

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

***APLICAÇÃO DE REDES NEURAS NA PREDIÇÃO DE DE-
MANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL***

JORCENILSON PEREIRA MAIA

ORIENTADOR: ALEXANDRE RICARDO SOARES ROMARIZ

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELETRICA

PUBLICAÇÃO: PPGENE.DM – 345/08

BRASILIA / DF: JULHO 2008

***APLICAÇÃO DE REDES NEURAIIS NA PREDIÇÃO DE DEMANDA DE
CRÉDITO NO SISTEMA FINANCEIRO NACIONAL***

JORCENILSON PEREIRA MAIA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

ALEXANDRE RICARDO SOARES ROMARIZ, Ph. D. (UNB)
(ORIENTADOR)

JANAINA GONÇALVES GUIMARÃES, Dra. (UNB)
(EXAMINADOR INTERNO)

JOÃO DE SOUZA NETO, Dr. (UNB)
(EXAMINADOR EXTERNO)

ADSON DA ROCHA, Ph. D. (UNB)
(SUPLENTE)

FICHA CATALOGRÁFICA

MAIA, JORCENILSON

APLICAÇÃO DE REDES NEURAIIS NA PREDIÇÃO DE DEMANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL. Brasília, 2008. 71p.

Dissertação (Mestrado) – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

1. Redes Neurais 2. Series Temporais 3. Crédito

REFERÊNCIA BIBLIOGRÁFICA

MAIA, J. P. (2008). APLICAÇÃO DE REDES NEURAIIS NA PREDIÇÃO DE DEMANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL, Publicação PPGENE. DM-345/08, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 71p.

CESSÃO DE DIREITOS

AUTOR: Jorcenilson Pereira Maia

TÍTULO: APLICAÇÃO DE REDES NEURAIIS NA PREDIÇÃO DE DEMANDA DE CREDITO NO SISTEMA FINANCEIRO NACIONAL.

GRAU: Mestre

ANO: 2008

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Jorcenilson Pereira Maia

Qd. 5 Conj. A Casa 35, Sobradinho.

73.030-051 Brasília – DF – Brasil.

AGRADECIMENTOS

Ao Prof. Dr. Romariz pelo apoio na realização do trabalho, pelo incentivo e, principalmente, pela orientação.

À Julia Cristina pelo apoio, paciência, amor e compreensão, durante esta caminhada.

Ao Departamento de Engenharia Elétrica pela utilização e facilidade de acesso aos seus laboratórios, bem como aos professores e funcionários que estiveram sempre dispostos a colaborar.

À memória de Wilson Pereira Maia.

RESUMO

O principal objetivo deste trabalho é avaliar o uso de técnicas de previsão de séries temporais com Redes Neurais Artificiais. Como aplicação, foi analisada a demanda do volume de crédito com recursos livres no Sistema Financeiro Nacional. Utilizaram-se técnicas de análise multivariadas para determinar as séries independentes que explicam a variável dependente (Recursos Livres). A sugestão deste trabalho é utilizar redes neurais em cooperação com métodos estatísticos, buscando alcançar melhores resultados que aqueles obtidos com técnicas estatísticas isoladamente. O trabalho inicia-se com considerações sobre previsões de séries temporais e uma caracterização breve sobre redes neurais. A seguir, o estudo de caso utilizando uma série histórica de recursos livres no Sistema Financeiro Nacional. Encerra-se com conclusões e recomendações.

Palavras-chave: *previsão de recursos livres, redes neurais artificiais.*

ABSTRACT

The main objective of this work is to evaluate the use Artificial Neural Networks in the time series prediction. Specifically, we study credit volume with free resources in the National Financial System. Multivariate data analysis was used to identify independent series that explain the dependent variable (Free Resources). Our approach is to use Artificial Neural Networks in cooperation with statistic techniques, looking for better results than those obtained with statistics alone. The work begins with an overview on prediction of time series and a brief review of Artificial Neural Networks. Next, a case study is presented, using a historical series of free resources in the National Financial System, followed by conclusions and recommendations.

Keywords: credit forecasting, artificial neural networks.

