previsão de séries temporais não-lineares.

Brace *et al.* (1991) concluíram que as redes neurais não são tão boas quanto os métodos estatísticos tradicionais ao avaliar a série carga elétrica. Caire *et al.* (1992) identificaram que para o consumo diário de eletricidade as redes neurais apresentam desempenho um pouco melhor que o ARIMA (*Autoregressive Integrated Moving*

¹ Em 1962, Windrow propôs uma das primeiras redes neurais em camadas treináveis com múltiplos elementos adaptativos, chamada *Madaline (Multipleadaline)*.

Average) na previsão um passo a frente, porém são bem mais confiáveis para previsões por horizonte de tempo maiores.

Sharda e Patil (1992), Tang *et al.* (1991) verificaram que para séries temporais com memória longa os modelos de redes neurais e Box-Jenkins produzem resultados equivalentes. No entanto, para séries temporais com memória curta, Tang *et al.* (1991) identificaram que as redes neurais são superiores ao modelo Box-Jenkins.

Abelém (1994) aplicou as redes neurais para previsão de séries temporais e comparou os resultados com Box-Jenkins através dos erros dos modelos. As variáveis utilizadas foram à série do preço do ouro no mercado internacional e realizadas previsões um passo a frente "*single*" e vários passos à frente "*multi-step*", além de previsões multivariadas utilizando taxas de câmbio de moedas estrangeiras.

Para o método Box-Jenkins o autor encontrou dificuldades na identificação do modelo mais adequado para a série. Alterações nos parâmetros auto-regressivo de ordem p , número de diferenças para tornar a série estacionária e média móvel de ordem q não provocaram nenhum efeito, fornecendo resultados praticamente iguais. Na maioria das vezes, os parâmetros estimados não foram significativos, indicando que a série não possuía regularidades.

A capacidade de previsão das redes neurais foi comprovada através de previsões "*single*" e "*multi-step*" do preço do ouro no mercado internacional. Nas previsões "*multi-step*", os resultados apresentados foram razoáveis em função da realimentação dos valores previstos pelas RNAs às suas entradas, introduzindo um erro que cresceu exponencialmente. Todavia, as RNAs comprovaram seu potencial fornecendo as indicações de tendências e de pontos de mudanças. As previsões "*single-step*" apresentaram resultados muito bons, tanto no que diz respeito à indicação de tendências como na precisão dos dados. (ABELÉM, 1994, p.65).

Faraway e Chatfield (1995) analisaram o desempenho preditivo dos modelos ARIMA e redes neurais através de dados extraídos do estudo de Tang *et al.* (1991) sobre o volume de passageiros em vôos dos Estados Unidos. Os resultados das previsões de curto, médio e longo prazo indicaram resultado superior dos modelos ARIMA.

Zhang *et al.* (1998) apresentaram um artigo referente ao estado da arte das redes neurais aplicadas no problema da previsão. Os autores concluíram que as características das redes neurais, como a adaptabilidade, não linearidade e o mapeamento de uma função trazem um desempenho satisfatório de previsão. Porém, para esse problema específico, resultados não são conclusivos se as redes neurais são melhores que os métodos estatísticos clássicos como Box-Jenkins.

Shabri (2001) comparou Box-Jenkins e redes neurais com algoritmo *backpropagation* para a previsão de cinco séries temporais distintas. Os resultados indicaram que para as séries com tendência e sazonalidade, tanto o Box-Jenkins quanto as redes neurais tiveram bom desempenho na previsão. No entanto, para séries que apresentam padrão irregular, Box-Jenkins superou a acurácia das redes neurais.

Bressan (2001) aplicou os modelos ARIMA, estruturais, bayesianos e redes neurais para a previsão de séries temporais de negociações de contratos futuros do boi gordo, café e soja, em operações de compra e venda de contratos nesses mercados. A precisão preditiva dos resultados ocorreu através do erro percentual de previsão e estatísticas relativas aos retornos financeiros em simulações de compra e venda de contratos como média, desvio padrão, assimetria, curtose e índice Sharpe.

Os resultados indicaram que para períodos de baixa volatilidade, os modelos possuem desempenho similar devido a facilidade de captar o padrão de comportamento dessas séries. Porém, para períodos de grande volatilidade, a rede neural não teve

resultados satisfatórios, diferente dos demais modelos, o que poderia ser associado ao critério de construção da rede e no processo de estimação e treinamento.

Por fim, o autor concluiu que, no período analisado, o melhor desempenho preditivo e operacional nos três mercados foi ao ARIMA em função de sua adaptabilidade e estrutura parcimoniosa, produzindo melhores previsões em termos agregados, com médias positivas em simulações de compra e venda de contratos futuros das commodities.

Gomes (2005) analisou o desempenho preditivo das redes neurais recorrentes para séries temporais de memória curta e longa, tendo como referência modelos ARIMA e ARFIMA (*AutoRegressive Fractionally Integrated Moving Average*). Os dados utilizados foram às séries espaciais (variáveis do solo) e séries temporais (inflações de países desenvolvidos). As redes neurais recorrentes apresentaram bons resultados para as séries espaciais devido a não linearidade. Para as séries econômicas, as redes recorrentes, redes *feedforward* e modelos lineares se alternaram como melhores modelos para previsão da taxa de inflação dependendo do país analisado e do horizonte de previsão.

Catalão *et al.* (2007) realizaram projeções do preço da energia para os mercados de curto prazo da Espanha e Califórnia através das redes neurais e os resultados foram comparados com os modelos ARIMA e Naive desenvolvidos pelos trabalhos de Contreras *et al.* (2003) e Conejo *et al.* (2005). O algoritmo de Levenberg-Marquardt foi aplicado para treinar a rede e a conclusão foi de que a rede neural foi superior a técnica ARIMA e Naive para todas as previsões realizadas.

Mileski Jr. (2007) determinou o método mais adequado para previsão da demanda por produtos de uma empresa do setor de perfumes e cosméticos com comportamento sazonal através da comparação entre modelos de suavização

exponencial, Box-Jenkins e redes neurais tipo *feedforward* com algoritmo *backpropagation* para ajuste dos pesos. Todos os métodos tiveram um bom desempenho para a previsão, porém quando comparado o erro absoluto médio percentual, o sinal de rastreabilidade, o intervalo de confiança, a média e o desvio padrão da amostra, as redes neurais tiveram melhores resultados que os demais.

Flores (2009) comparou o desempenho preditivo entre os modelos estatísticos ARIMA/SARIMA com redes neurais *multilayer perceptron* de múltiplas camadas. Foram analisadas as séries não lineares da produção industrial do Rio Grande do Sul e manchas solares. Os resultados foram comparados através dos erros de cada modelo, concluindo que Box-Jenkins foi superior as redes neurais para a série de produção industrial, enquanto que para os dados das manchas solares, a rede neural apresentou melhor capacidade preditiva.

Kirsten (2009) comparou os modelos de média móvel centrada com índice de sazonalidade, Holt-Winters, redes neurais *perceptron* de múltiplas camadas (RNA-MLP) e função de base radial (RNA-RBA) para as séries Bovespa, Dow Jones e Nasdaq. Os erros serviram para mensurar a acurácia de cada modelo e ao final concluiu que a RNA-MLP obteve melhor desempenho para a série Nasdaq, enquanto que modelos de média móvel centrada com índice de sazonalidade e o Holt-Winters foram mais adequados para série Dow Jones e os modelos lineares foram superiores aos modelos não lineares na série Ibovespa.

O modelo RN-MLP foi mediano para a série Dow Jones e ruim para Ibovespa. A justificativa para esses resultados foi que as amostras da fase de teste faziam parte de uma crise, onde os preços variavam bruscamente e durante a fase de estimação somente a série Nasdaq apresentou-se afetada por algum tipo de crise. Portanto, a RN-MLP

aprendeu com a crise durante a fase de estimação e apresentou um desempenho melhor na fase de teste, pois outra crise estava em andamento.

Este fato também explicaria o motivo de um método linear simples, como a média móvel centrada com índice de sazonalidade, obter melhor resultado que um não linear mais complexo, como a da série Ibovespa. Como a primeira crise praticamente não afetou esta série, nenhum dos modelos mais complexos conseguiu prever com mais acurácia o que estava acontecendo na fase de teste. (KIRSTEN, 2009, p.76).

Sobreiro *et al.* (2009) identificaram uma lacuna na previsão do preço dos combustíveis, que ficou mais evidente quando se observa o etanol no Brasil. Os autores compararam o desempenho das redes neurais de arquitetura *perceptron* de múltiplas camadas com o ARIMA. Os resultados obtidos através dos erros dos modelos indicaram que a rede neural apresentou melhor desempenho para previsão do preço do etanol quando comparado pelo modelo ARIMA.

Pasquotto (2010) utilizou as redes de Elman para previsão de séries temporais da demanda por produtos farmacêuticos, demanda por adubos e tráfego aéreo e comparou os resultados com modelos lineares sazonais através de Box-Jenkins. A qualidade de cada modelo foi verificada através do erro médio quadrático e do coeficiente de Theil. A conclusão foi de que a previsão de Box-Jenkins foi superior a rede de Elman para as três variáveis analisadas e a presença de tendência e sazonalidade nas séries temporais teve impacto negativo na qualidade de previsão das redes neurais.

Considerando todos os estudos apresentados e que os resultados obtidos não são conclusivos acerca da superioridade de previsão das redes neurais em relação aos modelos estatísticos de séries temporais, justifica-se o objetivo geral deste trabalho quanto à comparação entre métodos estatísticos e de inteligência artificial para predição de séries temporais.

A escolha da rede neural de Elman para a modelagem ocorreu em virtude da mesma se diferenciar da grande maioria das redes por apresentar uma estrutura de realimentação que permite que o fluxo de informações circule da camada escondida para a camada de entrada, realizando o processo de treinamento, de forma a introduzir memória a rede neural.

A série temporal analisada foi o preço do etanol no estado de Goiás, selecionada devido à importância da produção para o estado de Goiás que devido a seu desempenho no mercado nacional, consolidou-se como o segundo maior produtor de etanol na safra de 2010/2011, ficando atrás somente de São Paulo (CONAB, 2011). A figura abaixo demonstra o crescimento da produção do etanol em Goiás nas últimas safras.

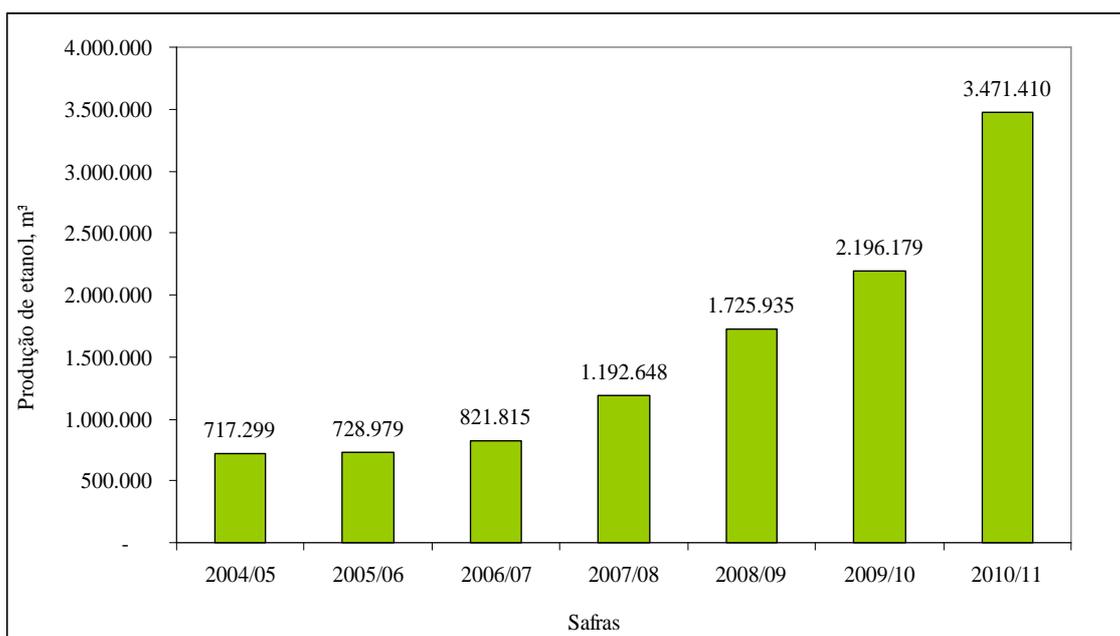


Figura 1 – Produção de etanol em Goiás, de 2004 até 2011

Fonte: SIFAEG (2011)

Segundo Bacchi e Alves (2004), um aspecto a ser considerado na análise de preços no setor sucroalcooleiro é a grande variabilidade ao longo do tempo que se refere a diferentes níveis de preços observados em anos-safra consecutivos e variações entre os

diferentes meses do ano-safra. Os níveis de preços no mercado doméstico de açúcar e álcool estão relacionados aos montantes de cana-de-açúcar produzidos e à alocação dessa matéria-prima entre o açúcar e álcool.

Para Moares (2002), o mercado de álcool é mais complexo que os demais, pois se relaciona com os mercados de combustíveis, açúcar e cana. A escolha da produção entre o açúcar e álcool depende dos preços relativos. A demanda pelo álcool hidratado depende do preço do álcool em relação à gasolina, cujo preço é fixado com base nos preços internacionais do petróleo que sofrem influência das políticas de preços dos principais países produtores do petróleo e na taxa de câmbio.

A oferta do álcool anidro ou hidratado também depende dos preços relativos. Se houver redução da demanda pelo álcool hidratado, ocorre queda dos preços desse produto e redirecionamento para a produção de álcool anidro ou açúcar dado a flexibilidade existente, o que pode ocasionar a escassez do álcool hidratado.

Dessa forma, a projeção de preços orienta os agentes econômicos que formam expectativas em relação ao preço e quantidade produzida, reduzindo risco e incertezas na tomada de decisões. Em relação à adoção de políticas públicas, a projeção de preço do etanol permite ao governo a análise prospectiva para entender o comportamento do setor sucroalcooleiro e de combustíveis ao longo da safra frente às variações dos preços tanto do álcool, açúcar e gasolina.

1.2 Definição do Problema de Pesquisa

Conforme Portugal e Fernandes (1996), em muitos casos, as redes neurais modelam irregularidades em séries temporais de forma superior aos modelos estatísticos tradicionais. Porém, esse método necessita um número maior de dados observados na série e nem sempre se consegue entender o que se passa na modelagem, pois não existe um modelo explícito.

Dessa forma, considerando a natureza cíclica e a elevada volatilidade dos preços do etanol em Goiás, a rede neural de Elman apresenta capacidade preditiva superior aos métodos estatísticos de séries temporais, considerando a previsão um passo a frente?

1.3 Objetivos

O objetivo geral deste trabalho foi comparar modelos estatísticos clássicos de séries temporais com o método de inteligência artificial para o problema de previsão do preço do etanol em Goiás. Os modelos analisados foram Holt-Winters, Box-Jenkins e a rede neural recorrente de Elman. Ao final, a capacidade preditiva dos modelos foi comparada através de medidas estatísticas do termo de erro.

Os objetivos específicos foram delimitados:

- i) Revisão de literatura referente aos principais modelos estatísticos de previsão de séries temporais;
- ii) Estudo da aplicação das redes neurais para os problemas de previsão;
- iii) Identificar e analisar as características da série temporal em estudo e desenvolver os modelos propostos;
- iv) Comparar os resultados entre os métodos estatísticos de séries temporais e a rede neural de Elman através erro médio (ME), média absoluta dos erros (MAE), média absoluta percentual dos erros (MAPE) e raiz do erro médio quadrático (RMSE).

1.4 Contribuições da Pesquisa

Este trabalho contribuiu para identificação de um método mais adequado para previsão de série temporal com tendência e sazonalidade que se refere aos preços do etanol no estado de Goiás, comparando modelos estatísticos de séries temporais como

Holt-Winters e Box-Jenkins com métodos de inteligência artificial através da rede neural recorrente de Elman.

1.5 Descrição dos Capítulos

A organização deste trabalho foi estruturada da seguinte forma:

O capítulo 2 apresentou a fundamentação teórica dos modelos estatísticos de previsão de séries temporais como o Holt-Winters e Box-Jenkins (ARIMA) e foi realizada revisão de literatura sobre as redes neurais artificiais, com ênfase nas redes neurais recorrentes.

O capítulo 3 definiu a metodologia, com caracterização da pesquisa, a série temporal escolhida, os modelos aplicados para gerar as previsões através de Holt-Winters (aditivo e multiplicativo), Box-Jenkins e rede neural de Elman, além dos índices de desempenho para mensurar os erros de cada modelo.

O capítulo 4 demonstrou os resultados das aplicações dos modelos de previsão, bem como comentários sobre desempenho dos mesmos.

O capítulo 5 trouxe as conclusões obtidas pela pesquisa e sugestões para trabalhos futuros.

2 REVISÃO DE LITERATURA

Neste capítulo faremos a revisão teórica sobre os métodos aplicados neste trabalho, com ênfase nos modelos estatísticos de séries temporais de suavização exponencial e Box-Jenkins e redes neurais, destacando a rede de Elman.

2.1 Modelos Estatísticos

Conforme Fischer (1982), uma série temporal é um conjunto de observações de uma variável dispostas sequencialmente no tempo. De acordo com o conjunto gerado, a série é classificada com discreta, quando o conjunto de observações no tempo for finito ou infinito enumerável. Caso contrário, se a série for infinita e não numerável é classificada como contínua.

Em relação à formação das séries, podemos classificá-las como determinística, ou seja, quando os valores futuros são estabelecidos através da relação funcional matemática do tipo $y = f(\textit{tempo})$ ou estocástica se esses valores são expostos em termos probabilísticos, uma vez que a série está descrita por meio de uma relação funcional que envolve não só o tempo, mas também uma variável aleatória do tipo $y = f(\textit{tempo}, \varepsilon_t)$, em que ε_t é o termo aleatório residual, necessário quando não se consegue explicar completamente algum movimento irregular da série através unicamente da relação matemática².

O objetivo da análise dessas séries é identificar um padrão de comportamento que possibilite previsões que, para abordagem econômica, pode ser realizada através de modelos de regressão de equação única; regressão com equações simultâneas; modelos

² Neste trabalho somente serão consideradas séries temporais estocásticas discretas, em que as observações serão geradas em intervalos de tempo contínuos.

auto-regressivos integrados de média móvel e auto-regressão vetorial (MORETTIN, 2008).

O processo estocástico é estacionário se suas médias e variâncias forem constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre dois períodos, e não do período de tempo efetivamente em que a covariância é calculada. Essa propriedade permite a realização de inferências estatísticas sobre os parâmetros estimados. Conforme Matos (2000), a não-estacionariedade constitui uma violação de pressuposto, cuja consequência é a possibilidade de se obter resultados espúrios.

Considera-se que um processo estocástico é uma família de variáveis aleatórias (v.a.) definidas no mesmo espaço de probabilidades. Seja T um conjunto arbitrário, um processo estocástico é definido por $\{y(t), t \in T\}$, tal que para cada $t \in T, y(t)$ é uma variável aleatória e t é o tempo. O conjunto de valores $\{y(t), t \in T\}$ é denominado de espaço dos estados e $y(t)$ representam os estados. Para cada $t \in T$, existe uma variável aleatória, com uma função de densidade de probabilidades $f_i(y)$ (MORETTIN, 2008).

Se as distribuições finito-dimensionais permanecerem as mesmas sob translações de tempo, o processo estocástico é estritamente estacionário, logo a média $\mu(t)$ e a variância σ^2 serão constantes para todo $t \in T$ e a covariância igual a 0, de forma que:

$$\text{Média: } E\{y(t)\} = \mu(t) = \mu \quad (2.1)$$

$$\text{Variância: } Var\{y(t)\} = \sigma^2(t) = \sigma^2 \quad (2.2)$$

$$\text{Covariância: } \gamma(t_1, t_2) = \gamma(t_1 - t_2, 0) = Cov\{y(t_1 - t_2), y(0)\} \quad (2.3)$$

Na estacionariedade estrita, a forma da distribuição conjunta do processo permanece sem variação mediante uma translação temporal. Porém, na prática é

complexa a especificação da distribuição conjunta de um processo estocástico, então existe a versão mais fraca, a qual denominaremos por processo fracamente estacionário, onde somente alguns momentos do processo permanecem inalterados no tempo, ou seja, um processo estocástico $\{y(t), t \in T\}$ é fracamente estacionário (ou estacionário de segunda ordem) se e somente se:

$$E\{y(t)\} = \mu(t) = \mu \quad (2.4)$$

$$E\{y^2(t)\} < \infty \quad (2.5)$$

$$\gamma(t_1, t_2) = \text{Cov}\{y(t_1), y(t_2)\} \text{ é uma função apenas de } |t_1 - t_2|. \quad (2.6)$$

Para Bueno (2008), além da estacionariedade, é necessário satisfazer a propriedade de ergodicidade que permite usar a série para calcular as médias em cada instante de tempo. Como todas as médias são iguais, basta uma única realização da série para viabilizar o cálculo. Se a média temporal convergir para a média não condicional, há ergodicidade e a série temporal pode ser estimada normalmente, mesmo com apenas uma realização do processo estocástico.

Outro processo importante é o ruído branco, caracterizado por uma sequência do termo de erro ε_t , com média zero, variância constante e não é correlacionado com qualquer realização da própria série, ou seja, autocorrelação igual a zero, $RB(0, \sigma^2)$, onde:

$$E(\varepsilon_t) = 0; \forall t \quad (2.7)$$

$$E(\varepsilon_t^2) = \sigma^2; \forall t \quad (2.8)$$

$$E(\varepsilon_t \varepsilon_{t-j}) = 0; \text{ para todo } j \neq 0 \quad (2.9)$$

Até o início da década de 1970, a análise de série temporal ocorria através da sua decomposição em tendência, ciclo, sazonalidade e variação irregular ou componente errático. Por essa abordagem, uma série pode resultar da combinação de todos os quatro componentes ou de um subconjunto deles.

De acordo com Fischer (1982), a tendência caracteriza um movimento regular e contínuo de longo prazo. Descreve as variações graduais que se mantêm em um longo período e se caracteriza pelo fato de que a variância se altera ao longo do tempo. Corrar e Theóphilo (2004) afirmam que a tendência é a componente mais importante da série temporal.

Os ciclos são variações que se referem às oscilações de longo prazo que caracterizam, em geral, os ciclos econômicos. São as flutuações em torno da curva de tendência e são mais difíceis de determinar, pois tanto o período considerado ou a causa do ciclo podem não ser conhecidos.

A sazonalidade corresponde às oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia. A diferença entre a sazonalidade e o ciclo é que o primeiro possui movimentos previsíveis, ocorrendo em intervalos regulares de tempo, enquanto que movimentos cíclicos tendem a ser irregulares.

Para Morettin e Tolói (2006), as variações irregulares são descolamentos esporádicos das séries, decorrente de causas naturais e sociais, provocadas por eventos imprevisíveis e não periódicos denominados ruídos, como as enchentes, eleições, greves, entre outros. Variações estas que não podem ser previstas, portanto não são incluídas no modelo.

2.1.1 Métodos de previsão

Conforme Brooks (2002), os modelos de séries univariadas de tempo representam uma classe específica, onde se objetiva modelar e prever variáveis utilizando apenas informações contidas em seus próprios valores passados e atuais e um termo de erro. São considerados ateóricos, o que implica que a sua construção e utilização não é baseada em qualquer teoria subjacente ao comportamento de outra

variável, ou seja, não existe qualquer independência e causalidade com outras séries temporais.

Esta prática pode ser contrastada com modelos estruturais, que são de natureza multivariada ou causal e tentam explicar as mudanças em uma variável dependente através de referência aos movimentos nos valores atuais ou passados de outras variáveis (explicativas). Porém, os modelos estruturais não serão abordados neste trabalho.

2.1.2 Alisamento exponencial

De acordo com Gelper *et al.* (2007), o alisamento exponencial é uma técnica usada para suavização e previsão de uma série temporal sem a necessidade de adaptação de um modelo paramétrico. É baseado em um esquema de computação recursiva onde as previsões são atualizadas a cada nova observação de entrada.

Esses métodos utilizam pesos exponenciais decrescentes no sentido das observações mais recentes em direção às observações mais antigas, onde existem um ou mais parâmetros de suavização a serem determinados e essa escolha desses parâmetros definirá os pesos a serem dados a cada observação. Os principais modelos são a média móvel simples, média móvel ponderada, suavização exponencial simples, dupla suavização exponencial e Holt-Winters.

Se os dados não possuem padrões de tendência ou sazonalidade, o alisamento exponencial simples é adequado. Para os dados que indicam tendência linear, o método Holt é apropriado. Porém, se os dados são sazonais é necessário aplicar o Holt-Winters para captar a sazonalidade (Mileski Jr., 2007), baseado em três equações de suavização: nível, que é estacionário e não exibe tendência ou sazonalidade, além dos componentes de tendência e sazonalidade e pode ser modelado de forma multiplicativa ou aditiva.

Os modelos aditivos são utilizados quando não existe interação entre tendência e sazonalidade, enquanto que a forma multiplicativa é aplicada quando as flutuações

sazonais aumentam ou diminuem proporcionalmente com a elevação ao decréscimo do nível da série. As equações de Holt-Winters estão apresentadas no quadro 1.

Quadro 1 – Equações recursivas para modelos de Holt-Winters

	Holt-Winters Aditivo	Holt-Winters Multiplicativo
Nível	$\ell_t = \alpha(y_t - s_{t-s}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ (2.10)	$\ell_t = \alpha\left(\frac{y_t}{s_{t-s}}\right) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ (2.11)
Tendência	$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$ (2.12)	$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1}$ (2.13)
Sazonalidade	$s_t = \gamma(y_t - \ell_t) + (1-\gamma)s_{t-s}$ (2.14)	$s_t = \gamma\left(\frac{y_t}{\ell_t}\right) + (1-\gamma)s_{t-s}$ (2.15)
Previsão	$F_{t+m} = (\ell_t + b_{t-m})s_{t-s+m}$ (2.16)	$F_{t+m} = (\ell_t + b_{t-m})s_{t-s+m}$ (2.17)

Fonte: Mileski Jr.(2007)

Onde:

ℓ_t = elemento de nível para o período t ;

b_t = elemento de tendência para o período t ;

s_t = elemento de sazonalidade para o período t ;

F_{t+m} = valor da previsão para o período m ;

y_t = valor observado no período t ;

s = intervalo de tempo da sazonalidade;

m = número do período previsto;

α = constante do ajuste de nível;

β = constante do ajuste de tendência;

γ = constante do ajuste de sazonalidade.

Hyndman *et al.*(2002) propuseram a taxonomia para os modelos de suavização exponencial de acordo com quadro 2, onde cada método é indicado por uma ou duas letras para tendência e outra escrita para sazonalidade. A nomenclatura NN descreve a suavização exponencial simples. AN se refere ao modelo Holt. O modelo aditivo de Holt-Winters é caracterizado por AA, enquanto o Holt-Winters multiplicativo é denotado por AM. As demais nomenclaturas representam modelos menos utilizados.

Quadro 2 – Classificação dos modelos de suavização exponencial

Componente de tendência	Componente de sazonalidade		
	N (nenhum)	A (aditivo)	M (multiplicativo)
N (nenhum)	N,N	N,A	N,M
A (aditivo)	A,N	A,A	A,M
A _d (aditivo amortecido)	A _d ,N	A _d ,M	A _d ,M
M (multiplicativo)	M,N	M,A	M,M
M _d (multiplicativo amortecido)	M _d ,N	M _d ,A	M _d ,M

Fonte: Hyndman *et al.* (2002)

Para cada um dos quinze métodos indicados, ainda existem duas possibilidades de modelagem no espaço de estados que correspondem ao modelo com erros aditivos e outro com erros multiplicativos. Dessa forma, existem 30 possibilidades de modelos de suavização exponencial que podem descrever uma série temporal (HYNDMAN *et al.* 2002, GARDNER JR, 2006)

Para Brooks (2002), as desvantagens dos métodos de suavização exponencial é o fato de que geralmente não são ideais para capturar qualquer dependência linear nos

dados. Além disso, as previsões desses modelos não convergem para a média de longo prazo da variável com o aumento do horizonte de tempo.

2.1.3 Modelos estacionários

2.1.3.1 Modelo auto-regressivo (AR)

O modelo auto-regressivo AR (p) é aquele em que a variável endógena de um período é explicada por seus valores passados (parte sistemática) e um termo de ruído branco (parte aleatória). Para processos estacionários com distribuição normal é expresso por:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.18)$$

Onde:

y_t = valor observado da série temporal no instante t ;

ϕ_i = i -ésimo parâmetro auto-regressivo estimado, $i = 1, 2, \dots, p$;

Assume-se que valor observado y_t é estacionário, $\phi_1, \phi_2, \dots, \phi_p$ são constantes ($\phi_p \neq 0$) e o termo de erro ε_t possui média $\mu = 0$ e variância constante σ_ε^2 . Se a média, μ , for diferente de zero, y_t deve ser substituído por $y_t - \mu$ na equação (2.18)

(SHUMWAY, STOFFER, 2010):

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \dots + \phi_p (y_{t-p} - \mu) + \varepsilon_t \quad (2.19)$$

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

Onde $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$

Uma forma útil para escrever a equação (2.19) é através do operador de defasagem B dado por:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \varepsilon_t \quad (2.20)$$

Onde:

$$B y_t = y_{t-1}$$

$$B^2 y_t = B(By_t) = By_{t-1} = y_{t-2}$$

Com o operador de defasagem, (2.18) é reescrita de forma mais simplificada:

$$\phi(B)y_t = \varepsilon_t \quad (2.21)$$

As propriedades de $\phi(B)$ são importantes para solução de (2.21). Isso leva a definição do operador auto-regressivo dado por:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.22)$$

A variância e as autocovariâncias do processo AR (p) são respectivamente:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 \quad (2.23)$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (2.24)$$

Para exemplificar, considere que y_t depende somente de y_{t-1} e do termo de erro ε_t , indicando um modelo auto-regressivo de ordem $p=1$, ou seja, AR (1), especificado por:

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (2.25)$$

Se $|\phi| < 1$ e y_t é estacionário, representa-se um modelo AR (1) como um processo linear determinado por:

$$y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \quad (2.26)$$

A figura 2 ilustra um exemplo de um processo aleatório simples de AR (1), também chamado de processo de Markov, com $\phi_1 = +0,9$ e $\phi_2 = -0,9$.

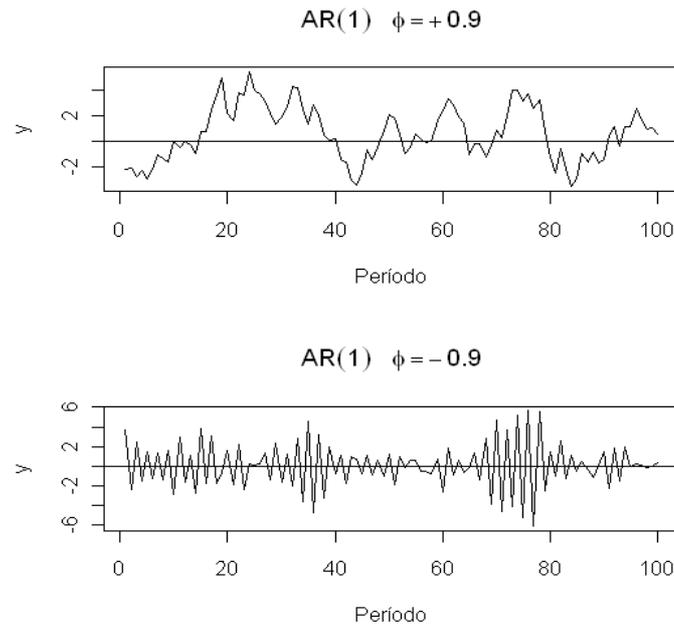


Figura 2 - Exemplo simulado de AR (1) com $\phi = +0,9$ e $\phi = -0,9$.

Fonte: Shumway e Stoffer (2010)

A condição de estacionariedade do AR(p), em que $|\phi| < 1$ estabelece que todas as p raízes da equação característica $\phi(B) = 0$ estão fora do círculo unitário e para que a variância de y_t seja não negativa e finita (FAVA, 2000).

2.1.3.2 Modelos de médias móveis (MA)

No modelo de médias móveis (MA), a série y_t resulta da combinação linear dos choques aleatórios (ruídos brancos) ocorridos no período corrente e passado. O modelo genérico envolve q valores defasados de ε_t e é indicado por MA(q) (FAVA, 2000), através da equação:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.27)$$

Onde:

y_t é o valor observado da série temporal no instante t ;

$\theta_1, \theta_2, \dots, \theta_q (\theta_q \neq 0)$ são parâmetros de média móvel estimado, $i = 1, 2, \dots, q$;

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, ou seja, ruído branco no instante t ;

q é o número de defasagens da média móvel;

É possível reescrever o processo MA (q), em (2.27), na forma equivalente através da equação abaixo:

$$y_t = \theta(B)\varepsilon_t \quad (2.28)$$

Em que o operador de média móvel é dado por:

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_p B^p \quad (2.29)$$

A variância do MA (q) é demonstrada por:

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma_\varepsilon^2 \quad (2.30)$$

Apenas as autocovariâncias de ordem menor ou igual a q são nulas e sua expressão é:

$$\gamma_k = (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q)\sigma_\varepsilon^2 \quad (2.31)$$

Para $k > q$, $\gamma_k = 0$. A condição de invertibilidade requer que todas as raízes da equação característica $\theta(B)=0$ caiam fora do círculo unitário.

Para exemplificar, considere o modelo de médias móveis de primeira ordem MA (1), em que:

$$y_t = \varepsilon_t - \theta\varepsilon_{t-1} \quad (2.32)$$

Onde θ é o parâmetro de média móvel estimado, com a média de y_t igual a zero e $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.

A figura 3 ilustra um exemplo de processo aleatório simples MA (1) com $\theta = +0,5$ e $\theta = -0,5$.

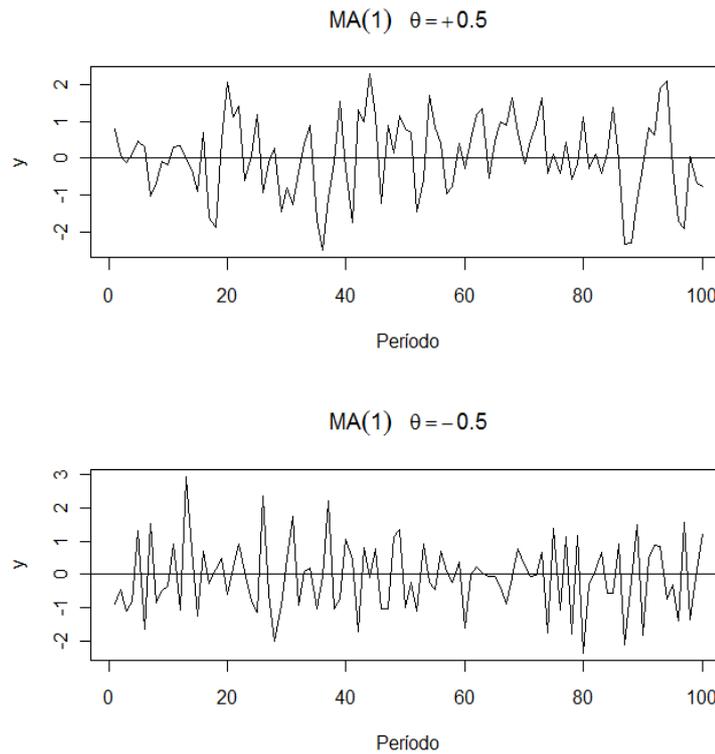


Figura 3 - Exemplo simulado de MA (1) com $\theta = +0,5$ e $\theta = -0,5$

Fonte: Shumway e Stoffer (2010)

A variância deste modelo é dada por (2.33):

$$\begin{aligned}\gamma_0 &= E[y_t^2] = E[(\varepsilon_t - \theta\varepsilon_{t-1})^2] \\ \gamma_0 &= \sigma_\varepsilon^2 + \theta^2\sigma_\varepsilon^2 \\ \gamma_0 &= \sigma_\varepsilon^2(1 + \theta^2)\end{aligned}\tag{2.33}$$

As autocovariâncias do MA (1) são definidas da forma seguinte:

$$\begin{aligned}\gamma_1 &= E[y_t y_{t-1}] = E[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-1} - \theta\varepsilon_{t-2})] \\ \gamma_1 &= -\theta\sigma_\varepsilon^2 \\ \gamma_2 &= E[y_t y_{t-2}] = E[(\varepsilon_t - \theta\varepsilon_{t-1})(\varepsilon_{t-2} - \theta\varepsilon_{t-3})]\end{aligned}\tag{2.34}$$

Portanto, $\gamma_2 = 0$.

A autocovariância de ordem maior ou igual a dois são nulas. As expressões da variância e autocovariância indicam que não é necessário impor nenhuma restrição ao parâmetro θ para obter a estacionariedade fraca (FAVA, 2000).

2.1.3.3 Modelo auto-regressivo de médias móveis (ARMA)

Este modelo é a combinação dos processos AR e MA, descrito por seus p valores passados e q ruídos brancos correntes e passados para séries temporais estacionárias, ilustrado a seguir (SHUMWAY, STOFFER, 2010):

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.35)$$

Em que:

ϕ_i = parâmetro auto-regressivo;

θ = parâmetro de média móvel;

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ = ruído branco, com média zero e variância constante.

Se y_t tem média μ diferente de zero, pode-se definir que $\alpha = \mu(1 - \phi_1 - \dots - \phi_p)$

e o modelo pode ser reescrito por:

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2.36)$$

Quando $q = 0$, o modelo é chamado de auto-regressivos de ordem p , AR (p), e quando $p = 0$, o modelo é chamado de média móvel de ordem q , MA (q). Para auxiliar na investigação dos modelos ARMA, é útil escrevê-los usando o operador de defasagem do processo AR (2.22) e operador da MA (2.29). Assim, o modelo ARMA (p, q) (2.35) pode ser descrito de forma simplificada:

$$\phi(B)y_t = \theta(B)\varepsilon_t \quad (2.37)$$

Para Morettin (2008), em um processo ARMA (p, q) a condição de estacionariedade é a mesma para o processo AR (p) e equivale que todas as raízes de $\phi(B)=0$ estão fora do círculo unitário e a condição de invertibilidade é a mesma para o

processo MA (q), ou seja, as raízes de $\theta(B)=0$ também devem estar fora do círculo unitário.

2.1.4 Modelo auto-regressivo integrado de médias móvel (ARIMA)

A relação temporal considerada por Box-Jenkins é representada por um conjunto de processos estocásticos denominado ARIMA (*Autoregressive Integrated Moving Average*), resultantes da combinação dos componentes auto-regressivo (AR), integração (I) e médias móveis (MA). Por envolverem apenas a série de tempo, eles são classificados como modelos univariados (GUJARATI, 2000).

Esses modelos são usados para analisar as séries temporais não estacionárias, transformando-as em estacionárias através da ordem de integração, ou seja, a diferenciação da variável envolvida. A razão para necessitar que os dados sejam estacionários é que qualquer modelo inferido a partir desses dados pode ser considerado como estável, fornecendo uma base válida para previsão.

É representado por ARIMA (p, d, q), onde p é o número de defasagens da série, d é a ordem de integração e q o número de defasagens dos erros aleatórios. As ordens de p e q do modelo, como se trata da combinação dos processos AR e MA, são determinadas com base em funções de autocorrelação parcial e de autocorrelação amostral, assim como exame de diagnósticos efetuados após estimação (MATOS, 2000).

Para representar o modelo ARIMA (p, d, q), a variável y_t é substituída por w_t que representa uma série estacionária, conforme equação abaixo:

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.38)$$

Portanto, $w_t = \nabla^d y_t = (1-B)^d y_t$, ou seja, a série é estacionária de ordem d. $\theta_1, \dots, \theta_q$ são os parâmetros de média móvel de ordem q e ϕ_1, \dots, ϕ_p são os parâmetros auto-regressivos de ordem p (SHUMWAY, STOFFER, 2010).

Generalizando, o modelo fica definido por:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad (2.39)$$

Se $E(\nabla^d y_t) = \mu$, o modelo é reescrito:

$$\phi(B)(1-B)^d y_t = \delta + \theta(B)\varepsilon_t \quad (2.40)$$

Em que $\delta = \mu(1 - \phi_1 - \dots - \phi_p)$.

De acordo com Gujarati (2000), na metodologia Box-Jenkins é necessário descobrir qual entre as várias versões do modelo ARIMA descreve o comportamento da série. Esta etapa é a identificação, onde se determina os valores de p, d e q do modelo. As fases posteriores são a estimação dos parâmetros e a validação, descritos a seguir.

2.1.4.1 Fase de identificação

O primeiro parâmetro a ser identificado é o grau de diferenciação d através do diagrama da função de autocorrelação amostral (FAC), no qual são apresentados os valores das autocorrelações em relação às defasagens³ k. Se as autocorrelações decrescem de forma exponencial, realizam-se diferenciações na série até que o diagrama apresente um corte abrupto para um valor qualquer de autocorrelação, quando a série será considerada estacionária.

Conforme Makridakis *et al.* (1998), considera-se d = 1 quando a série é não estacionária em relação ao processo, ou seja, quando a média durante um período é uma e depois desse período ocorre mudança na média. Se série é não estacionária quanto à inclinação, ou seja, oscila em uma direção em certo tempo e depois muda de direção, d = 2. O número de diferenciações d, em geral, é no máximo igual a 2.

Embora a FAC seja amplamente utilizada para definir a ordem de integração de y_t , pode trazer dúvidas em relação à correta ordem de integração. É interessante utilizar

³ Defasagens ou *lags* indicam intervalos de tempo em que ocorrem máximas correlações.

testes estatísticos para verificar se a diferenciação da série temporal foi suficiente para torná-la estacionária através de testes de raiz unitária. Esses testes representam um procedimento alternativo que consiste não em determinar, mas sim testar a ordem de integração da série y_t e decorre do fato de que o número de diferenças necessárias para tornar uma série temporal estacionária corresponde ao número de raízes sobre o círculo unitário presentes no processo gerador de y_t (FAVA, 2000). Mais informações sobre os testes de raiz unitária são realizadas no tópico 2.1.6, pág. 37.

Feito isto, determina-se o ARMA (p, q) em que as principais ferramentas para identificação são a função de autocorrelação (FAC) que demonstra a ordem q de um processo MA e a função de autocorrelação parcial (FACP) que indica a ordem p do processo AR e seus correlogramas resultantes, que são representações gráficas contra o tamanho da defasagem. O comportamento da FAC e FACP deve imitar os respectivos comportamentos das quantidades teóricas (GUJARATI, 2000).

Para o modelo AR (p) a função de autocorrelação decai exponencialmente com o aumento de defasagens e a função de autocorrelação parcial é truncada a partir da defasagem. Em relação ao MA (q), a função de autocorrelação é truncada na defasagem q e a autocorrelação parcial decai exponencialmente. Conforme Bueno (2008), no caso de ARMA (p, q), ambas as funções decaem exponencialmente a partir da defasagem de truncagem, dificultando o reconhecimento visual.

Para Fava (2000), a FAC e FACP são importantes no processo de identificação, pois como os modelos AR, MA e ARMA só se aplicam em séries originalmente estacionárias, a identificação tem início pela determinação da ordem de integração na série. Se $d = 0$, a série é estacionária e passa para a identificação dos filtros AR e MA. Caso contrário, aplicam-se quantas diferenças forem necessárias para tornar a série estacionária e trabalha-se com os resultados para a identificação de AR e MA.

Porém, esses padrões nem sempre são muito objetivos por serem estocásticos. Uma mesma série pode sugerir padrões alternativos de modelagem em razão da dificuldade em inferir qual é o padrão gerador daquela função. Dependendo da magnitude dos coeficientes dos parâmetros AR e MA, o padrão que emerge da FAC é diferente do verdadeiro padrão do processo gerador de dados.

Visualmente, pode ser difícil identificar quando se inicia o decaimento exponencial. Outro problema ocorre para estimar os valores da variância da autocorrelação e autocorrelação parcial, pois ambas dependem dos verdadeiros valores dos coeficientes, os quais são desconhecidos (BUENO, 2008).

O coeficiente de autocorrelação (FAC) ou correlação serial é dado por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.41)$$

Em que γ_k é a covariância (y_t, y_{t-k}) e γ_0 é a variância de y_t . A sequência de pares (k, ρ_k) , $k = 1, 2, \dots$, é denominada função de autocorrelação. Porém, como esse coeficiente envolve o cálculo de parâmetros desconhecidos, é necessário trabalhar com o coeficiente de correlação amostral, expresso por:

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, k = 1, 2, \dots, \quad (2.42)$$

Onde:

n = tamanho da amostra

Conhecida a distribuição de r_k é possível determinar os intervalos de confiança e os testes de hipóteses usuais para verificar se cada coeficiente de autocorrelação é nulo. Também é possível testar se os primeiros coeficientes de autocorrelação são conjuntamente iguais a zero, ao invés de testar as autocorrelações residuais individuais, através das estatísticas Q . Box e Jenkins (1976) sugeriram o uso do teste de Box-Pierce:

$$Q = n \sum_{k=1}^k r_k^2 \quad (2.43)$$

Onde:

k = número de defasagens do modelo;

n = número de observações;

r_k = coeficiente de correlação amostral;

Porém, o teste Box-Pierce não tem bom desempenho em amostras pequenas ou moderadas no sentido de que a distribuição se afasta da χ^2 . Neste caso, utiliza-se o teste Ljung-Box, na qual a estatística é descrita por:

$$Q(K) = n(n+2) \sum_{k=1}^k \frac{r_k^2}{n-k} \quad (2.44)$$

As equações (2.43) e (2.44) têm distribuição χ^2 com n graus de liberdade e hipótese nula $H_0: \sum_{k=1}^n r_k = 0$ contra $H_1: \sum_{k=1}^n r_k \neq 0$. Para análise dos testes, escolhe-se o nível de significância α , se $Q > \chi_k^2$, rejeita-se a hipótese nula de que os k primeiros coeficientes de autocorrelação são nulos. Os testes acima são válidos se a distribuição é normal e estacionária. Caso haja violação dessas premissas, a potência do teste é reduzida (FAVA, 2000).