Sumário

<i>APLICAÇÃO DE REDES NEURAIIS NA PREDIÇÃO DE DEMANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL</i>	ii
FICHA CATALOGRÁFICA.....	iii
REFERÊNCIA BIBLIOGRÁFICA.....	iii
CESSÃO DE DIREITOS.....	iii
AGRADECIMENTOS.....	iv
RESUMO	vi
ABSTRACT.....	vii
LISTA DE FIGURAS.....	x
LISTA DE TABELAS.....	xi
LISTA DE ABREVIATURAS.....	xii
1. INTRODUÇÃO	1
1.1. REDES NEURAIIS NO PROCESSAMENTO DE SÉRIES TEMPORAIS.....	2
1.2. OBJETIVOS DA PESQUISA.....	3
1.3. ORGANIZAÇÃO DA DISSERTAÇÃO	4
2. SÉRIES TEMPORAIS	5
2.1. TÉCNICAS DESCRITIVAS	6
2.1.1. Coeficiente de correlação de Pearson.....	6
2.1.2. Autocorrelação	7
2.2. PROGNÓSTICO.....	9
2.3. MÉTODOS SIMPLES DE PREVISÃO DE SÉRIES TEMPORAIS	10
2.3.1. Média móvel.....	10
2.3.2. Alisamento exponencial simples	11
2.3.3. Alisamento exponencial linear	11
2.3.4. Alisamento Exponencial Sazonal e Linear de Winter	12
2.4. MÉTODOS AVANÇADOS DE PREVISÃO DE SÉRIES TEMPORAIS	13
2.4.1. Modelos Autoregressivo e de Médias Móveis	13
2.4.2. Modelo de Box e Jenkins	14
2.5. CONSIDERAÇÕES SOBRE PREVISÃO DE SÉRIES TEMPORAIS	15
2.6. ANÁLISES DOS RESÍDUOS	16
2.7. PREVISÃO	18
2.8. PERFORMANCE PREDITIVA	18

2.9.	CRITÉRIOS DE INFORMAÇÃO	19
3.	REDES NEURAIS ARTIFICIAIS.....	20
3.1.	HISTÓRICO	20
3.2.	MODELO DE NEURÔNIO ARTIFICIAL	21
3.3.	FUNÇÕES DE ATIVAÇÃO	22
3.4.	ARQUITETURAS DE REDES NEURAIS ARTIFICIAIS.....	23
3.5.	TIPOS DE TREINAMENTO.....	24
3.6.	REDES NEURAIS ARTIFICIAIS DE MÚLTIPLAS CAMADAS	26
3.7.	O ALGORITMO DE RETRO PROPAGAÇÃO DE ERROS	27
4.	APLICAÇÃO DE REDES NEURAIS NA PREDIÇÃO DE DEMANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL.....	31
4.1.	TRATAMENTOS ESTATÍSTICOS.....	31
4.1.1.	Análise de correlação e regressão	32
4.2.	REGRESSÃO LINEAR MÚLTIPLA.....	33
4.2.1.	Suposições em análise de regressão múltipla.....	34
4.2.2.	Estimação do modelo de regressão e avaliação do ajuste geral do modelo	35
4.2.3.	Avaliação da significância estatística do modelo na previsão da variável dependente.....	42
4.2.4.	Interpretação da variável estatística de regressão.....	43
4.3.	REDES NEURAIS.....	44
4.3.1.	Pré-processamento das séries	46
4.3.2.	Pós-processamento das séries.....	46
4.4.	ESTUDOS DE CASOS.....	46
4.5.	TESTES E RESULTADOS	46
5.	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	51
	APÊNDICES	54

LISTA DE FIGURAS

Figura 1-1 Modelo de um neurônio biológico	2
Figura 3-1 Modelo não-linear de um neurônio – Haykin 2001	21
Figura 3-2 Exemplos de tipos de função de ativação.....	22
Figura 3-3 Camadas de uma rede – Fontes (1998)	29
Figura 4-1 Recursos Livres no Sistema Financeiro Nacional.....	31
Figura 4-2 Analise resíduos	33
Figura 4-3 Utilização de séries sem tratamento estatísticos - treinamento	47
Figura 4-4 Utilização de séries sem tratamento estatísticos - teste.....	47
Figura 4-5 Utilização de séries sem tratamento estatísticos - métricas	48
Figura 4-6 Utilização de séries com tratamento estatísticos - treinamento	48
Figura 4-7 Utilização de séries com tratamento estatísticos - teste	49
Figura 4-8 Utilização de séries com tratamento estatísticos - métricas.....	49
Figura 4-9 Método estatístico analise multivariada de dados – métrica.....	50

LISTA DE TABELAS

Tabela 4-1 Avaliação qualitativa de r (correlação) quanto à intensidade.....	32
Tabela 4-2 Variáveis selecionadas regressão múltipla	35
Tabela 4-3 Demais variáveis econômicas iniciais	35

LISTA DE ABREVIATURAS

A

AR (Auto Regressive)	12, 13, 14, 15, 17
ARIMA (Auto Regressive Integrated Moving Average)	12, 13, 14
ARMA (Auto Regressive Moving Average).....	12, 13, 14, 16, 19

M

MA (Moving Average)	12, 13, 14, 15
MAPE (Mean Absolute Percentage Error).....	44, 50
MLP (Muti Layer Perceptron).....	2, 3, 4, 44, 46, 50

P

PIB (Produto Interno Bruto).....	1, 4, 37, 38
----------------------------------	--------------

R

RMSE (Root Mean Squared Error).....	44, 45, 50
-------------------------------------	------------

S

SELIC (Taxa de Financiamento no Mercado Interbancário de Um Dia)	31
SFN (Sistema Financeiro Nacional)	1, 31, 33, 44, 53

T

TDNN (Time Delay Neural Network)	4
----------------------------------------	---

U

U de Theil (Métrica Verificação Previsão Trivial)	19, 44, 45
---------------------------------------------------------	------------

1. INTRODUÇÃO

Este estudo examina, por meio de uma ferramenta de inteligência computacional, a relação entre o volume de crédito com Recursos Livres (crédito ofertado pelos bancos independentemente das políticas de direcionamento de crédito do governo federal) no Sistema Financeiro Nacional e variáveis econômico-financeiras. Com a estabilização da economia as operações de crédito do sistema financeiro apresentaram crescimento expressivo nos últimos cinco anos. A crescente oferta de crédito, principalmente devido à demanda reprimida em décadas anteriores. Verifica-se no final de 2007 que os Recursos Totais (livres + controlados) alocados no SFN têm um percentual de aproximadamente 34,5% do PIB, valor este ainda muito abaixo dos países com economia estabilizada, cerca de 60% do PIB e cerca de 100% do PIB em países de primeiro mundo. Em maio de 2008 as operações realizadas com recursos livres, 71,5% da carteira total do sistema financeiro, somaram R\$ 746,5 bilhões. Esse desempenho foi determinado, sobretudo pelo aumento do crédito consignado, habitacional e arrendamento mercantil, alavancando em grande parte, o desempenho das carteiras de pessoas físicas. O segmento de pessoas jurídicas apresentou crescimento mais modesto em relação ao segmento pessoa física. Assinale-se que a evolução maior do segmento pessoa jurídica concentra-se na modalidade capital de giro para empresas. A consolidação de um cenário macroeconômico estável vem favorecendo o aumento da previsibilidade para os tomadores de crédito, condição essencial para o planejamento de longo prazo. A crescente oferta de crédito e a respectiva diminuição dos *spreads* recebidos pelas instituições financeiras demonstram uma iminente necessidade de ferramentas que possibilitem a verificação antecipada da alocação de recursos. Neste intuito, este trabalho compara métodos estatísticos com Redes neurais na previsão de séries temporais de crédito. As redes neurais têm sido utilizadas com sucesso na previsão de séries financeiras. Isto está relacionado ao fato de que uma rede, composta por uma ou mais camadas ocultas com um número adequado de neurônios, tende a aproximar qualquer função linear ou não-linear contínua, num dado intervalo. A seção 1.1 apresenta redes neurais no processamento de séries temporais. A seção 1.2 estabelece os objetivos do trabalho. A seção 1.3 descreve os passos tomados para alcançar os objetivos e a seção 1.4 fornece uma visão dos capítulos da dissertação.

1.1. REDES NEURAIS NO PROCESSAMENTO DE SÉRIES TEMPORAIS

Rede Neural Artificial é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Inspira-se em alguns aspectos do sistema nervoso central, dentre os quais:

- 1- O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem [1].
- 2- Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido [1].