O coeficiente de autocorrelação parcial de ordem k representado por ϕ_{kk} mede a correlação entre y_t e y_{t-k} depois que a influência de $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ sobre y_t foi descontada, conforme segue o sistema de equações:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_k \rho_{k-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \dots + \phi_k \rho_{k-2} \\ (\dots) & \\ \rho_k &= \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_k \end{aligned} \quad (2.45)$$

Para Bueno (2008), formas alternativas de identificação foram propostas através de critérios de informação que são mecanismos para encontrar o número ideal

de parâmetros de um modelo, visto que podem ocorrer vários problemas de identificação através da FAC e FACP. A cada regressor adicional, a soma dos resíduos não vai aumentar e, frequentemente, irá diminuir. A redução ocorre em função de mais regressores e o critério de informação associa uma penalidade a esse aumento.

Se a penalidade for menor que a diminuição da soma dos resíduos, o regressor adicional deve ser incorporado ao modelo, caso contrário, se a penalidade for maior que a redução da soma, o regressor adicional traz mais custos que benefícios. Portanto, o critério de informação busca minimizar uma função baseada nos resíduos penalizada pelo número de regressores. Dentre os principais, cita-se o AIC e o BIC.

O critério de AIC, também chamado de *Akaike Information Criterion* e dado respectivamente por:

$$AIC = \log \hat{\sigma}_k^2 + \frac{n+2k}{n} \quad (2.46)$$

$$\text{Onde: } \hat{\sigma}_k^2 = \frac{SSE}{n}; \quad SSE = \sum_{t=1}^n (y_t - \bar{y})^2$$

SSE = soma do quadrado dos resíduos;

$\hat{\sigma}_k^2$ = estimador da máxima verossimilhança da variância;

n = tamanho da amostra;

k = n° de parâmetros do modelo.

Para Shumway e Stoffer (2010), várias correções ocorrem para melhorar o comportamento do AIC, que pode ser tendencioso, diminuindo a probabilidade de se selecionar uma ordem maior que a verdadeira. Em (2.47) apresenta-se a equação corrigida.

$$AICc = \log \hat{\sigma}_k^2 + \frac{n+2k}{n-k-2} \quad (2.47)$$

Também é possível derivar um termo de correção com base em argumentos bayesianos, como em Schwarz (1978), que leva ao BIC, *Bayesian Information Criterion*, representado por (2.48):

$$BIC = \log \hat{\sigma}_k^2 + \frac{k + \log n}{n} \quad (2.48)$$

O critério BIC é consistente assintoticamente, tendendo a escolher um modelo mais parcimonioso que o AIC. Já o AIC funciona melhor em pequenas amostras, não obstante seja viesado para escolher modelos parametrizados (MORETTIN, 2008). Porém, importante observar que estas medidas individualmente, considerando um único modelo, não têm nenhum significado. O AIC e BIC podem assumir quaisquer valores, inclusive negativos, já que eles dependem da função de verossimilhança. Porém, para comparar esses critérios, a estimação precisa ser feita no mesmo período amostral (EHLERS, 2009).

Alguns trabalhos propuseram a abordagem de redes neurais para identificação dos parâmetros do ARIMA, como Jhee *et al.* (1992) que desenvolveram duas redes neurais usadas para modelar a função de autocorrelação (FAC) e função de autocorrelação parcial (FACP), sendo que as saídas dessas RNAs indicaram as ordens de um modelo ARMA. Lee e Jhee (1994), estudaram um sistema de RNAs para identificação automática dos parâmetros de Box-Jenkins usando a função de autocorrelação da amostra através de rede *multilayer perceptron* para filtragem dos ruídos e identificar corretamente os parâmetros do modelo ARMA.

Reynolds *et al.* (1995) também utilizaram as RNAs para o problema de identificação do modelo Box-Jenkins. Duas redes foram criadas, a primeira usada para determinar o número de diferenças necessárias para que uma série temporal se tornasse estacionária, enquanto a segunda teve o objetivo de identificar o modelo ARMA com base nas informações da FAC e FACP da série.

2.1.4.2 Estimação

A segunda etapa da metodologia Box-Jenkins é a estimação dos parâmetros do modelo identificado, ϕ se houver parâmetro auto-regressivo e θ se houver o filtro de médias móveis e a variância do ruído branco. Para um bom ajuste do ARIMA é necessário utilizar técnicas em que a estrutura residual seja um ruído branco, equivalente a uma variável aleatória independentemente e identicamente distribuída, ou seja, $\varepsilon_t \sim i.i.d(0, \sigma^2)$ (BUENO, 2008).

A estimativa dos parâmetros pode ocorrer através do método dos momentos, mínimos quadrados ou a máxima verossimilhança (condicional ou exata). O método dos momentos não tem boas propriedades quando comparados com os demais. A verossimilhança condicional é assintoticamente equivalente ao da máxima verossimilhança exata, mas com a vantagem de especificação e estimar a função, porém não é tão eficiente quanto ao método exato principalmente para pequenas amostras.

2.1.4.3 Validação dos parâmetros estimados

Para Fava (2000), a validação ocorre através da análise dos resíduos e da avaliação da ordem do modelo. Se o modelo identificado for adequado, o mesmo pode ser utilizado para realizar previsões, caso contrário, outra especificação deve ser escolhida para modelar a série temporal.

Além de verificar se os resíduos se comportam como um ruído branco através dos testes Box-Pierce e Ljung-Box, equações (2.43) e (2.44), no qual os coeficientes de autocorrelação amostrais dos erros ε_t devem ser estatisticamente iguais à zero, é necessário identificar se os mesmos apresentam distribuição normal. A identificação da normalidade do ε_t pode ser realizada através dos testes de Jarque-Bera e Shapiro-Wilk.

O teste Jarque-Bera verifica se os momentos da série estimada são iguais aos da normal. Sob essa hipótese, a assimetria é igual a zero e a curtose igual a três.

$$H_0: E(\varepsilon_t^s)^3 = 0 \wedge E(\varepsilon_t^s)^4 = 3$$

$$H_1: E(\varepsilon_t^s)^3 \neq 0 \wedge E(\varepsilon_t^s)^4 \neq 3$$

Em que a estatística do teste Jarque-Bera é dada por:

$$JB = \frac{n}{6} \left[\frac{\sum_{t=1}^n (\hat{\varepsilon}_t^s)^3}{n} \right]^2 + \frac{n}{24} \left[\frac{\sum_{t=1}^n (\hat{\varepsilon}_t^s)^4}{n} - 3 \right]^2 \quad (2.49)$$

Onde:

$$\hat{\varepsilon}_t^s = \text{resíduo padronizado, } \hat{\varepsilon}_t^s = (\hat{\varepsilon}_t - \bar{\varepsilon}_t) / \hat{\sigma};$$

n = tamanho da amostra;

A rejeição da hipótese nula indica não normalidade, porém a não rejeição não indica normalidade. A não rejeição indica que apenas o terceiro e quarto momentos da distribuição empírica coincidem com os da normal (BUENO, 2008).

Teste Shapiro-Wilk é baseado na estatística W . As hipóteses propostas foram:

H_0 : amostra provém de uma população normal;

H_1 : amostra não é de uma população normal.

De acordo com Shapiro e Wilk (1965), dada uma amostra aleatória y_1, y_2, \dots, y_n , de tamanho n , deve-se proceder da seguinte forma para calcular o valor de W :

i) Ordene a amostra, tal que $y_1 \leq y_2 \leq \dots \leq y_n$;

ii) Calcule a estatística $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2$;

iii) Calcule o valor de b , em que:

$$\text{Se } n=2k, \text{ então } b = \sum_{i=1}^k a_{n-i+1} (y_{n-i+1} - y_i)$$

$$\text{Se } n=2k+1, \text{ então } b = a_n (y_n - y_1) + \dots + a_{k+2} (y_{k+2} - y_k)$$

Em que i representa a observação atual, n é o tamanho da amostra, o coeficiente $\{a_{n-i+1}\}$ do teste W é dado pela tabela desenvolvida por Shapiro-Wilk, demonstrada abaixo para $n = 10$ e $i = 5$.

Quadro 3 - Coeficientes $\{a_{n-i+1}\}$ para o teste W de normalidade

i	n								
	2	3	4	5	6	7	8	9	10
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2	-	0,0000	0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3	-	-	-	0,0000	0,0875	0,1401	0,1743	0,1976	0,2141
4	-	-	-	-	-	0,0000	0,0561	0,0947	0,1224
5	-	-	-	-	-	-	-	0,0000	0,0399

Fonte: Shapiro e Wilk (1965)

- iv) Calcule $W = b^2 / S^2$, dado em (ii) e (iii);
- v) Tomar a decisão: rejeitar H_0 ao nível de significância α se $W_{\text{calculado}} < W_{\alpha}$
(valores críticos tabelados da estatística W de Shapiro-Wilk).

Como exemplo, considere uma amostragem de $\{6, 1, -4, 8, -2, 5, 0\}$.

- i) Para a amostra ordenada, obtém-se:

$$y_1 = -4, y_2 = -2, y_3 = 0, y_4 = 1, y_5 = 5, y_6 = 6, y_7 = 8.$$

ii) $S^2 = 146 - \left(\frac{1}{7}\right)(14^2) = S^2 = 118;$

- iii) Do quadro 3, para $n=7$, verifica:

$$a_7 = 0,6233, a_6 = 0,3031, a_5 = 0,1401, a_4 = 0,0000$$

$$b = 0,6233(8+4) + 0,3031(6+2) + 0,1401(5+0)$$

$$b = 10,6049$$

$$W = 0,9530$$

- iv) Para $\alpha = 0,50$ o valor crítico tabelado da estatística W de Shapiro-Wilk é de 0,928 e $W_{\text{calculado}} = 0,9530$. Como $W_{\text{calculado}} > W\alpha$, não existem evidências para rejeitar H_0 ao nível de significância de 0,50.

2.1.5 Modelos sazonais

Várias modificações são feitas no ARIMA para explicar o comportamento sazonal e não-estacionário das séries temporais, presente principalmente em séries econômicas. Devido a essas flutuações é apropriado introduzir polinômios auto-regressivos e de média móvel com defasagens sazonais (SHUMWAY, STOFFER, 2010).

O resultado do modelo auto-regressivo de média móvel sazonal, dito como ARMA (P, Q)_s, assume a seguinte forma:

$$\Phi_P(B^s)y_t = \Theta_Q(B^s)\varepsilon_t \quad (2.50)$$

Em que o operador auto-regressivo sazonal e operador de média móvel sazonal de ordens P e Q, respectivamente, com período sazonal s, é dado por:

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (2.51)$$

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \quad (2.52)$$

Abaixo, exemplo de um modelo ARMA (P, D) sazonal de primeira ordem.

$$(1 - \Phi B^{12})y_t = (1 + \Theta B^{12})\varepsilon_t \quad (2.53)$$

$$y_t = \Phi y_{t-12} + \varepsilon_t + \Theta \varepsilon_{t-12} \quad (2.54)$$

Em geral, é possível combinar operadores sazonal e não-sazonal em um modelo sazonal multiplicativo denotado por ARMA (p, q) (P, Q) através da equação:

$$\Phi_P(B^s)\phi(B)y_t = \Theta_Q(B^s)\theta(B)\varepsilon_t \quad (2.55)$$

Os modelos sazonais anteriores consideram as observações consecutivas não correlacionadas; a correlação só existe entre t e $t-s$, $t-2s$ A incorporação da

correlação entre instantes de tempo sucessivos resulta no modelo sazonal multiplicativo geral denominado ARIMA (p, d, q) (P, D, Q)_s (FAVA, 2000).

O ARIMA (p, d, q) (P, D, Q)_s é demonstrado por (2.56):

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d y_t = \delta + \Theta_Q(B^s)\theta(B)\varepsilon_t \quad (2.56)$$

Onde:

$$\nabla^d = (1 - B)^d$$

$$\nabla_s^D = (1 - B^s)^D$$

$$\Phi_P(B^s) = \text{equação (2.51)}$$

$$\Theta_Q(B^s) = \text{equação (2.52)}$$

$$\phi(B) = \text{equação (2.22)}$$

$$\theta(B) = \text{equação (2.29)}$$

2.1.6 Raiz unitária

Para Margarido e Anéfalos (1999), quando a série é estacionária os resultados da estatística tradicional são válidos, no entanto, quando a série apresenta raiz unitária há estimadores viesados, comprometendo a validade dos resultados. Por isso, a aplicação do teste de raiz unitária é importante para verificar se a integração foi suficiente para transformar a série não-estacionária em estacionária.

Segundo Brooks (2002), o teste Dickey-Fuller (DF) tem como objetivo básico examinar a hipótese nula de que $\varphi = 1$, contra a hipótese alternativa $\varphi < 1$, na equação abaixo:

$$\Delta y_t = \varphi y_{t-1} + \varepsilon_t \quad (2.57)$$

Portanto, as hipóteses de interesse são H₀: a série contém uma raiz unitária, contra hipótese alternativa H₁: a série é estacionária. Então, o teste DF verifica a

existência de uma raiz unitária em y_t , quando o processo gerador da série é expresso por um dos três modelos apresentados abaixo:

$$\Delta y_t = \varphi y_{t-1} + \varepsilon_t \quad (2.58)$$

$$\Delta y_t = \alpha + \varphi y_{t-1} + \varepsilon_t \quad (2.59)$$

$$\Delta y_t = \alpha + \beta t + \varphi y_{t-1} + \varepsilon_t \quad (2.60)$$

A equação (2.58) demonstra um processo estocástico com caminho aleatório puro. O modelo (2.59) adiciona um intercepto α e (2.60) inclui o intercepto α e uma tendência linear βt . O coeficiente φ estimado é dividido por seu erro-padrão para calcular a estatística τ de DF (τ , τ_μ e τ_τ , para as equações (2.58), (2.59) e (2.60) respectivamente). O valor obtido é comparado com o valor tabelado de DF para verificar se a hipótese nula $\varphi = 1$ é rejeitada, ou seja, se a série é estacionária. Hipóteses sobre a constante α e da tendência linear βt podem ser testadas por meio de testes individuais ($H_0: \alpha = 0$ e $H_0: \beta t = 0$) e de testes conjuntos $H_0: (\alpha, \beta t, \varphi) = (0, 0, 0)$.

Em síntese, os testes DF consistem em estimar as equações (2.58) a (2.60) por mínimos quadrados ordinários e comparar as estatísticas t resultantes, aos valores críticos gerados por Dickey-Fuller. Para as hipóteses conjuntas, a estatística do teste é construída a partir da soma dos quadrados dos resíduos das equações de regressão (FAVA, 2000).

Para Brooks (2002), os testes acima são válidos somente quando o termo de erro ε_t for um ruído branco não-correlacionado. Caso contrário, o verdadeiro tamanho do teste, ou seja, a proporção de vezes que a hipótese nula é incorretamente rejeitada é maior do que o tamanho nominal utilizado, geralmente 5%. A solução foi aumentar o teste, usando p defasagens das variáveis dependentes, resultando no teste Dickey-Fuller Aumentado (ADF). As equações (2.58), (2.59) e (2.60) são reescritas:

$$\Delta y_t = \varphi y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-1} + \varepsilon_t \quad (2.61)$$

$$\Delta y_t = \alpha + \varphi y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-1} + \varepsilon_t \quad (2.62)$$

$$\Delta y_t = \alpha + \beta_t + \varphi y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-1} + \varepsilon_t \quad (2.63)$$

2.2 Rede Neural Artificial (RNA)

Conforme Fausett (1994), as redes neurais artificiais têm sido desenvolvidas como generalizações de modelos matemáticos de cognição humana ou biologia neural, identificando a similaridade entre o comportamento de um neurônio biológico e um modelo de neurônio computacional. A motivação para o desenvolvimento da rede neural artificial ocorre em função do processamento de informações que ocorre nos neurônios biológicos.

Os neurônios biológicos são compostos por um corpo celular denominado soma e vários ramos. Esses ramos que conduzem informações em uma célula (estímulo) são os dendritos, e o ramo que conduz a informação para fora da célula (reação) é denominado axônio. Uma ativação de um neurônio, chamado potencial de ação, é transmitido a outros neurônios através de seu axônio, em primeira instância.

Para Kasabov (1998), a interação entre os neurônios que ocorre em pontos de contato entre a terminação axônica de um neurônio e o dendrito de outro denomina-se de sinapse. Em uma sinapse, duas partes podem ser distinguidas: uma membrana pré-sináptica, pertencente ao neurônio transmissor, e outra pós-sináptica, que pertence ao neurônio receptor. É através das sinapses que os nodos se unem funcionalmente, formando as redes neurais.

Dessa forma, a semelhança entre a RNA e o cérebro se dá em função de que o conhecimento adquirido ocorre através de um processo de aprendizado e as transmitâncias das conexões inter-neurônios, ou os pesos sinápticos, às quais estão submetidos os fluxos de informações através da rede, são utilizadas para armazenar o conhecimento obtido.

Segundo Portugal e Fernandes (1996), a rede neural formada por unidades de processamento, chamadas de nós, neurônios ou células, interconectadas por arcos

unidirecionais, denominadas de ligações, conexões ou sinapses, que tem capacidade de armazenar conhecimento experimental e torná-lo disponível para uso. Cada neurônio tem um estado interno chamado de ativação que é uma função das entradas que recebeu. O neurônio envia sua ativação como um sinal para vários outros neurônios (Fausett, 1994). Porém, um neurônio pode enviar apenas um sinal de cada vez, apesar de que o sinal é transmitido para vários outros neurônios .

A descrição matemática do modelo de neurônio artificial, proposto inicialmente por McCulloch e Pitts, como um circuito binário simples que combinava várias entradas e gerava um sinal de saída resultou em um modelo com n terminais de entrada x_1, x_2, \dots, x_n , simulando os dendritos do neurônio biológico e apenas um terminal de saída y , representando o axônio (BRAGA *et al.*, 2007).

O comportamento das sinapses é realizado através de terminais de entradas do neurônio que têm pesos acoplados w_1, w_2, \dots, w_n e o efeito de uma sinapse particular i no neurônio pós-sináptico resultavam em $x_i w_i$. Uma representação desse modelo é ilustrada pela figura 4.

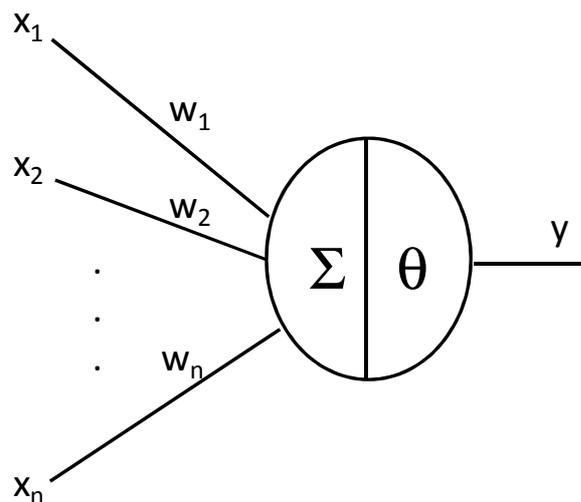


Figura 4 – Neurônio de McCulloch e Pitts

Fonte: Braga *et al.* (2007)

Porém, a rede desenvolvida por McCulloch e Pitts possui apenas uma camada e programa funções linearmente separáveis, além de ser proposta com pesos fixos, não ajustáveis, representando limitações ao modelo. A figura 5 ilustra uma generalização desse modelo, onde os pares de entrada são combinados através de uma função de soma ou produto para produzir um estado de ativação que produzirá a saída de ativação do neurônio.

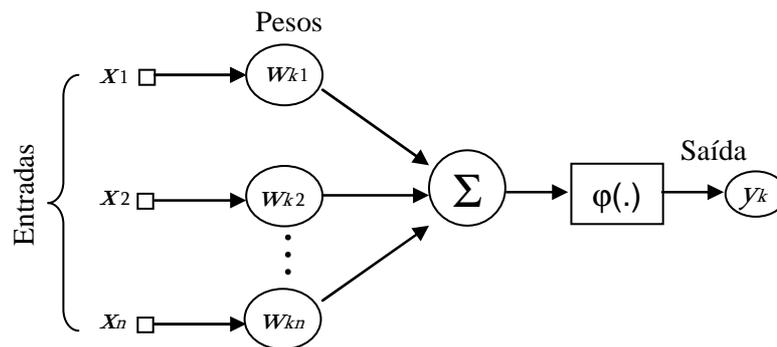


Figura 5 - Generalização do modelo McCulloch e Pitts

Fonte: Haykin (2001)

Um sinal x_j na entrada da sinapse j conectado ao neurônio k é multiplicado pelo peso sináptico w_{kj} . O primeiro índice do peso sináptico se refere ao neurônio em questão e o segundo indica o terminal de entrada da sinapse à qual o peso se refere. Um somador é utilizado para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio, enquanto a função de ativação restringe a amplitude da saída de um neurônio, sendo que o intervalo normalizado da amplitude da saída de um neurônio é escrito no intervalo $[0, 1]$ ou $[-1, +1]$. Os bias, representados por b_k , tem o objetivo de aumentar ou diminuir a entrada líquida da função de ativação (HAYKIN, 2001).

Formalmente, escreve-se um neurônio k da figura 5 através das equações:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.64)$$

$$y_k = \varphi(u_k + b_k) \quad (2.65)$$

Onde:

x_1, x_2, \dots, x_j = sinais de entrada;

$w_{k1}, w_{k2}, \dots, w_{km}$ = pesos sinápticos do neurônio k ;

u_k = saída do combinador linear devido aos sinais de entrada;

$\varphi(\cdot)$ = função de ativação;

b_k = polarização ou bias;

y_k = sinal de saída do neurônio

O uso do bias b_k aplica uma transformação afim à saída u_k do combinador linear no modelo da figura acima, identificado por:

$$v_k = u_k + b_k \quad (2.66)$$

Considerando v_k como potencial de ativação ou campo local induzido, a demonstração da polarização também pode ser realizada combinando as equações (2.64) e (2.66) onde se tem:

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (2.67)$$

$$y_k = \varphi(v_k) \quad (2.68)$$

Uma rede neural é caracterizada por sua função de ativação, o padrão de suas conexões entre os neurônios (arquitetura) e o método que determina os pesos das conexões, denominado treinamento ou aprendizagem. A seguir descrição dessas características.

2.2.1 Funções de ativação

A função de ativação $\varphi(v_k)$ define a saída de um neurônio em termos do campo do local induzido v . Essa função pode apresentar características lineares ou não lineares, sendo que essa última simula com maior precisão os neurônios biológicos. As principais funções lineares são a limiar (denominada na literatura de engenharia como

função de Heaviside, função degrau ou degrau unipolar) e a função linear por partes, enquanto que a principal função não-linear é a sigmoide (OLIVEIRA *et al.*, 2009).

A ativação da função limiar e a saída k do neurônio que emprega essa função são expressas como:

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad (2.69)$$

$$y^k = \begin{cases} 1 & \text{se } v_k \geq 0 \\ 0 & \text{se } v_k < 0 \end{cases} \quad (2.70)$$

Onde v_k é o campo do local induzido do neurônio, escrito por:

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad (2.71)$$

Para a função linear por partes, assumem-se duas formas específicas, a primeira considera que se a região linear de operação é mantida sem entrar em saturação, surge um combinado linear. A segunda indica que a função linear por partes se reduz a função de limiar, se o fator de amplificação da região linear é feito infinitamente grande (HAYKIN, 2001).

$$\varphi(v) = \begin{cases} 1, & v \geq +1/2 \\ v, & +1/2 > v > -1/2 \\ 0, & v \leq -1/2 \end{cases} \quad (2.72)$$

De acordo com Kasabov (1998), a função sigmóide exibe um balanceamento adequado entre o comportamento linear e não linear. As características gerais indicam que essa função é limitada, ou seja, seus valores estão restritos entre dois limites, por exemplo, $[0, 1]$, $[-1, 1]$, contínua e suave, portanto pode ser diferenciada em todo seu domínio e crescente.

O principal exemplo é a função logística, definida por:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (2.73)$$

Onde a é o parâmetro de inclinação da função sigmóide que na origem é igual a $a/4$.

2.2.2 Arquitetura

Para Kasabov (1998), os principais fatores responsáveis pela determinação do comportamento de um neurônio são a sua função de ativação e o padrão de conexão dos pesos dos quais ele envia e recebe sinais. Dentro de cada camada, os neurônios geralmente têm a mesma função de ativação e o mesmo padrão de conexões para outros neurônios. Esse arranjo de neurônios em camadas e os padrões de conexão dentro e entre as camadas representam a arquitetura da rede.

A arquitetura restringe o tipo de problema que pode ser tratado pela rede. RNA com uma camada única de nodos, como a desenvolvida por McCulloch e Pitts só conseguem resolver problemas linearmente separáveis, enquanto que as redes recorrentes são apropriadas para resolver problemas com processamento temporal (BRAGA *et al.*, 2007).

A definição da arquitetura de uma rede depende no número de camadas, classificados como camada única ou múltipla (as unidades de entrada não são contadas como camada, pois não realizam nenhum tipo de computação), o número de nós em cada camada, tipo de conexão entre os nodos e topologia da rede. Em geral, as arquiteturas são classificadas em três classes distintas:

- i) Redes realimentadas adiante com camada única;

Essa rede é estritamente do tipo alimentada adiante ou acíclica, ou seja, os neurônios estão organizados na forma de camadas, que na forma mais simples, possui uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de neurônios (HAYKIN, 2001). A figura 6 ilustra a rede acíclica com uma única camada de neurônios.

Também são definidas como *feedforward* já que os grafos não têm ciclos. Os neurônios que recebem sinais de excitação são chamados de camada de entrada, ou primeira camada. Neurônios que têm sua saída, como saída da rede, pertencem à camada de saída ou última camada (BARRETO, 2002).

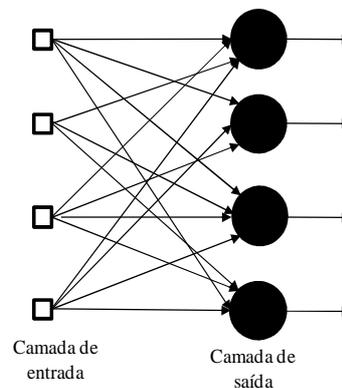


Figura 6 – Rede acíclica com única camada de neurônios

Fonte: Haykin (2001)

ii) Redes alimentadas diretamente com múltiplas camadas;

Conforme Haykin (2001), essas redes possuem uma ou mais camadas ocultas, cujos nós computacionais são chamados de neurônios ocultos ou unidades ocultas que tem como objetivo intervir entre a entrada externa e a saída da rede. Adicionando-se uma ou mais camadas ocultas, a rede é capaz de extrair estatísticas de ordem elevada. Os nós da camada de entrada fornecem os elementos do padrão de ativação (vetor de entrada), que formam os sinais de entrada aplicados aos neurônios (nós) na segunda camada. Os sinais de saída da segunda camada são utilizados como entradas para a terceira camada e assim por diante para o resto da rede.

A figura 7 indica um exemplo de rede 10-4-2, ou seja, 10 neurônios de fonte (entrada), 4 neurônios ocultos e 2 neurônios de saída. Essa rede é considerada totalmente conectada, no sentido de que cada um dos nós de uma camada de rede está conectado a todos os nós da camada adjacente seguinte.

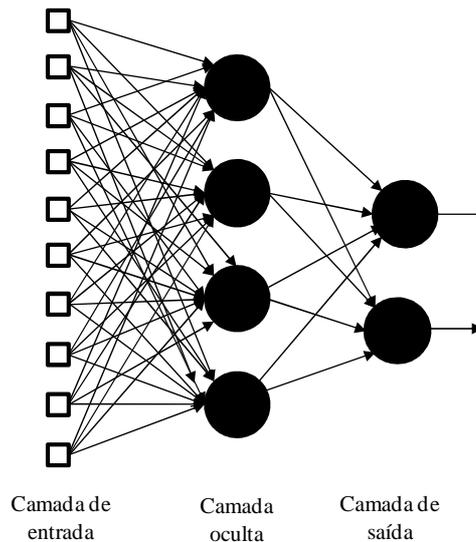


Figura 7 – Rede com múltiplas camadas

Fonte: Haykin (2001)

iii) Redes recorrentes;

Para Barreto (2002), redes recorrentes são aquelas que, além de apresentarem ciclos, envolvem neurônios dinâmicos. Os ciclos podem ser com realimentação, ou com retroação, ou com *feedback*. Nessa estrutura não existem laços de auto-realimentação na rede, isto é, uma situação onde a saída de um neurônio é realimentada para a sua própria entrada (HAYKIN, 2001).

2.2.3 Processos de aprendizado

Conforme Braga *et al.* (2007), na fase de aprendizado, a rede extrai informações relevantes sobre os padrões de informações apresentados para ela, criando a representação própria para o problema. Nesse processo, os parâmetros da RNA são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando, sendo o tipo específico de aprendizagem realizada pela maneira como ocorrem os ajustes realizados nos parâmetros.

Uma rede neural pode ser configurada de modo que a aplicação de um conjunto de fatores produz, diretamente ou através de um processo de relaxamento, o

conjunto de saídas desejadas. Existem vários métodos para definir os pontos fortes das conexões existentes, dentre elas definir os pesos de forma explícita, através de um conhecimento a priori ou treinar a rede, alimentando os padrões de ensino e deixá-lo mudar conforme alguma regra de aprendizagem.

Segundo Haykin (2001), essas regras definidas para a solução de um problema de aprendizado é denominada de algoritmo de aprendizagem. É importante observar que existem vários algoritmos de aprendizagem que diferem entre si através do mecanismo de como é ajustado o peso sináptico de um neurônio e a maneira pela qual a rede neural se relaciona o ambiente no qual ela opera.

A rede neural aprende acerca de seu ambiente através de um processo interativo de ajustes dos pesos sinápticos e níveis de bias, tornando-se mais instruída após cada iteração do processo de aprendizagem. Ela é estimulada por um ambiente, sofre modificações nos seus parâmetros livres como resultados desta estimulação e responde de uma maneira nova ao ambiente devido as modificações ocorridas na sua estrutura interna.

Existem cinco principais regras de aprendizado:

i) Aprendizagem por correção de erros;

Essa regra minimiza a diferença entre a saída calculada pela rede e a saída desejada, ou seja, o erro da resposta atual da rede neural, através da equação:

$$e_k(n) = d_k(n) - y_k(n) \quad (2.74)$$

Onde:

$e_k(n)$ = sinal de erro;

$y_k(n)$ = sinal de saída do neurônio k ;

$d_k(n)$ = resposta desejada ou saída-alvo;

O sinal de erro $e_k(n)$ aciona um mecanismo de controle que tem como objetivo ajustar os pesos sinápticos do neurônio k através de correções que aproximam o sinal de saída $y_k(n)$ com a resposta desejada $d_k(n)$. Esse propósito é obtido minimizando uma função de custo ou índice desempenho $\xi(n)$, definida por:

$$\xi(n) = \frac{1}{2} e_k^2(n) \quad (2.75)$$

A minimização da função de custo $\xi(n)$ resulta na regra de aprendizagem denominada delta ou Widrow-Hoff. Para $w_{kj}(n)$ que representa o valor do peso sináptico w_{kj} do neurônio k excitado por um elemento $x_j(n)$ do vetor de sinal $\mathbf{x}(n)$ no tempo n , o ajuste $\Delta w_{kj}(n)$ aplicado ao peso sináptico é representado por:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \quad (2.76)$$

Em que η é uma constante positiva denominada taxa de aprendizado. Portanto, de acordo com a equação acima, a regra delta determina que o ajuste feito em um peso sináptico de um neurônio é proporcional ao produto do sinal de erro ao sinal de entrada da sinapse em questão (HAYKIN, 2001).

ii) Aprendizagem baseada em memória;

As experiências passadas são armazenadas em uma memória de exemplos de entrada-saídas classificadas como $\{(x_i, d_i)\}_{i=1}^N$, em que x_i representa o vetor de entrada e d_i é a resposta desejada correspondente.

iii) Aprendizagem hebbiana;

Hebb propôs uma mudança no aprendizado associativo (nível celular) de forma que se neurônio pré-sináptico apresentar grande influência na ativação do neurônio pós-sináptico, a conexão entre eles deve ser reforçada. Portanto, a regra de Hebb postula que se os dois neurônios de uma sinapse (conexão) são ativados simultaneamente, ou seja,

sincronamente, então a força desta sinapse é aumentada. Caso contrário, se eles forem ativados assincronamente, a sinapse é enfraquecida ou eliminada.

iv) Aprendizagem competitiva;

Os neurônios de saída de uma rede competem entre si para se tornar ativos. Esse aprendizado é adequado para classificar um conjunto de padrões de entrada. As principais características são a presença de um conjunto de neurônios idênticos entre si, exceto por alguns pesos sinápticos distribuídos aleatoriamente. Existe um limite imposto sobre a força de cada neurônio, além de um mecanismo que permite a competição para resposta a um dado conjunto de entradas, de forma que somente um neurônio de saída esteja ativo em um determinado instante.

v) Aprendizagem de Boltzmann

Os neurônios formam uma estrutura recorrente e operam de maneira binária, visto que eles estão em estado “ligado” ou em um estado “desligado”.

2.2.3.1 Situações de aprendizado

As situações de aprendizado são caracterizados por três tipos distintos: supervisionado, com a presença de um supervisor ou professor externo, não supervisionado, com ausência desse professor externo e o aprendizado por reforço que combina as duas características anteriores.

No aprendizado supervisionado a entrada e saída desejadas para a rede são fornecidas por um supervisor externo com objetivo de ajustar os parâmetros da rede, de forma a encontrar alguma ligação entre os pares de entradas e saídas fornecidos (Braga *et al.*, 2007). A rede tem sua saída corrente comparada com a saída desejada, recebendo informações do supervisor sobre o erro da resposta atual. A cada padrão de entrada submetido à rede compara-se a resposta desejada que representa uma ação ótima para

ser realizada pela rede com a resposta calculada, ajustando os pesos das conexões para minimizar o erro.

A desvantagem do aprendizado supervisionado é que na ausência do professor, a rede não aprende novas estratégias para situações que não foram apresentadas pelos exemplos do treinamento da rede. Os exemplos mais conhecidos para esse tipo de aprendizado são regra delta e sua generalização para redes de múltiplas camadas denominada algoritmo de *backpropagation*.

No aprendizado não supervisionado a partir do momento em que a rede estabelece harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se uma habilidade de formar representações internas para codificar características de entrada e criar novas classes ou grupos automaticamente. Isso só se torna possível quando existe redundância de dados de entrada, visto que sem redundância seria impossível encontrar quaisquer padrões ou características de entrada de dados de entrada.

Segundo Kasabov (1998), o aprendizado por reforço é baseado na combinação do aprendizado supervisionado e não supervisionado. O algoritmo de aprendizado apresenta o vetor de entrada á rede e a analisa o vetor de saída gerado pela rede. Se esse vetor apresenta resultados satisfatórios, os pesos sinápticos são aumentados. Caso contrário, os pesos sofrem um decréscimo.

2.2.4 Redes neurais recorrentes

A rede neural recorrente é uma adaptação do padrão de rede *feedforward* para permitir um modelo de dados sequenciais. Em cada iteração, a rede recebe uma entrada, atualiza sua camada oculta e faz uma previsão (figura 8). A presença de realimentação de informações permite as representações internas e dispositivos de memória que são capazes de processar e armazenar informações temporais e sinais sequenciais. Porém, a

presença de conexões recorrentes ou realimentação de informação conduz a comportamentos complexos, mesmo com um número reduzido de parâmetros (ZUBEN, 1996).

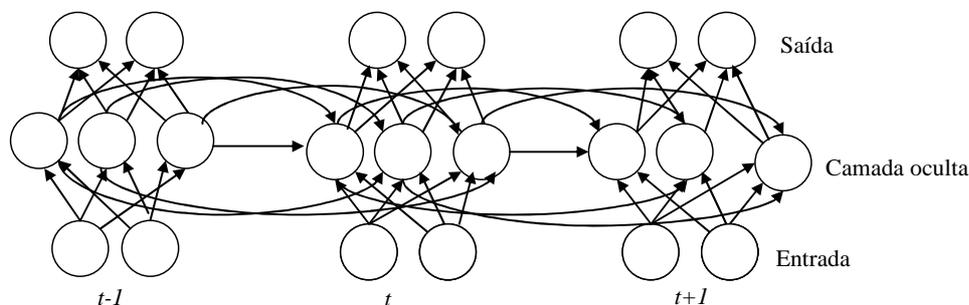


Figura 8 – Rede neural recorrente

Fonte: Sutskever *et al.* (2011)

Essas redes possuem um ou mais laços de realimentação que podem ser de natureza local ou global. A realimentação permite que essas redes adquiram representações do estado, transformando-as em dispositivos apropriados para diversas aplicações como previsão não-linear e modelagem, processamento de voz, controle de instalações industriais, entre outros.

Haykin (2001) indica que as principais características de arquiteturas das redes recorrentes é que todas elas incorporam um perceptron de múltiplas camadas estático ou parte dele e exploram a capacidade de mapeamento não linear do perceptron de múltiplas camadas. Existem quatro principais arquiteturas de redes pertencentes às redes recorrentes com realimentação global:

- i) Modelo recorrente de entrada-saída, que são redes auto-regressivas não lineares, com entradas exógenas usando realimentação da camada de saída para a camada de entrada;

- ii) Modelo de espaço de estados que são redes totalmente conectadas com realimentação da camada oculta para a camada de entrada. Os neurônios ocultos definem o estado da rede, como a rede de Elman;
- iii) Perceptron de múltiplas camadas recorrentes com mais de uma camada oculta, com realimentação da saída de camada computacional para a sua própria entrada;
- iv) Rede de segunda ordem usando neurônios de segunda ordem.

2.2.4.1 Rede recorrente de Elman

Para Braga *et al.* (2007), a rede Elman é caracterizada como uma rede recorrente simples e a realimentação capacita à realização de tarefas que se estendem no tempo. As unidades intermediárias podem ter funções de ativação lineares ou não lineares, e as unidades de contexto são usadas apenas para memorizar as ativações anteriores das unidades intermediárias e podem ser consideradas como atraso no tempo em um passo. As conexões *feedforward* são modificáveis e as conexões recorrentes são fixas, motivo pelo qual a rede de Elman é apenas parcialmente recorrente.

Essa rede possui conexões recorrentes dos neurônios ocultos para uma camada de unidades de contexto que consistem em atrasos unitários. Essas unidades de contexto armazenam as saídas dos neurônios ocultos por um passo de tempo, e então realimenta de volta para a camada de entrada. Os neurônios ocultos têm um registro de suas ativações passadas, o que capacita a rede a realizar tarefas de aprendizado que se estendem no tempo. (HAYKIN, 2001, p. 790).

Segundo Elman (1990), a realimentação ocorre da saída de cada neurônio da camada oculta para todos os neurônios da mesma camada. A camada de contexto, também oculta, composta por unidades de contexto simula a memória da rede. O processamento da rede consiste nos eventos: no instante t (inicial), o sinal é propagado pela rede e as unidades de contexto, inicializadas com a saída da camada oculta com o

valor zero, não influenciarão na saída da rede, ou seja, na primeira iteração a rede se comportará como uma rede *feedforward*.

Nessa primeira iteração os neurônios ocultos ativarão os neurônios da camada de contexto e esses armazenarão a saída desta iteração que será utilizada no próximo ciclo. O algoritmo *backpropagation* é aplicado para a correção dos pesos sinápticos, com exceção as sinapses recorrentes que são fixas em 1. No instante $t+1$ o processo é repetido. Agora os neurônios ocultos serão ativados pelas unidades de entrada e pelas unidades de contexto que possuem o valor de saída dos neurônios ocultos no instante t . A figura 9 descreve esse processo.

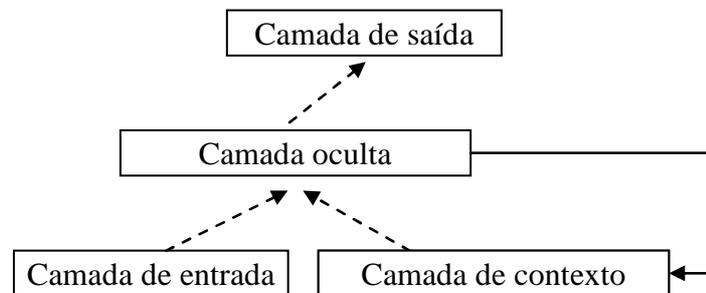


Figura 9 - Rede genérica de Elman

Fonte: Elman (1990)

A camada de entrada possui neurônios sensoriais que recebem um sinal externo e o propagam sem alterá-lo. A camada de saída possui neurônios lineares em que as saídas são somas ponderadas de seus respectivos sinais de entrada. A camada intermediária tem neurônios que apresentam funções de ativação lineares ou não-lineares e a camada de contexto é formada por neurônios que apenas memorizam as ativações dos neurônios da camada intermediária, atuando com operadores de defasagens em um instante de tempo (ZUBEN, 1996).

De acordo com Oliveira *et al.* (2009), a rede Elman é aplicada em situações onde se deseja que a rede utilize as entradas precedentes, considerando-as com a entrada

atual para gerar uma resposta esperada, permitindo que a rede execute a aprendizagem das tarefas que se estendam no tempo.

A arquitetura dessa rede permite que os neurônios não dependam apenas dos valores de entrada, mas também desses valores defasados no tempo. Esse tipo de rede é semelhante ao modelo média móvel (MA) das séries temporais McNelis (2005). Representa uma ferramenta para capturar a memória de séries temporais com grande variabilidade ao longo de um período e o uso de defasagens é uma maneira para capturar a memória em séries temporais, porém a memória também pode apresentar uma estrutura não linear devido a mudanças ocorridas através do tempo. Portanto, a rede Elman, na qual os neurônios defasados realimentam os neurônios atuais é um mecanismo útil para modelar este tipo de estrutura de memória.

A figura 10 apresenta a arquitetura detalhada da rede Elman.

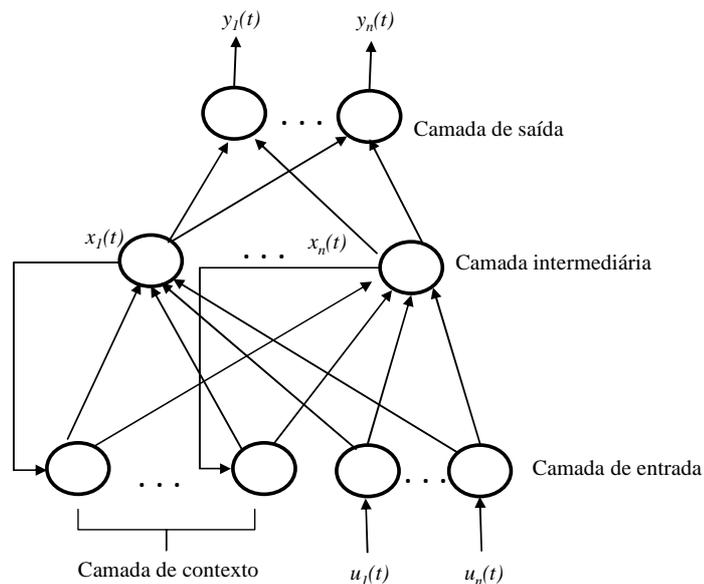


Figura 10 – Arquitetura detalhada da rede Elman

Fonte: Zuben (1996)

3 METODOLOGIA

3.1 Caracterização da Pesquisa

Esta pesquisa representou um procedimento formal, com tratamento científico específico que teve como objetivo apresentar respostas ao problema proposto. O trabalho foi desenvolvido com base em conhecimentos disponíveis e através da utilização cuidadosa de métodos, técnicas e outros procedimentos científicos que envolveram várias fases, desde a formulação do problema até a apresentação de resultados.

Do ponto de vista da natureza, foi uma pesquisa aplicada, pois produziu conhecimentos para aplicação dirigida na solução de um problema específico (Marconi e Lakatos, 2009). Neste trabalho em particular realizou-se uma análise comparativa entre métodos estatísticos (Holt-Winters e ARIMA) e inteligência artificial (rede neural de Elman) aplicado ao problema da previsão do preço do etanol no estado de Goiás, representando uma importante ferramenta de gestão, pois reduz o risco na tomada de decisão dos agentes econômicos envolvidos.

Ao final, a decisão para selecionar um método específico considerou os erros produzidos pelos modelos, dentre eles o erro médio (ME), a média absoluta dos erros (MAE), média absoluta percentual dos erros (MAPE) e a raiz do erro médio quadrático (RMSE).

Em relação aos objetivos gerais, tratou-se de uma pesquisa exploratória já que teve como finalidade desenvolver e esclarecer conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores, proporcionando uma visão geral, de tipo aproximativo, acerca de determinado fato (GIL, 2002).

Foi realizada pesquisa bibliográfica com análise da literatura publicada sobre os métodos de suavização exponencial, Box-Jenkins e redes neurais, traçando um quadro teórico que permitiu a estruturação conceitual que sustentou o desenvolvimento da pesquisa, mostrando quais as lacunas existentes e onde se encontram os principais entraves teóricos ou metodológicos.

3.2 Dados

Os preços do etanol em Goiás foram coletados no sítio eletrônico da Agência Nacional do Petróleo (ANP, 2012), que em cumprimento ao artigo 8º da Lei 9.478/1997 (Lei do Petróleo) tem o papel de implementar, na esfera de suas atribuições, a política nacional de petróleo e gás natural, com ênfase na proteção dos interesses dos consumidores quanto a preço, qualidade e oferta dos produtos.

A ANP acompanha o comportamento dos preços praticados pelas distribuidoras e pelos postos revendedores de combustíveis com pesquisas semanais dos preços da gasolina, etanol hidratado, óleo diesel, gás natural veicular e gás liquefeito de petróleo, coletados em 555 localidades, aproximadamente 10% dos municípios brasileiros, de acordo com os procedimentos estabelecidos pela Portaria ANP nº 202 de 2002⁴, através de empresas ou consórcio contratado pela ANP.

Em síntese, ocorre a coleta semanal dos preços de venda ao consumidor final e os correspondentes preços de aquisição por parte dos agentes econômicos selecionados para integrar a amostra definida pela ANP. A coleta de preços em Goiás é realizada em 19 municípios. O quadro 4 mostra a relação das localidades pesquisadas em Goiás e o tamanho da amostra extraída nas coletas dos dados.

⁴ Regulamenta os procedimentos para levantamento de preços e margens de comercialização de combustíveis praticados em estabelecimentos de agentes econômicos autorizados pela ANP.

Quadro 4 – Coleta de dados da ANP em Goiás

Localidade	Postos revendedores	Tamanho da amostra	
		Número	Percentual
Águas Lindas de Goiás	9	5	55,56%
Anápolis	65	16	24,62%
Aparecida de Goiânia	52	13	25,00%
Caldas Novas	21	12	57,14%
Catalão	22	10	45,45%
Formosa	22	13	59,09%
Goiânia	254	65	25,59%
Goiatuba	14	10	71,43%
Itumbiara	25	18	72,00%
Jataí	23	16	69,57%
Luziânia	24	13	54,17%
Mineiros	10	7	70,00%
Morrinhos	17	11	64,71%
Planaltina de Goiás	5	5	100,00%
Porangatu	11	7	63,64%
Rio Verde	35	18	51,43%
Trindade	12	7	58,33%
Valparaíso de Goiás	15	10	66,67%

Fonte: ANP (2012)

Portanto, a amostra extraída do sítio da ANP se refere aos preços do etanol em Goiás (R\$/litro) ao consumidor final, no período de janeiro de 2005 até dezembro de 2011, totalizando 84 observações mensais. Optou-se por trabalhar com os preços

medianos, já que essa medida de tendência central não é afetada pelos valores extremos. Posteriormente, houve a transformação logarítmica da amostragem com o objetivo de estabilizar a variância para os métodos estatísticos e normalização dos dados para a rede de Elman.

3.3 Programa R e Modelos de Previsões

3.3.1 Programa R

O uso de *softwares* estatísticos é de grande importância para análise e interpretação dos resultados, porém eles apresentam um custo de aquisição elevado. Alternativamente, tem-se os softwares de domínio público como o R que é uma linguagem orientada a objetos, disponível como Software Livre sob os termos da *GNU da Free Software Foundation General Public License* em forma de código de fonte aberto. Esse código pode ser modificado ou implementado com novos procedimentos desenvolvido por qualquer usuário a qualquer momento. Ele compila e roda em diversas plataformas como UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS (R DEVELOPMENT CORE TEAM, 2010).

O R tem vários pacotes com conjuntos de funções que permitem a realização de análises estatísticas, além apresentar ajuda para suas funções e demonstrações de execução. Dentre eles, cita-se o pacote *Forecast* que tem métodos e ferramentas para exibir e analisar séries temporais univariadas, incluindo a suavização exponencial através da modelagem do espaço e estados e a seleção automática dos parâmetros do ARIMA. Hyndman e Khandakar (2008) descreveram esse pacote.

Outro pacote importante neste trabalho foi o *RSNNS* desenvolvido por Bergmeir e Benitez (2010), baseado no simulador de rede neural *SNNS (Stuttgart Neural Network Simulator)*. O *RSNNS* envolve a funcionalidade do *SNNS* para torná-lo

disponível a partir do R, permitindo a integração de diversas topologias e algoritmos de treinamento para o desenvolvimento das redes neurais.

3.3.2 Suavização exponencial

As fórmulas recursivas para realizar as previsões dos principais métodos de suavização exponencial estão indicados no quadro 5, com modelagem no espaço de estados onde a sazonalidade, tendência e os erros podem ser modelados de forma aditiva ou multiplicativa. As denominações A e Ad indicam um comportamento aditivo e outro aditivo suavizado, respectivamente e M e Md um comportamento multiplicativo e multiplicativo suavizado da série.

Quadro 5 – Fórmulas recursivas para previsão dos modelos de suavização exponencial

Tendên- -cia	Sazonalidade	
	A	M
A	$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.77)</p>	$\begin{aligned} \ell_t &= \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t / \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.78)</p>
Ad	$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.79)</p>	$\begin{aligned} \ell_t &= \alpha(y_t / s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1}) \\ b_t &= \beta(\ell_t - \ell_{t-1}) + (1-\beta)\phi b_{t-1} \\ s_t &= \gamma(y_t / \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.80)</p>
M	$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1} \\ b_t &= \beta(\ell_t / \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.81)</p>	$\begin{aligned} \ell_t &= \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1} \\ b_t &= \beta(\ell_t / \ell_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(y_t / \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.82)</p>
Md	$\begin{aligned} \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^\phi \\ b_t &= \beta(\ell_t / \ell_{t-1}) + (1-\beta)b_{t-1}^\phi \\ s_t &= \gamma(y_t - \ell_{t-1}b_{t-1}^\phi) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.83)</p>	$\begin{aligned} \ell_t &= \alpha(y_t / s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^\phi \\ b_t &= \beta(\ell_t / \ell_{t-1}) + (1-\beta)b_{t-1}^\phi \\ s_t &= \gamma(y_t / \ell_{t-1}b_{t-1}^\phi) + (1-\gamma)s_{t-m} \end{aligned}$ <p style="text-align: center;">(2.84)</p>

Fonte: Hyndman e Khandakar (2008)

Onde:

l_t = elemento de nível para o período t ;

b_t = elemento de tendência para o período t ;

s_t = elemento de sazonalidade para o período t ;

y_t = valor observado no período t ;

s = intervalo de tempo da sazonalidade;

m = número do período previsto;

α = constante do ajuste de nível;

β = constante do ajuste de tendência;

γ = constante do ajuste de sazonalidade;

ϕ = constante de amortecimento.

A estimação dos parâmetros $\theta=(\alpha,\beta,\gamma,\phi)$ ocorre através da função de verossimilhança e os valores estão restritos entre 0 e 1. O algoritmo do R determina o número de parâmetros dos modelos de suavização de acordo com o critério de AIC. Esse critério de seleção é preferível se comparado com outras medidas de previsão de erro como o MSE ou MAE, pois estes penalizam os modelos com muitos parâmetros. Além disso, o AIC fornece um método para selecionar entre o erro multiplicativo e aditivo, pois se baseia na probabilidade e não na previsão um passo a frente.

O R possui uma cadeia de três caracteres representados pela nomenclatura (ETS) que indicam qual o método de suavização exponencial será utilizado. A primeira letra é o componente de erro (A, M ou Z), a segunda se refere ao tipo de tendência (N, A, M ou Z) e a terceira letra representa o tipo de sazonalidade (N, A, M ou Z). Em todos os casos, “A” é o termo aditivo, “M” equivale ao termo multiplicativo, “N” representa nenhuma característica e “Z” significa selecionado automaticamente.

3.3.3 Modelo ARIMA

A identificação do modelo Box-Jenkins segue um ciclo iterativo realizado a partir de análises de funções de autocorrelação e autocorrelações parciais, seguido da fase de estimação e testes. A validação ocorre a partir da análise dos resíduos e o mesmo será considerado admissível conforme medidas de desempenho através do erro. A fase final consiste na utilização do modelo estimado para realizar previsões.

Neste trabalho serão aplicadas as três fases da metodologia Box-Jenkins, ilustrado pela figura 11.

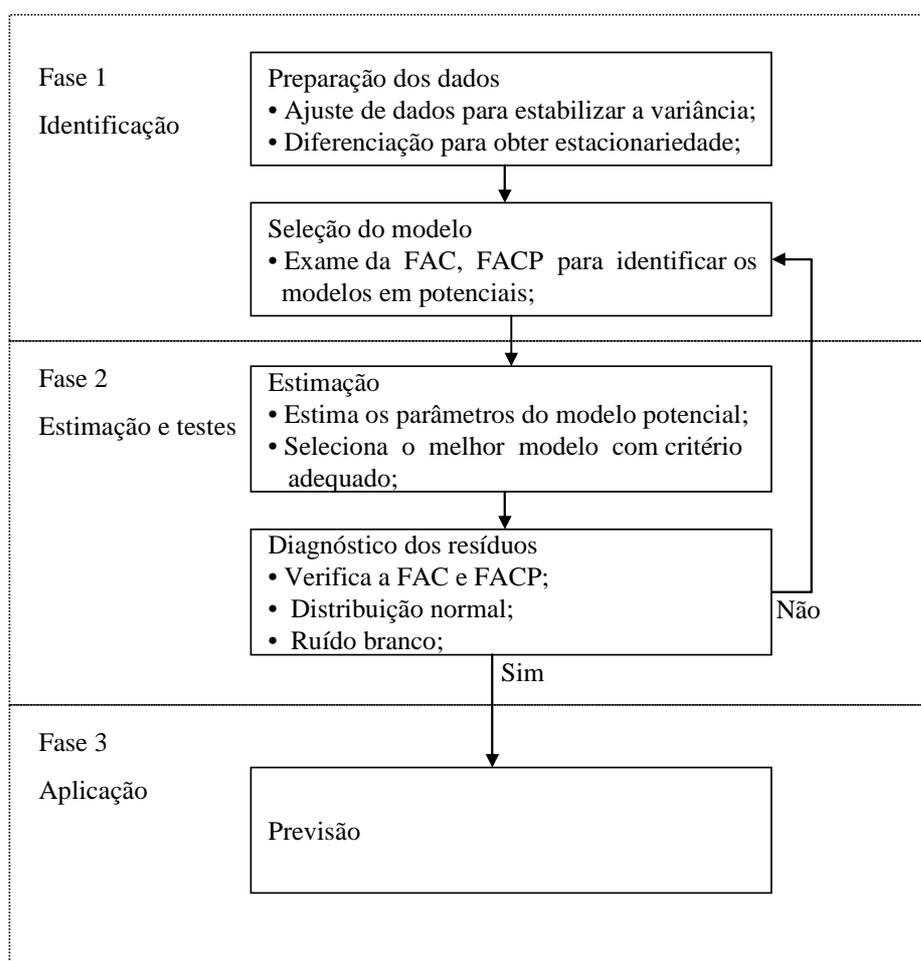


Figura 11– Representação esquemática do algoritmo Box Jenkins

Fonte: Flores (2009)

Para Hyndman e Khandakar (2008), um dos obstáculos no uso ARIMA é o processo de seleção de ordem dos parâmetros (p,d,q) que é considerado subjetivo e

difícil de se verificar. Ocorreram várias tentativas para automatizar a modelagem ARIMA nos últimos 25 anos, dentre eles citam-se os métodos propostos por Hannan e Rissanen, G'omez e posteriormente Liu.

Hannan e Rissanen (1983), propuseram um método para identificar a ordem de um modelo ARMA para uma série estacionária através do ajustamento de um modelo auto-regressivo, então a probabilidade de possíveis modelos é calculado através de uma série de regressões padrão. Eles estabeleceram as propriedades assintóticas do procedimento em condições muito gerais.

Enquanto Liu, em 1989, propôs um método para identificação de modelos ARIMA sazonais usando um método de filtragem e regras heurísticas através de algoritmo usado pelo software SCA-Expert. Posteriormente, G'omez (1998), estendeu o método de identificação de Hannan-Rissanen para incluir identificação do modelo ARIMA sazonal multiplicativo. Eles aplicaram este processo de identificação automática no software TRAMO e SEATS. Para uma determinada série, o algoritmo tenta encontrar o modelo com o BIC mínimo.

A automatização da modelagem ARIMA no R tem como objetivo selecionar um modelo adequado com a escolha dos parâmetros. Para dados não-sazonais, considera-se ARIMA (p, d, q) onde d, ou seja, a quantidade de vezes que a série deve ser diferenciada até se tornar estacionária, é selecionada com base em sucessivos testes KPSS de raiz unitária. Depois da identificação de d, seleciona-se as ordens p e q através do critério de AIC.

Para os dados sazonais, considera-se ARIMA (p, d, q) (P, D, Q), onde se aplica o teste Canova-Hansen para verificar se a mudança no padrão sazonal ao longo do tempo é suficiente para justificar uma raiz unitária sazonal, ou se um padrão estável de

sazonalidade pode ser modelado através de variáveis dummy. Se d e D são conhecidos, seleciona-se as ordens p , q , P , Q com base no critério de AIC.

Por exemplo, para dados sazonais, considera-se $ARIMA(p, d, q)(P, D, Q)_m$ onde p e q podem assumir valores de 0 a 3 e P e Q variam de 0 a 1 e m equivale ao tamanho da sazonalidade. Quando $c = 0$ (constante, com $\mu = 0$), existem 288 modelos prováveis e se $c \neq 0$, tem-se 192 possibilidades, totalizando 480 modelos. Portanto, é necessário um algoritmo eficiente para percorrer esse espaço de busca, a fim de obter um modelo com menor AIC.

O algoritmo implementado no R define quatro modelos possíveis:

- i) $ARIMA(2,d,2)$ se $m = 1$ e $ARIMA(2, d, 2)(1,D,1)$, se $m > 1$;
- ii) $ARIMA(0,d,0)$ se $m = 1$ e $ARIMA(0, d, 0)(0,D,0)$, se $m > 1$;
- iii) $ARIMA(1,d,0)$ se $m = 1$ e $ARIMA(1, d, 0)(1,D,0)$, se $m > 1$;
- iv) $ARIMA(0,d,1)$ se $m = 1$ e $ARIMA(0, d, 1)(0,D,1)$, se $m > 1$

Se $d + D \leq 1$, considera-se $c \neq 0$, caso contrário, $c = 0$. Dos quatro modelos acima, escolhe-se aquele com menor valor de AIC e ele passa a ser chamado de “modelo atual” denotado por $ARIMA(p, d, q)$ se $m = 1$ ou $ARIMA(p, d, q)(P, D, Q)_m$ se $m > 1$.

Na segunda etapa, considera-se até treze variações do modelo atual. Sempre que um modelo com menor AIC é encontrado ele se torna o modelo atual. Esse processo é repetido e termina quando não se identifica um modelo próximo ao atual com menor AIC.

3.3.4 Redes neurais

A rede neural foi implementada pelo pacote *RSNNS* do R (Bergmeier e Benitez, 2010), através das etapas abaixo:

- a) Organização da base de dados com as informações relevantes;

- b) Separação dessas informações em dois grupos, o primeiro destinando à fase de aprendizagem da rede, enquanto o segundo será utilizado na fase de teste;
- c) Transformar a informação primária para servir de entrada à rede;

Antes de iniciar o processo de treinamento foi realizada a normalização dos dados primários com objetivo de aumentar o desempenho das RNAs. A equação (2.85) foi aplicada para realizar a padronização dos dados.

$$y_t^* = \frac{(y_t - \bar{y})}{\sigma_y} \quad (2.85)$$

Em que:

y_t^* = valor normalizado;

y_t = valor real ou observado;

σ_y = desvio padrão da amostra;

\bar{y} = média amostral;

- d) Selecionar, treinar e testar a rede: esta fase, bem como a eventual reformulação das anteriores, deverá ser repetida até que os resultados sejam considerados adequados;

Os parâmetros referentes à taxa de aprendizado, os pesos iniciais e o número de ciclos (iterações) foram modificados ao longo do treinamento com objetivo de identificar um modelo com um resultado mais satisfatório. Para realizar essas simulações foi criado um *script* no R que permitiu alterar esses diferentes parâmetros.

Como a taxa de aprendizado tem grande influência durante o processo de treinamento da rede neural optou-se por variá-la entre 0,1 e 1,0, pois se a taxa for muito baixa, o aprendizado da rede será muito lento, porém se essa taxa for muito alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado.

Para Montana e Davis (1989), a aplicação das redes neurais tem sido dificultada pela falta de um algoritmo de treinamento que encontre um conjunto global de pesos ideais em um tempo relativamente curto. Esse treinamento representa um problema de otimização em que o espaço de busca é o espaço de pesos da rede e a função objetivo é uma função do desempenho obtido pela rede para um determinado conjunto de padrões.

Os pesos iniciais das redes neurais foram determinados aleatoriamente com distribuição uniforme entre -1,0 até +1,0. Essa distribuição de probabilidade reflete observações empíricas com buscas de soluções ótimas que tendem a conter pesos com pequenos valores reais, embora os pesos também possam ser arbitrados com grandes valores reais.

O número de ciclos (iterações) foi considerado como critério de parada do treinamento. Observando que um número excessivo de ciclos pode fazer com que a rede diminua sua capacidade de generalização (*overfitting*) e que um pequeno número de ciclos a rede pode não atingir seu melhor desempenho (*underfitting*).

- e) Aplicar à rede neural desenvolvida para o problema proposto.

3.3.5 Medidas de erros

O desempenho de um modelo se relaciona com a diferença entre os valores observados e previstos, ou seja, as medidas de erros. A escolha de uma medida específica depende se o objetivo é selecionar um método mais preciso para previsão ou calibrar um determinado modelo.

Como o objeto de estudo é a comparação entre métodos de previsão, optou-se por utilizar quatro medidas específicas: ME (erro médio), MAE (média absoluta dos erros), RMSE (raiz do erro quadrático médio) e MAPE (média absoluta percentual dos erros).

a) Erro médio (*mean error*)

$$ME = \frac{1}{n} \sum_{t=1}^n e_t \quad (2.86)$$

b) Raiz do erro quadrático médio (*root mean square error*)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2.87)$$

c) Média absoluta dos erros (*mean absolute error*)

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (2.88)$$

d) Média absoluta percentual dos erros (*mean average percentage error*)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right| \times 100 \quad (2.89)$$

O termo de erro e_t equivale ao valor observado menos o estimado, ou seja,

$$e_t = y - \hat{y}.$$

4 RESULTADOS E DISCUSSÃO

Uma análise cuidadosa da figura 12 permitiu destacar algumas características evidenciadas pela série de preços do etanol no estado de Goiás, sugerindo o tipo de modelação mais adequado ao seu estudo e previsão. Um diagnóstico preliminar desses dados indicou a existência de um padrão sazonal dos preços. Notou-se ainda a existência de algumas oscilações periódicas ao longo dos vários anos analisados.

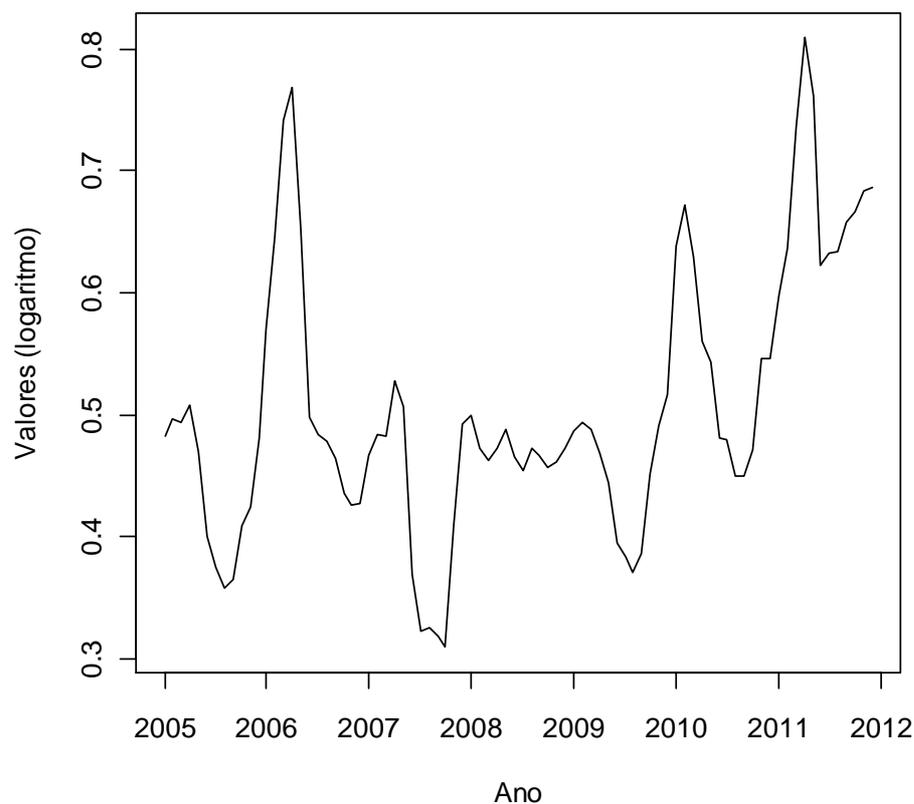


Figura 12 – Evolução do preço do etanol em Goiás

Através do *boxplot* (figura 13), os dados foram agrupados mensalmente, percebendo-se que a variabilidade em alguns meses é bem maior, além de uma tendência de queda no preço no último trimestre do ano. Identificou-se a presença de *outliers* nos meses de setembro, outubro, novembro e dezembro.

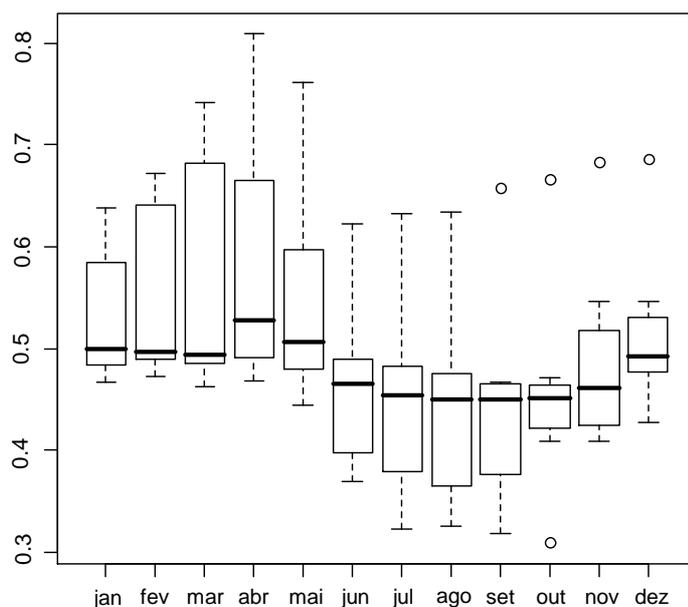


Figura 13 – Boxplot do preço do etanol

A tabela 1 mostra as principais medidas descritivas obtidas pela série de preços após transformação logarítmica, possibilitando uma visão geral do comportamento dessa variável ao longo do tempo. Percebe-se uma grande variabilidade nos dados ao longo do período avaliado, com coeficiente de variação de 21,79%. Os valores de 0,4496 e 0,5495 deixam 25% e 75% da amostragem abaixo deles, respectivamente, enquanto o valor mediano de 0,4815 indica que 50% das observações estão abaixo desse valor e, por consequência, 50% acima dele.

Tabela 1 – Medidas descritivas da série logarítmica de preços

Parâmetros	Média	Mediana	Desvio-padrão	Mínimo	Máximo	Quartil (1°)	Quartil (3°)	Coeficiente de variação
Valores	0,506	0,4815	0,110	0,3090	0,8092	0,4496	0,5495	21,79%

Através da análise gráfica da decomposição da série de preços, nota-se claramente um comportamento sazonal homogêneo e estável em todos os anos, o que já era esperado, tendo em vista a sazonalidade da produção da cana-de-açúcar, que implica

no preço dos produtos sucroalcooleiros. Também existe uma tendência de crescimento justificada, dentre outros fatores, pelo aumento da frota de veículos *flexfuel* e consequente elevação da demanda pelo produto, o que ao final pressiona os níveis de preços. A decomposição da série de preços evidencia a presença de tendência e sazonalidade conforme figura 14

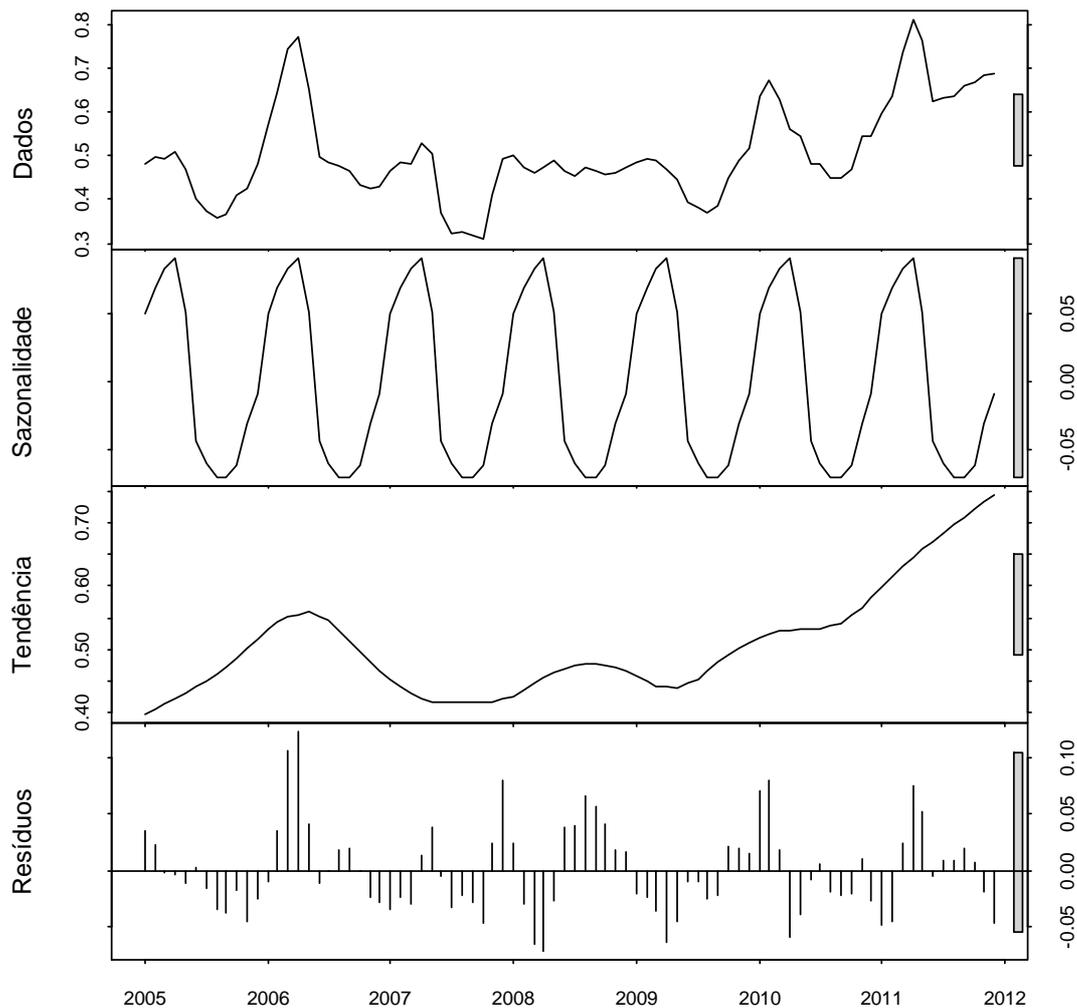


Figura 14 - Decomposição da série

4.1 Método ARIMA

4.1.1 Análise dos dados

O correlograma da série mostra uma forte persistência, que declina lenta e gradualmente à medida que o número de defasagens aumenta, caracterizando um

processo de média móvel. Ao mesmo tempo, o decaimento razoavelmente rápido das autocorrelações parciais indica um comportamento característico de séries com termo auto-regressivo. A figura 15 indica que a série não é estacionária, sendo necessário diferenciá-la.

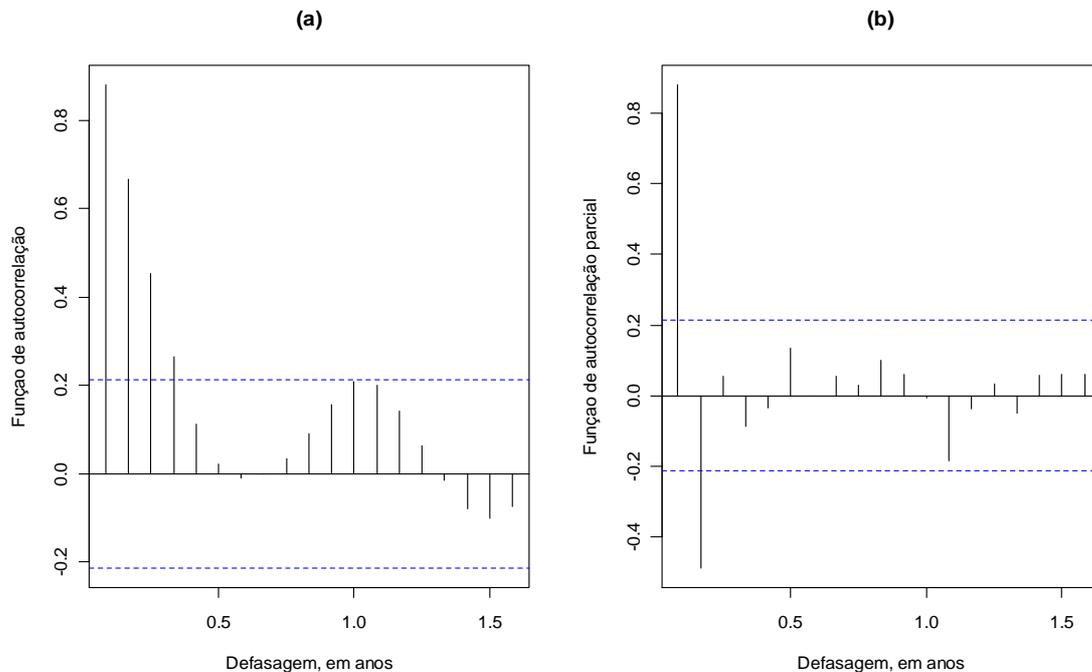


Figura 15 - Autocorrelação (ACF) (a) e autocorrelação parcial (PACF) (b)

Ao analisar a dependência entre as defasagens (figura 16), verifica-se que a dispersão de pontos sobre a primeira defasagem sugere que a relação entre as observações é linear e os aumentos das dispersões observadas nas defasagens seguintes indicam que elas são bem menos dependentes que a primeira. Portanto, em princípio, podemos inferir que o número de defasagens necessárias para tornar a série estacionária será $d=1$, o que será averiguado na fase de identificação dos parâmetros através da previsão automática do modelo ARIMA.

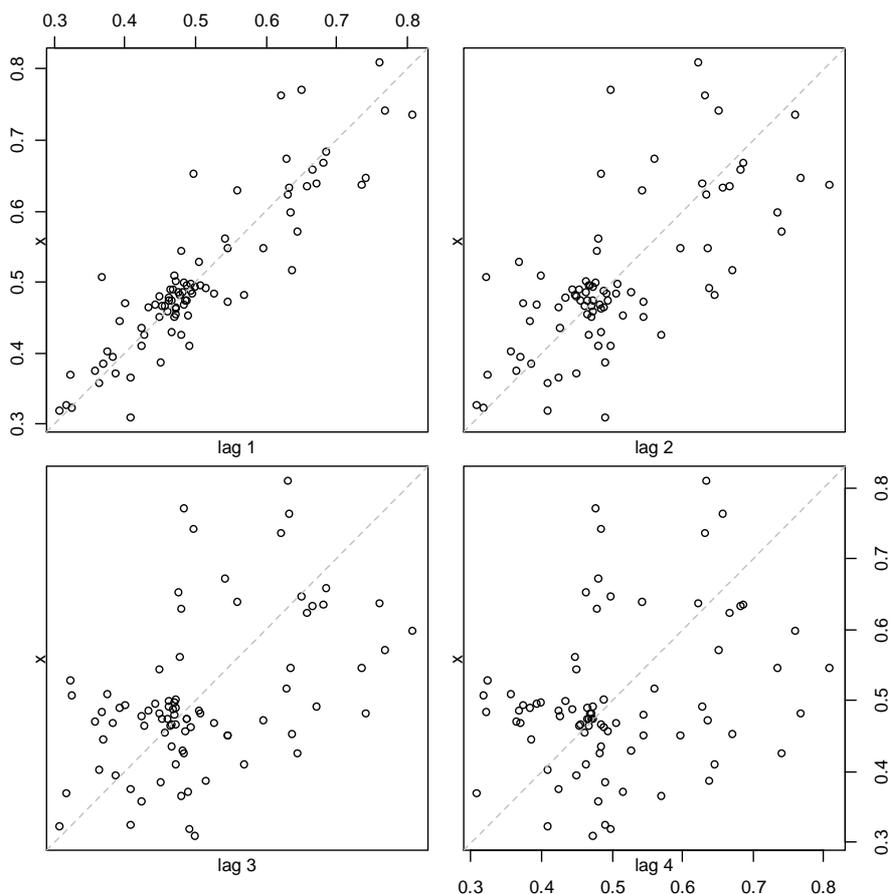


Figura 16 – Dependência entre defasagens

4.1.2 Identificação

A modelagem de Box-Jenkins foi realizada em três fases, conforme explicitado na metodologia. Para a etapa de identificação, utilizou-se a seleção automática dos parâmetros p , d e q do programa R. Como os dados são sazonais, considerou-se ARIMA $(p, d, q) (P, D, Q)$, onde se aplicou o teste Canova-Hansen para verificar se a mudança no padrão sazonal ao longo do tempo é suficiente para justificar uma raiz unitária sazonal, ou se um padrão estável de sazonalidade pode ser modelado através de variáveis *dummy*.

O primeiro parâmetro identificado foi o número de diferenciações necessárias para tornar a série estacionária, escolhido com base no teste de raiz unitária KPSS (*Kwiatkowski, Phillips, Schmidt, Shin*), posteriormente determinaram-se os parâmetros

auto-regressivo e de média móvel, os operadores auto-regressivo sazonal e de média móvel sazonal de ordens P e Q.

O algoritmo do R percorreu um espaço de busca para identificar os possíveis modelos e ao final selecionou aquele que apresentou menor critério de AIC que foi o ARIMA(2,1,3)(0,0,1)[12], o que equivale dizer que este modelo foi formado por dois parâmetros auto-regressivos, três parâmetros de média móvel e uma única diferenciação para tornar a série estacionária, além um parâmetro de média móvel sazonal, com tamanho da sazonalidade de doze meses. Após a fase de identificação, determinaram-se os valores desses parâmetros através do método da verossimilhança.

Os principais resultados do ARIMA (2,1,3)(0,0,1)[12] são apresentados na tabela 2. É possível verificar que todos os parâmetros estimados se encontram dentro do intervalo de confiança que contém os verdadeiros valores desses parâmetros. Outro ponto observado foram os valores muito próximos dos critérios de informação AIC e BIC, que apresentaram resultados negativos, o que é plausível, já que eles dependem a função de verossimilhança.

Tabela 2 – Parâmetros estimados do ARIMA (2,1,3)(0,0,1)[12]

Descrição	Parâmetro estimado	Erro padrão	Intervalo de confiança	
			2,50%	97,50%
ϕ_1	0,077	0,147	-0,210	0,364
ϕ_2	0,555	0,156	0,249	0,861
θ_1	0,603	0,101	0,404	0,801
θ_2	-0,724	0,111	-0,941	-0,506
θ_3	-0,739	0,083	-0,902	-0,576
Θ_1	0,393	0,114	0,169	0,617

Tabela 3 – Critérios de informação

Critérios de informação	AIC	AICc	BIC
	-303,05	-301,55	-286,11

4.1.3 Testes

Pela figura seguinte, percebe-se que o histograma dos resíduos é simétrico (a) e que a maioria dos pontos se encontram sobrepostos à reta (b), representando indícios de normalidade.

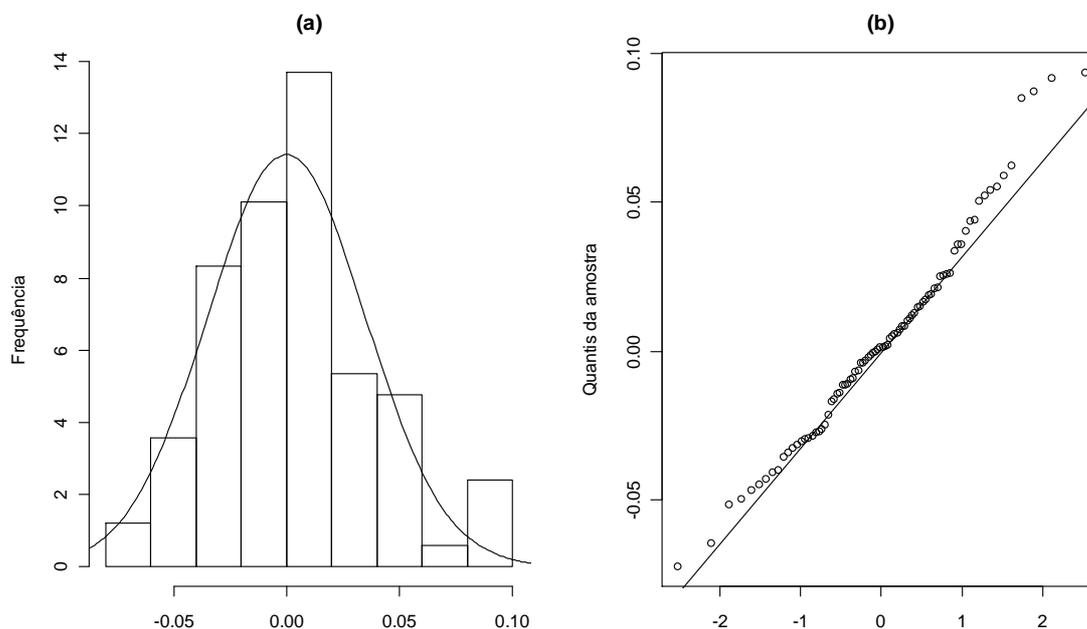


Figura 17 – Histograma (a) e QQ-Plot dos resíduos (b)

O teste Shapiro-Wilk indicou que não existem evidências para rejeitar a hipótese de normalidade, com nível de significância de 5%. Paralelamente, o teste Jarque-Bera caracterizou que o terceiro e quarto momentos da distribuição empírica coincidem com os da normal, ao nível de significância de 5%. Portanto, a análise gráfica preliminar de normalidade dos resíduos foi confirmada pelos resultados dos testes demonstrados a seguir.

Tabela 4 – Testes de normalidade

Modelo	Teste de Normalidade	
ARIMA(2,1,3)(0,0,1)[12]	Shapiro-Wilk	Jarque-Bera
Estatística	W = 0,9771	X ² = 3,1985
p-value	0,139	0,202

A análise conjunta dos testes Ljung-Box e Box Pierce (tabela 5) mostraram que não há indícios para rejeitar a hipótese nula de ruído branco, ao nível de significância de 5%. Ou seja, pode-se afirmar que os resíduos não se correlacionam, portanto mais um dos pressupostos da modelagem Box-Jenkins foi satisfeito, conforme resultados apresentados abaixo.

Tabela 5 – Testes de autocorrelação

Modelo	Autocorrelação	
ARIMA (2,1,3)(0,0,1)[12]	Box-Pierce	Ljung-Box
Estatística	X ² = 0,0046	X ² = 0,0047
p-value	0,9461	0,9451

4.1.4 Modelo ARIMA (2,1,3)(0,0,1)[12]

Depois da fase de identificação e validação conforme testes realizados que informaram que os pressupostos exigidos pelo método Box-Jenkins foram atendidos, portanto o ARIMA (2,1,3)(0,0,1)[12] pode ser utilizado para realizar previsões.

4.2 Holt-Winters

O método de alisamento exponencial é considerado como um procedimento automático de previsão, portanto não é necessário aplicar nenhuma estratégia de modelagem, como por exemplo, transformar a série em estacionária, que é um dos pressupostos do ARIMA. Porém, para garantir a validade desse modelo, deve-se realizar testes para verificar se os resíduos apresentam distribuição normal, pois se essa propriedade não foram satisfeita, os resultados serão espúrios e a capacidade preditiva do modelo será prejudicada.

Como a série observada apresenta níveis de sazonalidade e tendência, o método de suavização exponencial mais adequado é o Holt-Winters (aditivo ou multiplicativo). Assim, adotou-se a nomenclatura proposta por Hyndman *et al.* (2002), no qual foram desenvolvidos três modelos com o auxílio do software R, em que a estimação dos parâmetros de suavização e o estado inicial são determinados pela função de verossimilhança e o critério de seleção é realizado pelo AIC. Dessa forma, temos os seguintes modelos:

- i) ETS (M, Ad, M), equivale ao Holt-Winters Multiplicativo, com erro multiplicativo, tendência aditiva suavizada e sazonalidade multiplicativa;
- ii) ETS (A, Ad, A), representa o Holt-Winters Aditivo, com erro aditivo, tendência aditiva suavizada e sazonalidade aditiva;
- iii) ETS (Z, Z, Z), modelo selecionado automaticamente pelo R, em que o erro foi multiplicativo, a tendência multiplicativa amortecida e a sazonalidade multiplicativa, ou seja, ETS (M, Md, M).

As tabelas seguintes indicam os valores iniciais para as equações de recorrência, os parâmetros de suavização de cada um dos modelos e a tabela 8 demonstra os critérios de informação AIC, AICc e BIC.

Tabela 6 – Valores iniciais para as equações de recorrência

Modelo	ℓ_0	b_0	s_0
ETS (M, Ad, M)	0,3909	0,0106	1,0111
ETS (A, Ad, A)	0,4087	0,0021	-0,0135
ETS (Z, Z, Z)	0,4397	1,0129	1,0042

Tabela 7 - Parâmetros de suavização

Modelo	Parâmetros de suavização				σ
	α	β	γ	ϕ	
ETS (M, Ad, M)	0,9578	5,00E-04	3,40E-03	0,9700	0,0681
ETS (A, Ad, A)	0,9959	2,00E-04	3,40E-03	0,9621	0,0352
ETS (Z, Z, Z)	0,9994	2,00E-04	1,00E-04	0,9800	0,0665

Tabela 8 - Critérios de informação

Modelo	Critérios de informação		
	AIC	AICc	BIC
ETS (M, Ad, M)	-163,2207	-153,9479	-121,8968
ETS (A, Ad, A)	-155,8644	-146,5917	-114,5405
ETS (Z, Z, Z)	-167,3830	-158,1103	-126,0591

A figura 18 ilustra o histograma e QQ-Plot dos resíduos para o modelo ETS (M, Ad, M), que quando submetidos ao teste de Shapiro-Wilk para normalidade apresentaram como resultado $W = 0,9800$ e $p\text{-valor} = 0,2179$, comprovando que se aproximam suficientemente da distribuição normal. O teste Jarque-Bera com $X^2 = 3,6878$ e $p\text{-valor} = 0,1582$ também demonstrou que o terceiro e quarto momentos da distribuição empírica coincidem com os da normal, ao nível de significância de 5%.

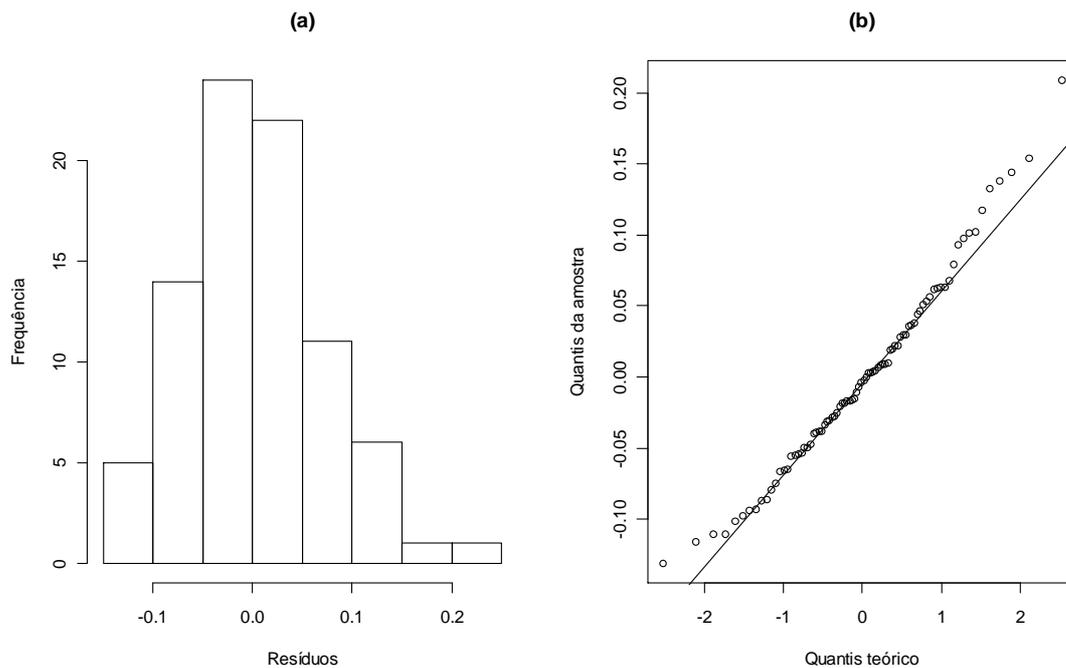


Figura 18 – Histograma (a) e QQ-Plot dos resíduos (b) do ETS (M, Ad, M)

Tabela 9 – Testes de normalidade do modelo ETS (M, Ad, M)

Modelo	Teste de Normalidade	
	ETS (M, Ad, M)	Shapiro-Wilk
Estatística	$W = 0,9800$	$X^2 = 3,6878$
p-value	0,2179	0,1582

O segundo modelo ETS (A, Ad, A) apresentou indícios de normalidade, conforme demonstrado pelo histograma e QQ-Plot (figura 19), confirmado pelos resultados do teste Shapiro-Wilk de $W=0,9890$ e $p\text{-valor}=0,6984$, ou seja, os resíduos apresentam distribuição normal, com nível de significância de 5%. De forma semelhante, o Jarque-Bera de $X^2 = 0,1540$ e $p\text{-valor} = 0,9259$ indicam que o terceiro e quarto momento da distribuição empírica dos resíduos coincidem com os da normal, com nível de significância de 5%.

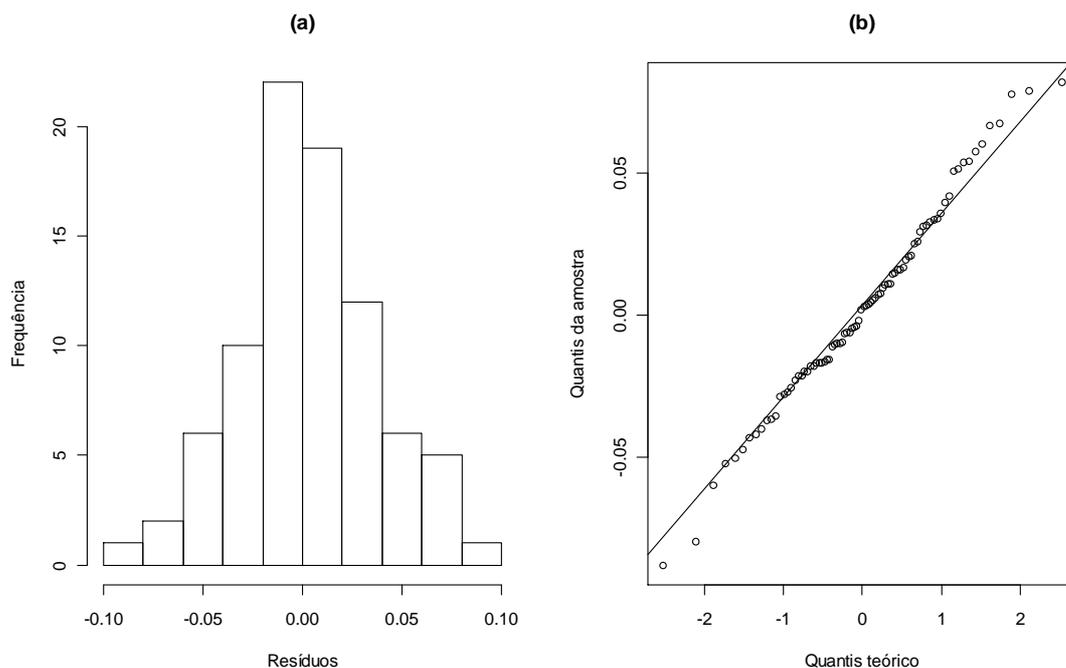


Figura 19 – Histograma (a) e QQ-Plot dos resíduos (b) do ETS (A, Ad, A)

Tabela 10 – Testes de normalidade do modelo ETS (A, Ad, A)

Modelo	Teste de Normalidade	
ETS (A,Ad,A)	Shapiro-Wilk	Jarque-Bera
Estatística	W = 0,9890	$X^2 = 0,1540$
p-value	0,6984	0,9259

Os resíduos do modelo ETS (Z, Z, Z) também apresentaram distribuição normal, conforme teste Shapiro-Wilk ($W=0,9838$), em que o p-valor= $0,3751$ é superior ao nível de significância de 5%, portanto não existem evidências para rejeitar a hipótese nula de que os resíduos provêm de uma distribuição normal. O teste Jarque-Bera ($X^2 = 2,7783$) resultou no p-valor= $0,2493$, ou seja, superior ao nível de significância de 5%, de forma que a hipótese nula de que a assimetria é igual a zero e a curtose igual a três também não pode ser rejeitada. Essas conclusões podem ser visualizadas na figura 20.

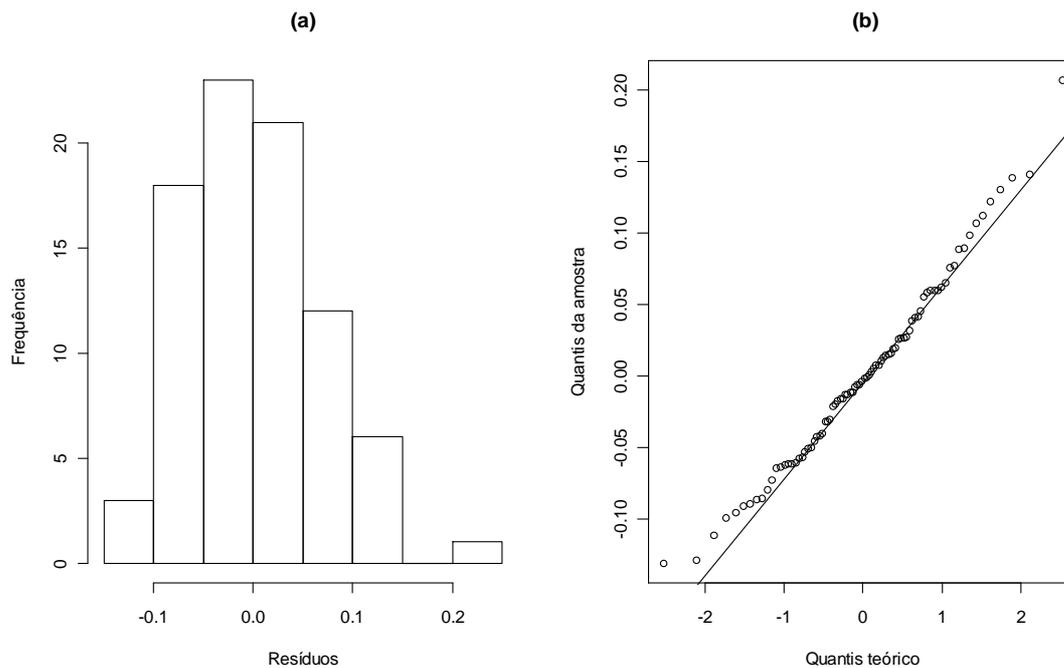


Figura 20 – Histograma (a) e QQ-Plot dos resíduos (b) do ETS (Z, Z, Z)

Tabela 11 – Testes de normalidade do modelo ETS (Z, Z, Z)

Modelo	Teste de Normalidade	
ETS (Z, Z, Z)	Shapiro-Wilk	Jarque-Bera
Estatística	W = 0,9838	X ² = 2,7783
p-value	0,3751	0,2493

Como os modelos de Holt-Winters foram validados e estão aptos para realizar previsões, os mesmos foram descritos após a determinação dos parâmetros através das fórmulas recursivas apresentadas no quadro 5, pág. 60.

a) ETS (A, Ad, A)

$$\ell_t = 0,9959(y_t - s_{t-m}) + 0,0041(\ell_{t-1} + 0,9621b_{t-1})$$

$$b_t = 0,0002(\ell_t - \ell_{t-1}) + (0,9998)0,9621b_{t-1}$$

$$s_t = 0,0034(y_t - \ell_{t-1} - 0,9621b_{t-1}) + 0,9966s_{t-m}$$

b) ETS (M, Ad, M)

$$\ell_t = 0,9578(y_t / s_{t-m}) + (1 - 0,9578)(\ell_{t-1} + 0,9700b_{t-1})$$

$$b_t = 0,0005(\ell_t - \ell_{t-1}) + (0,9995)(0,9700b_{t-1})$$

$$s_t = 0,0034(y_t / \ell_{t-1} - 0,9700b_{t-1}) + 0,9966s_{t-m}$$

c) ETS (M, Md, M)

$$\ell_t = 0,9949(y_t / s_{t-m}) + 0,0006\ell_{t-1}b_{t-1}^{0,98}$$

$$b_t = 0,0002(\ell_t / \ell_{t-1}) + 0,9998b_{t-1}^{0,98}$$

$$s_t = 0,0001(y_t / \ell_{t-1}b_{t-1}^{0,98}) + 0,9999s_{t-m}$$

Os erros dos modelos de Holt-Winters são apresentados na tabela 12. O erro médio (ME) nos mostrou se o modelo de previsão está superestimando ou subestimando o preço do etanol, sendo que o ME positivo indica superestimativa e ME negativo informa que os valores estão subestimados. Apenas para ETS (Z, Z, Z) o erro médio foi negativo, embora muito próximo de zero.

O MAPE indicou o valor médio do erro percentual das previsões sobre todo o conjunto de observações. Todos os modelos de suavização exponencial apresentaram valores muito próximos. Os demais índices também indicaram resultados satisfatórios.

Tabela 12 – Erros dos modelos de suavização exponencial

Modelos	ME	MAE	RMSE	MAPE
ETS (M, Ad, M)	0,01%	2,72%	3,50%	5,33
ETS (A, Ad, A)	0,29%	2,78%	3,52%	5,53
ETS (Z, Z, Z)	-0,02%	2,64%	3,39%	5,20

Como as medidas de erros foram muito semelhantes, adotou-se um segundo critério para seleção que foi o menor desvio padrão dos parâmetros de suavização (tabela 7, pág. 77). Desta forma, ETS (A, Ad, A) foi escolhido dentre os demais para ser comparado com os outros métodos de previsão.

4.3 Rede Neural de Elman

Para Zhang *et al.* (1998), a questão inicial para o desenvolvimento de uma rede neural é a divisão dos dados para o treinamento e conjuntos de teste. Embora não haja uma solução geral para este problema, vários fatores tais como as características problema, o tipo de dados e o tamanho da amostra devem ser considerados na tomada de decisão.

O treinamento deve ser caracterizado por possuir uma boa representação da população em análise. Estudos apresentados por Granger (1993), Tang e Fishwick, (1993) indicam que pelo menos 20% da amostra deve ser aplicada para validar o modelo obtido na fase de treinamento. Com base nesses trabalhos, os padrões de treinamento corresponderam aos dados entre janeiro de 2005 até julho de 2010, totalizando 67 observações e os testes foram realizados com os demais dados, equivalente ao período agosto de 2010 até dezembro de 2011, com 17 padrões observados.

A etapa seguinte foi à normalização estatística dos dados que teve como objetivo adaptar os dados de entrada a faixa dinâmica das funções de ativação da rede neural. Essa normalização está descrita em Weigend *et al.* (1992) e apresentada na metodologia.

Para o problema de previsão das séries temporais, o número de neurônios da camada de entrada corresponde à escolha das defasagens utilizadas para descobrir o padrão existente da série. Porém, *“através da literatura, não encontramos resultados consistentes para a questão da determinação deste importante parâmetro”* Zhang *et al.* (1998). Dessa forma, a rede foi simulada com diferentes números na camada de entrada e aquela que apresentou melhor erro foi considerada, equivalente a três neurônios nessa camada.

Já a quantidade de neurônios na camada de saída indica o horizonte de previsão, que pode ocorrer um passo a frente, ou seja, quando a camada de saída possui apenas um neurônio. Para previsões maiores, a solução é realimentar as informações da rede ou apresentar mais neurônios na camada de saída, o que não exige que a rede seja realimentada. Para a camada de saída, considerou-se apenas um neurônio, o que representa a previsão um passo a frente.

Após várias simulações, os seguintes parâmetros foram definidos:

- Taxa de aprendizado: 0,40;
- Número de ciclos para parada do treinamento: 3.000 iterações;
- Pesos iniciais determinados de forma aleatória: 0,60, -0,15; -0,55 e 0,12;
- Ativação da camada oculta: ocorreu através da função sigmoide logística e a camada de saída utilizou a função identidade.

Para a função de aprendizado foi utilizado o algoritmo *backpropagation*. A importância desse algoritmo é que ele identifica iterativamente a diferença mínima entre as saídas desejadas e as saídas obtidas pela rede neural, de acordo com um erro mínimo, ajustando os pesos entre as camadas através da retropropagação do erro encontrado em cada iteração.

A figura 21 mostra o processo de aprendizado durante o treinamento da rede através da soma dos quadrados dos resíduos. O treinamento ficou estabilizado nas proximidades da época 1100. A linha preta indica a fase de treino e a vermelha representa a fase de testes.

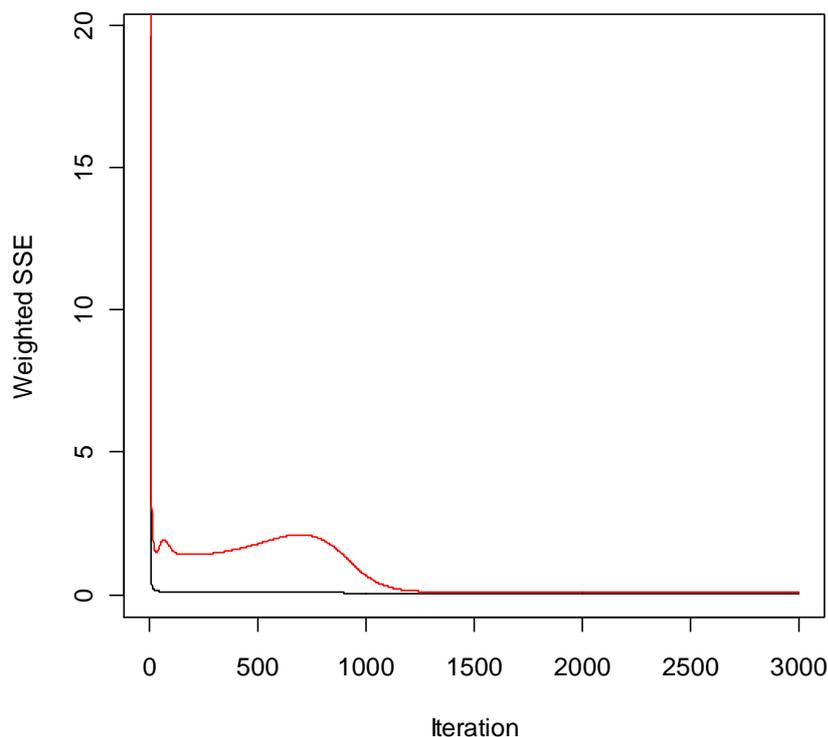


Figura 21 – Processo de aprendizado da rede de Elman

Os erros da rede neural são apresentados abaixo.

Tabela 13 – Erros da rede de Elman

Erros	ME	MAE	RMSE	MAPE
Treino	-0,02%	5,23%	7,52%	3,15%
Teste	0,44%	5,86%	8,76%	3,10%

4.4 Análise Comparativa dos Modelos

Conforme descrito na seção 1.3, o objetivo deste trabalho foi comparar as metodologias estatísticas tradicionais de previsão de séries temporais, especificamente os modelos de suavização exponencial e Box-Jenkins, com técnicas de inteligência artificial caracterizada pela rede neural de Elman. Após várias análises, modelagens e

previsões obtidas pelas técnicas investigadas neste trabalho foi possível comparar as metodologias propostas.

Os modelos estatísticos permitiram identificar os padrões de comportamento como tendência e sazonalidade presentes na série estudada. Após a validação dos modelos de suavização exponencial e Box-Jenkins, os mesmos apresentaram bons resultados de previsão. Já a aplicação da rede neural de Elman para previsão do preço do etanol em Goiás se mostrou uma ferramenta relevante, embora esta técnica não permita identificar padrões de comportamentos existentes na série.

A tabela 14 indica os resultados dos erros de cada modelo proposto. Embora a validação da rede neural seja realizada com base na fase de testes, optou-se por calcular os erros sobre as etapas de validação e teste para que as medidas de erro tivessem o mesmo número de observações daquelas apresentadas pelos métodos estatísticos, visto que a comparação entre modelos com diferenças de tamanho entre as amostras fica prejudicada.

Tabela 14 – Resumo das medidas de erros dos modelos propostos

Descrição	ARIMA (2,1,3)(0,0,1)[12]	ETS (A, Ad, A)	Rede Elman
ME	0,36%	0,29%	0,07%
MAE	2,67%	2,78%	5,37%
RMSE	3,49%	3,52%	7,79%
MAPE	5,23%	5,53%	3,14%

Verificou-se que o desempenho obtido pelo ARIMA (2,1,3)(0,0,1)[12] foi melhor que os demais modelos em relação ao MAE e RMSE. Importante observar que, embora o método de Box-Jenkins foi superior em dois indicadores, os erros do modelo

ETS (A, Ad, A) ficaram muito próximos, o que nos leva a concluir que, em termos de acurácia, a metodologia de suavização exponencial e Box-Jenkins foram equivalentes para previsão do preço do etanol em Goiás.

A rede neural apresentou melhor desempenho através do MAPE, embora as demais medidas de erro também foram consideradas toleradas. A figura 22 ilustra que todas as metodologias encontraram resultados bem próximos aos valores observados para a série estudada.

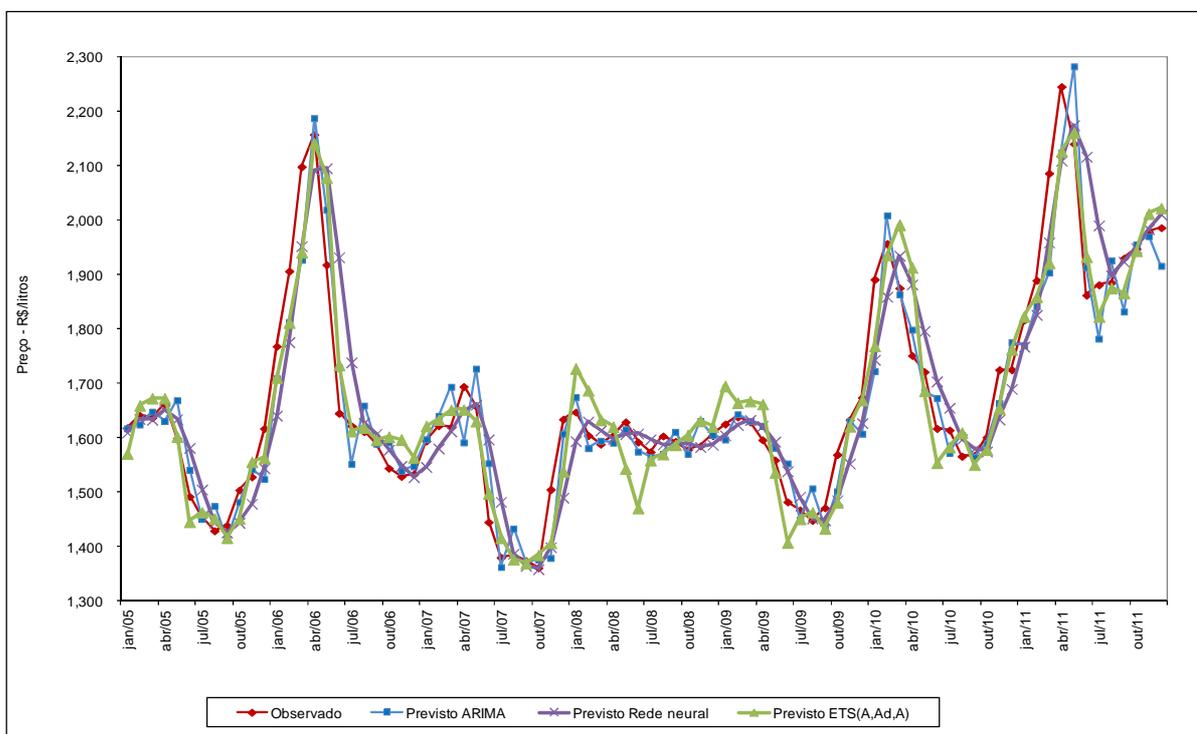


Figura 22 – Análise comparativa entre valores observados e preditos pelos modelos propostos para o preço do etanol em Goiás, de janeiro de 2005 até dezembro de 2011

Ainda, com base na figura 22, verifica-se que os resultados indicaram que para períodos de baixa variação de preços, como por exemplo, janeiro de 2008 até janeiro de 2009, os modelos de Box-Jenkins e a rede neural de Elman apresentaram desempenho similar devido a facilidade de captar o padrão de comportamento da série temporal.

5 CONCLUSÃO

Neste trabalho foram apresentadas metodologias estatísticas e de inteligência artificial para o problema da previsão do preço do etanol no estado de Goiás. Os resultados indicaram que todos os métodos desenvolvidos se mostraram competitivos em termos de predição um passo à frente, destacando-se, no entanto os modelos de suavização exponencial e Box-Jenkins como os mais adequados em termos de parcimônia entre desempenho e complexidade.

Apesar dos resultados satisfatórios da rede neural, há que se observar que ela não oferece conhecimento sobre os dados, pois não identifica padrões de comportamento da série temporal, embora a rede seja um sistema adaptativo que capta as relações funcionais entre os dados através de um processo de treino e aprendizado. Outra questão é que o desenvolvimento de uma rede neural incorpora dificuldades e limitações como à identificação da taxa de aprendizado, pesos iniciais, definição de número de camadas e a topologia da rede, refletindo incertezas quanto à determinação desse modelo.

A principal dificuldade para aplicação da rede neural neste trabalho foi à falta de um algoritmo de treinamento que encontrasse um conjunto global de pesos ideais em um tempo relativamente curto de tempo. Para contornar esse problema, os pesos iniciais foram identificados através da abordagem por tentativa-erro já que não existe uma metodologia específica para este fim.

Como esse treinamento representa um problema de otimização em que o espaço de busca é o espaço de pesos da rede e a função objetivo é uma função do desempenho obtido pela rede para um determinado conjunto de padrões (Montana e Davis, 1989), a sugestão para trabalhos futuros é a combinação das redes neurais com os algoritmos genéticos para problemas de otimização.

Dentre esses problemas de otimização, temos a identificação dos parâmetros das redes como o número de camadas, funções de ativação, taxa de aprendizado, número de neurônios em camadas intermediárias e os pesos iniciais, já que os algoritmos representam uma classe de procedimentos para otimização que buscam explorar um espaço amplo e complexo de forma a encontrar valores próximos ao ótimo global através de regras de transição de estado probabilístico, conseguindo sair de situações de mínimos locais com mais facilidade que as redes neurais.

Outra sugestão é o desenvolvimento de sistemas híbridos de redes neurais que incorporem técnicas estatísticas, como o número de diferenciação ou o tamanho da sazonalidade para determinar o número de camadas de entradas da rede para que a estatística possa justificar a definição dos parâmetros da rede e para que eles não sejam realizados através da abordagem tentativa-erro.

6 BIBLIOGRAFIA

ABELÉM, A. J. G., 1994, **Redes Neurais Artificiais na Previsão de Séries Temporais**. Dissertação de M.Sc., Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

AGÊNCIA NACIONAL DO PETRÓLEO (ANP). Sistema de Levantamento de Preços. Disponível em: <http://www.anp.gov.br/preco>. Acesso em: 15 de jan. 2012.

BACCHI, M.R.P., ALVES, L.R.A., “Formação do Preço do Açúcar Cristal Empacotado ao Varejo da Região Centro-Sul do Brasil”. In: **Agric.**, v. 51, n. 1, pp. 5-22, Jan-Jun, 2004.

BARRETO, J. M., **Introdução às Redes Neurais Artificiais**. Santa Catarina, Laboratório de Conexão e Ciências Cognitivas, 2002.

BERGMEIER, C., BENITEZ, J.M., Neural Networks in R using the Stuttgart Neural Network Simulator: RSNNS. 2010. Disponível em <http://cran.r-project.org/web/packages/RSNNS/citation.html>. Acesso em 05.06.2010.

BOX, G. E. P., JENKINS, G. M., **Times Series Analysis: Forecasting and Control**. São Francisco: Holden-Day, 1976.

BRACE, M.C., SCHMIDT, J., HADLIN, M., “Comparison of the Forecasting Accuracy of Neural Networks With Other Established Techniques”. In: **Proceedings of the First Forum on Application of Neural Networks to Power Systems**, pp. 31-35, Seattle, 1991.

BRAGA, A. P., CARVALHO, A. P. de L. de C., LUDEMIR, T. B., **Redes Neurais Artificiais Teoria e Aplicações**. Rio de Janeiro, LTC, 2007.

BRESSAN, A. A., 2001, **Modelos de Previsão de Preços Aplicados aos Contratos Futuros Agropecuários**. Tese de D.Sc., Universidade Federal de Viçosa, Viçosa, MG, Brasil.

BRESSAN, A.A., LIMA, J.E., “Modelos de Previsão de Preços Aplicados aos Contratos Futuros de Boi Gordo na BM&F”. In: **Nova Economia**, pp. 117-140, Belo Horizonte, Jan - Jun, 2002.

BROOKS, C., **Introductory Econometrics for Finance**. New York, Cambridge University Press, 2002.

BUENO, R. de L. da S., **Econometria de Séries Temporais**. São Paulo, Cengage Learning, 2008.

CAIRE, P., HATABIAN, G., MULLER, C., “Progress in Forecasting by Neural Networks”. In: **Proceedings of the International Joint Conference on Neural Networks**, pp. 540-545, Jun, 1992.

CATALÃO, J. P. S., MARIANO, S. J. P. S., MENDES, V. M. F., FERREIRA, L. A. F.M., “Short-Term Electricity Prices Forecasting in a Competitive Market: A Neural Network Approach.” In: **Electric Power Systems Research**, v. 77, pp. 1297-1304, Aug, 2007.

COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da Safra Brasileira: cana-de-açúcar, terceiro levantamento. Brasília: Conab, 2011.

- CONEJO, A.J., PLAZES, M.A., ESPINOLA, R., MOLINA, A.B. “Day-Ahead Electricity Price Forecasting Using The Wavelet Transform and ARIMA Models”. In: **IEEE Transactions on Power Systems**. v. 20, n. 2, pp. 1035-1042, 2005.
- CONTRERAS, J., ESPINOLA, R., NOGALES, F.J., CONEJO, A.J. “ARIMA Models to Predict Next-Day Electricity Prices”. **IEEE Transactions Power Systems**. v. 18, n. 3, pp.1014 - 1020, 2003.
- CORRAR, L. J., THEÓPHILO, C. R., **Pesquisa Operacional para Decisão em Contabilidade e Administração: Contabilometria**. São Paulo, Atlas, 2004.
- EHLERS, R.S. *Análise de Séries Temporais*. [S.L]: USP, 2009. Disponível em: <http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>. Acesso em 15 jan.2009.
- ELMAN, J. L., “Finding structure in time”. **Cognitive Science**, v. 14, pp. 179-211, 1990.
- ENGLE, R.F., “Autoregressive Conditional Heterocedasticity With Estimates of the Variance of United Kingdom Inflation”. **Econometrica**, v. 50, pp. 987-1008, 1982.
- FARAWAY, J., CHATFIELD, C., “Time Series Forecasting With Neural Networks: Case Study”. **Research Report 95-06**, London: UK, University of Bath, 1995. 20p.
- FAUSETT, L., **Fundamentals of Neural Networks Architectures, Algorithms And Applications**. [S.L], Prentice Hall, 1994.
- FAVA, V.L., “Metodologia de Box-Jenkins para Modelos Univariados”. In. VASCONCELOS, M. A. S., ALVES, D., (orgs). *Manual de Econometria*. São Paulo: Atlas, 2000, pp. 205 - 231.
- FISCHER, S., 1982, **Séries Univariantes de Tempo: Metodologia de Box & Jenkins**. Tese de D.Sc., Fundação de Economia e Estatística, Porto Alegre, RS, Brasil.
- FLORES, J.H.F., 2009, **Comparação de Modelos MLP/RNA e Modelos Box-Jenkins em Séries Temporais Não Lineares**. Dissertação de M.Sc., Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil.
- GARDNER JR, E.S., “Exponential Smoothing: The state of the art - Part II”. In: **International Journal of Forecasting**, v. 22, ed. 4, pp.637 - 666, Out-Dez 2006.
- GELPER, S., FRIED, R., CROUX, C., “Robust Forecasting With Exponential an Holt-Winters Smoothing”. Faculty of Economics and Applied Economics, 2007.
- GIL, A. C., **Como Elaborar Projetos de Pesquisa**. 4 ed. São Paulo, Atlas, 2002.
- GOMES, D. T., 2005, **Modelos de Redes Neurais Recorrentes Para Previsão de Séries Temporais de Memórias Curta e Longa**. Dissertação de M.Sc., Universidade Estadual de Campinas, Campinas, SP, Brasil.
- GÓMES, V., “Automatic Model Identification in the Presence of Missing Observations and Outliers”. **Working paper D-98009**, Ministerio de Economia y Hacienda, Dirección General de Análisis y Programación Presupuestaria, Madri, 1998.
- GRANGER, C.W.J., “Strategies for Modelling Nonlinear Time Series Relationships”. **The Economic Record**, 69 (206), pp. 233 - 238, 1993.
- GUJARATI, D. N., **Econometria Básica**. Rio de Janeiro, Elsevier, 2000.
- HAYKIN, S., **Redes Neurais: Princípios e Práticas**. 2 ed. Porto Alegre, Bookman, 2001.

HANNAN, E.J., RISSANEN, J., “Recursive Estimation of Mixed Autorregressive Moving Average Estimation”. **Biometrika**, n.69, pp. 81-94, 1983.

HU, M.J.C., “Application of the Adaline System to Weather Forecasting”. **Master Thesis**, Technical Report 6775-1, Stanford Electronic Laboratories, Stanford, CA, Jun, 1964.

HYNDMAN, R.J., KHANDAKAR, Y. “Automatic Time Series Forecasting: The forecast Package for R”. **Journal of Statistical Software**, v. 27, ed. 3, Jul, 2008.

HYNDMAN, R.J., KOEHLER, A.B., SNYDER, R.D., GROSE, S., “A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods”. **International Journal of Forecasting**, v. 18, ed. 3, pp. 439 - 454, Jul-Set, 2002.

JHEE, W.C., LEE, K.C., LEE, J.K., “A Neural Network Approach for the Identification of the Box-Jenkins Model. Network”. In: **Computation in Neural Systems 3**, pp. 323-339, 1992.

KASABOV, N. K., **Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering**. USA, Bradford Book, 1998.

KIRSTEN, H.A., 2009, **Comparação Entre Modelos Holt-Winters e Redes Neurais Para Previsão de Séries Temporais Financeiras**. Dissertação de M.Sc., Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brasil.

LAPEDES, A.; FARBER, R., “How Neural Nets Work”. In: Anderson, D.Z., (Ed.), **Neural Information Processing Systems, American Institute of Physics**, pp. 442-456, New York, 1988.

LEE, J.K., JHEE, W.C., “A Two-Stage Neural Network Approach for ARMA Model Identification With ESACF”. **Decision Support Systems**, v. 11, pp. 461 - 479, 1994.

LIU, L.M., “Identification of Seasonal Arima Models Using a Filtering Method”. **Communications in Statistics: Theory and Methods**, v.18, pp. 2279 - 2288, 1989.

MAKRIDAKIS, S. G., WHEELWRIGHT, S. C., HYNDMAN, R. J., **Forecasting: Methods and Applications**. 3 ed. New York, Wiley, 1998.

MARCONI, M. de A., LAKATOS, E. M., **Fundamentos de Metodologia Científica**. 6 ed. São Paulo, Atlas, 2009.

MARGARIDO, M.A.; ANEFALOS, L.C., “Testes de Raiz Unitária e o Software SAS”. In: **Agric**, 46(2), pp.19 - 45, São Paulo, 1999.

MATOS, O.C., **Econometria Básica: Teoria e Aplicações**. 3 ed. São Paulo, Atlas, 2000.

MCNELIS, P. D., **Neural Networks in Finance: Gaining Predictive Edge in the Market**. [S.L], Elsevier, 2005.

MILESKI JR., 2007, **Análise de Métodos de Previsão de Demanda Baseados em Séries Temporais em Uma Empresa do Setor de Perfumes e Cosméticos**. Pontifícia Universidade Católica do Paraná, Curitiba, PR, Brasil.

MORAES, M.A.F.D. “Desregulamentação da Agroindústria Canavieira: Novas Formas de Atuação do Estado e Desafios do Setor Privado”. In: MORAES, M.A.F.D; SHIKIDA, P.F.A (orgs.). **Agroindústria Canavieira no Brasil: Evolução, Desenvolvimento e Desafios**. São Paulo: Atlas, 2002. p. 21-41.

- MORETTIN, P.A., **Econometria Financeira: Um Curso de Séries Temporais Financeiras**. São Paulo, Blucher, 2008.
- MORETTIN, P.A., TOLOI, C.M.C., **Análise de Séries Temporais**. 2 ed. São Paulo, Blucher, 2006.
- OLIVEIRA, M.A.A., AMARAL, J.L.M., AMARAL, J.F.M., “Algoritmos para Auto-Ajuste de Sensores Utilizando Redes Neurais”. **VIII Semetro**. João Pessoa, Paraíba, 17-19 Jun 2009.
- PASQUOTTO, J.L.D., 2010, **Previsão de Séries Temporais no Varejo Brasileiro: Uma Investigação Comparativa das Redes Neurais Recorrentes de Elman**. Dissertação de M.Sc., Universidade de São Paulo, São Paulo, SP, Brasil.
- PORTUGAL, M.S., FERNANDES, L.G.G. “Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução”. In: **Revista Nova Economia**, v. 6, n.1, pp. 51-73, Jul.1996.
- R DEVELOPMENT CORE TEAM (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- REYNOLDS, S.B., MELLICHAMP, J.M., SMITH, R.E., “Box-Jenkins Forecast Model Identification”. In: **AI Expert**, pp. 15-28, Jun. 1995.
- SCHARWZ, G., “Estimating the Dimension of a Model”. In: **Annals of Statistics** 6, pp. 461 - 464, 1978.
- SHABRI, A., “Comparision of Time Series Forecasting Methods Using Neural Networks and Box-Jenkins Model”. In: **Matematika**, pp. 52-32, 2001.
- SHAPIRO, S.S, WILK, M.B., “An Analysis of Variance Test for Normality (Complete Samples)”. In: **Biometrika**, v.52, n.3/4, pp.591-611, Dez. 1965.
- SHARDA, R., “Neural Networks for the MS/OR Analyst: An Application Bibliography”. **Interfaces**, v. 24, pp. 116 - 130, 1994.
- SHARDA, R., PATIL, R.B., “Connectionist Approach to Time Series Prediction: An Empirical Test”. In: **Journal of Intelligent Manufacturing** 3, pp. 317 - 323, 1992.
- SHUMWAY, R.H., STOFFER, D.S., **Time Series Analysis And Its Applications: With R Examples**. 3ed. New York, Springer, 2010.
- SINDICATO DA INDÚSTRIA E FABRICAÇÃO DO AÇÚCAR NO ESTADO DE GOIÁS (SIFAEG). Números do setor. Disponível em: <http://www.sifaeg.com.br/index.html>. Acesso em: 30 de mar. 2011.
- SOBREIRO, V. A., ARAÚJO, P. H. de S. L., NAGANO, M.S., “Precificação do Etanol Utilizando Técnicas de Redes Neurais Artificiais”, In: **Revista de Administração**, v. 44, n.1, pp. 46 - 58, Jan, 2009.
- SUTSKEVER, I., MARTENS, J., HINTON, G., “Generating Text With Recurrent Neural Networks”. In: 28 International Conference on Machine Learning, Bellevue, WA, USA, 2011.
- TANG, Z., ALMEIDA, C., FISHWICH, P.A., “Time Series Forecasting Using Neural Networks vs Box-Jenkins Methodology”. In: **Simulation**, v.57 n. 5, pp. 303 - 310, 1991.

TANG, Z., FISHWICK, P.A., "Feedforward Neural Nets as Models for Time Series Forecasting". **ORSA Journal on Computing**. v.5, n. 4, pp. 374 - 385, 1993.

WEIGEND, A.S., HUBERMAN, B.A., RUMELHART, D.E., "Predicting Sunspots and Exchange Rates With Connectionist Networks". In: Casdagli, M., Eubank, S. (Eds.), **Nonlinear Modeling and Forecasting**. Addison-Wesley, Redwood City, CA, pp. 395–432, 1992.

WERBOS, P.J., "Generalization of Backpropagation With Application to a Recurrent Gas Market Model". In: **Neural Networks**, n. 1, pp. 339 - 356, 1988.

ZHANG, G., PATUWO, B.E., HU, M.Y., "Forecasting With Artificial Neural Networks: The State of The Art". In: **International Journal of Forecasting**. v. 14, ed. 1, pp. 35 - 62, 1998.

ZUBEN, F.J.V., 1996, **Modelos Paramétricos e Não-Paramétricos de Redes Neurais**. Tese de D.Sc., Universidade Estadual de Campinas, Campinas, SP, Brasil.

APÊNDICE A – Códigos no R**Modelo de Box-Jenkins**

```
### Pacotes necessários

library(forecast)

library(fUnitRoots)

library(TSA)

library(tseries)

library(urca)

y=ts(preco,frequency=12,start=c(2005, 1))

log=log(y)

x=ts(log,frequency=12,start=c(2005, 1))

plot(x, main=" ", ylab="Valores (logaritmo)", xlab="Ano")

class(x)

start(x); end(x); frequency(x)

summary(x)

boxplot(x ~ cycle(x), main="(b)", las=3) ### análise variância

### las=ordem da escala 1=horizontal; 3=vertical

decomp<-stl(x,s.window='periodic') ### decomposição da série

plot(decomp)

### Estatística descritiva -

summary(x)

quantile(x)

max(x)-min(x) ### amplitude

var(x) ### variância
```

```
sd(x) ### desvio padrão

100*sd(x)/mean(x) ### Coeficiente de variação, em porcentagem

subset(table(x),table(x)==max(table(x)))### modo

### Análise preliminar

lag.plot(x,lags=4,do.lines=FALSE)

### Função de autocorrelação e autocorrelação parcial

par(mfrow=c(1,2))

acf(x,ylab="Função de autocorrelação", xlab="Defasagem, em anos", main="(a)")

pacf(x, ylab="Função de autocorrelação parcial", xlab="Defasagem, em anos",
main="(b)")

### Modelo Arima ###

library(forecast)

arimafit<-auto.arima(x)

arimafit

summary(arimafit)

confint(arimafit) ### interval de confiança

### Resíduos

residuos<-c(arimafit$residuals)

residuos

Box.test(residuos, type='Box-Pierce')

tsdiag(arimafit)

Box.test(residuos,lag=1,type='Ljung-Box')

### Teste Shapiro Wilk e Jarque Bera

shapiro.test(residuos)

jarque.bera.test(residuos)
```

```
### histograma com curva normal superposta

par(mfcol=c(1,2))

hist(residuos,freq=F, xlab="Resíduos", ylab="Frequência",main="(a)")

d = seq(range(residuos)[1]-3*sd(residuos),range(residuos)[2]+3*sd(residuos),0.001)

lines(d,dnorm(d,0,sd(residuos)))

### gráfico de probabilidades normais

qqnorm(residuos, xlab="Quantis teórico", ylab="Quantis da amostra", main="(b)")

qqline(residuos)

### Análise Real x Calculado

plot(x, main="Real x Calculado", ylab="log", xlab="Período")

lines(x-arimafit$resid,col="red")

previsãoARIMA<-c(x-arimafit$resid) ### Demonstrar valores previstos do MODELO

previsãoARIMA
```

```
#####
```

Modelos de Holt-Winters

```

### Modelo Holt Winters Multiplicativo – ETS(M,A,M)

modeloMAM<- ets(x,model="MAM",damped=TRUE)

modeloMAM

accuracy(modeloMAM) ### medidas de desempenho

### Análise dos resíduos

par(mfcol=c(1,2))

modeloMAM$residuals

hist(modeloMAM$residuals, xlab="Resíduos", ylab="Frequência", main="(a)")

qqnorm(modeloMAM$residuals, xlab="Quantis teórico", ylab="Quantis da amostra",
main="(b)")

qqline(modeloMAM$residuals)

jarque.bera.test(modeloMAM$residuals)

shapiro.test(modeloMAM$residuals)

### Previsão

MAM<- forecast(modeloMAM, h=12)

MAM

plot(MAM, main="(a) ETS (M,A,M)")

seqplot.ts(x,fitted(MAM))### Gráfico ajustado com o modelo

previsãoMAM<-c(x- modeloMAM$residuals) ### Demonstrar valores previstos do
MODELO

previsãoMAM

#####

```

```

### Modelo Holt Winters Aditivo – ETS(A,A,A)

modeloAAA <- ets(x,model="AAA",damped=TRUE)

modeloAAA

accuracy(modeloAAA) ### medidas de desempenho

### Análise dos resíduos

par(mfcol=c(1,2))

modeloAAA$residuals

hist(modeloAAA$residuals, xlab="Resíduos", ylab="Frequência", main="(a)")

qqnorm(modeloAAA$residuals, xlab="Quantis teórico", ylab="Quantis da amostra",
main="(b)")

qqline(modeloAAA$residuals)

jarque.bera.test(modeloAAA$residuals)

shapiro.test(modeloAAA$residuals)

### Previsão

AAA<- forecast(modeloAAA, h=12)

AAA

plot(AAA, main="(a) ETS (A,A,A)")

seqplot.ts(x,fitted(AAA)) ### Gráfico ajustado com o modelo

previsãoAAA<-c(x- modeloAAA$residuals) ### Demonstrar valores previstos do
MODELO

previsãoAAA

```

```
#####
```

```

### Suavização Exponencial Automática– ETS(Z,Z,Z)

modeloZZZ<- ets(x,model="ZZZ",damped=TRUE)

modeloZZZ

accuracy(modeloZZZ)### medidas de desempenho

### Análise dos resíduos

par(mfcol=c(1,2))

modeloZZZ$residuals

hist(modeloZZZ$residuals, xlab="Resíduos", ylab="Frequência", main="(a)")

qqnorm(modeloZZZ$residuals, xlab="Quantis teórico", ylab="Quantis da amostra",
main=" (b)")

qqline(modeloZZZ$residuals)

jarque.bera.test(modeloZZZ$residuals)

shapiro.test(modeloZZZ$residuals)

### Previsão

ZZZ<- forecast(modeloZZZ, h=12)

ZZZ

plot(ZZZ, main="(a) ETS (Z,Z,Z)")

seqplot.ts(x,fitted(ZZZ)) ### Gráfico ajustado com o modelo

previsãoZZZ<-c(x- modeloZZZ$residuals) ### Demonstrar valores previstos do
MODELO

previsãoZZZ

#####

```

Rede Neural de Elman

```
### Pacotes necessários
```

```
library(AER)
```

```
library(strucchange)
```

```
library(survival)
```

```
library(splines)
```

```
library(RSNNS)
```

```
library(tseries)
```

```
### Entrada de dados
```

```
P=read.table('C:/Documents and  
Settings/anasousa/Desktop/p_7.csv',h=T,sep=";",dec=",")
```

```
### Normalização de dados
```

```
pn=normalizeData(P,type="norm")
```

```
### Separar dados para treinamento:
```

```
###Tabela de preços normalizados
```

```
treinoXP=pn[c(1:67),c(1:3)]
```

```
treinoXP=matrix(treinoXP,ncol=3)
```

```
treinoYP=pn[c(1:67),4]
```

```
treinoYP=matrix(treinoYP,ncol=1)
```

```
### Separar dados para teste:
```

```
testeXP=pn[c(68:84),c(1:3)]
```

```
testeXP=matrix(testeXP,ncol=3)
```

```
testeYP=pn[c(68:84),4]
```

```
testeYP=matrix(testeYP,ncol=1)
```

```
### Configurando a RNA:
```

```

alfa=c(0.4) #Pode ser feito alterações desse parâmetro

nn=3 #Number of units in the hidden layer(s)

n=3000 #maximum of iterations to learn

pesos=c(0.6055604,-0.1535941,-0.5543522,0.1268725)

TreinoErr=0 #Matriz para armazenar os números relativos ao treinamento da RNA

TesteErr=0 #Matriz para armazenar os números relativos ao teste da RNA

TreinoErr=matrix(TreinoErr,nrow=n,ncol=2)

TesteErr=matrix(TesteErr,nrow=n,ncol=2)

par(mfrow = c(1,1))

### Rede de Elman:

ElmanP=elman(treinoXP,treinoYP,                                size=nn,
maxit=n,initFunc="JE_Weights",initFuncParams=pesos,learnFunc="JE_BP",
learnFuncParams=alfa,      updateFunc="JE_Order",      updateFuncParams=c(0),
shufflePatterns=FALSE,linOut=TRUE, outContext=FALSE,testeXP,testeYP)

plotIterativeError(ElmanP)

### Valores previstos (treino e teste):

ElmanP$fitted.values ### valores previstos de treino

ElmanP$fittedTestValues ### valores previstos de teste

ElmanP

summary(ElmanP)

```