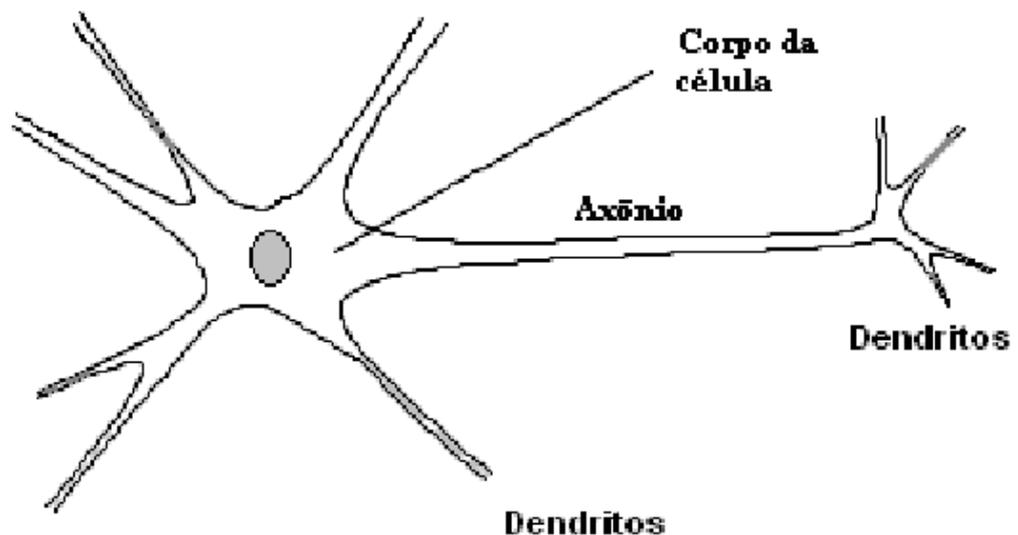


Figura 1-1 Modelo de um neurônio biológico

Arquiteturas de redes neurais diferem pela maneira que os neurônios são organizados, pelos métodos pelos quais os pesos são determinados e pelo domínio das aplicações. Entretanto, em todas as arquiteturas os pesos devem ser aprendidos e não programados [17]. Na maioria das vezes a rede é treinada inserindo-se vetores de exemplos de entrada e saída. A rede MLP (*Multilayer Perceptron*), treinada com o algoritmo de retro-propagação é um modelo adaptativo não linear, sendo uma ferramenta altamente flexível para aplicações em problemas de reconhecimento de padrões, classificação e outros de natureza estática ou estacionária. Em problemas temporais, como previsão de séries temporais, a rede MLP é muito utilizada com

um "artifício" na camada de entrada da rede - o mecanismo da janela. A janela é um mecanismo através do qual se submete à rede os valores prévios da série nos quais é baseada a previsão [17]. Esta janela constitui em realidade a incorporação de "memória" na entrada da rede, permitindo assim à rede MLP o aprendizado do comportamento das séries temporais estacionárias. Com este método, as séries não-estacionárias têm que ser previamente transformadas em estacionárias através de algum método estatístico ou matemático, como a diferenciação. Esta maneira de utilizar a rede neural dá bons resultados na solução de alguns problemas de previsão de séries temporais, mostrando em alguns casos um melhor rendimento que os modelos puramente estatísticos.

Em geral, para que uma rede neural se torne dinâmica, esta deve ter memória. Incorporando memória à estrutura da rede estática, a saída da rede passa a ser função do tempo. Este enfoque de construção de um sistema dinâmico não linear é muito utilizado, pois mostra uma clara separação de responsabilidades: a rede estática toma conta da não linearidade e a memória toma conta do tempo.

O mecanismo da janela, utilizado intuitivamente, pode ser formalizado matematicamente, e generalizado para incorporar memória também nas camadas escondidas, fazendo assim com que a mesma rede MLP seja capaz de aprender o comportamento de sinais temporais não estacionários. Estes tipos de redes são as denominadas "redes neurais com atrasos no tempo".

Outro método de incorporar memória à rede neural, é agregando laços de realimentação das saídas das unidades processadoras até suas entradas, arquitetura conhecida como rede recorrente [1][17].

1.2. OBJETIVOS DA PESQUISA

O objetivo principal deste trabalho é avaliar o desempenho de Redes Neurais Artificiais na predição de demanda de crédito, identificando características de semelhança e diferença com modelos estatísticos.

Este trabalho consistiu das seguintes etapas: a pesquisa bibliográfica das metodologias desenvolvidas para Redes Neurais Artificiais; a seleção do modelo para realizar os testes; e o estudo de casos, em previsão de séries temporais.

A pesquisa compreendeu uma revisão dos modelos propostos na bibliografia para o tratamento de sinais temporais. Verificou-se que a maioria das soluções com redes neurais, dadas a problemas de processamento de séries temporais, faz uso da rede MLP (Multilayer Perceptron) estática, sem laços de realimentação, empregando o algoritmo de retro-propagação para

o aprendizado. No caso de previsão de séries temporais, esta MLP estática é fornecida com "memória" na camada de entrada mediante o denominado "mecanismo da janela" [17].

Este mecanismo de janelamento, em redes neurais artificiais, para tratar problemas dinâmicos, pertence a um tipo geral encontrado na bibliografia, citado como rede neural com atrasos no tempo, ao qual pertencem também a rede TDNN (*Time-Delay Neural Network*) e a rede FIR (*Finite-response Impulse Response*). Nos modelos deste tipo é imprescindível determinar previamente o número de atrasos necessários para a aplicação específica. Uma vez escolhido, este parâmetro permanece fixo. Terminada a pesquisa das metodologias, procurou-se escolher a mais representativa. A rede MLP foi escolhida. Escolhida a rede, desenvolveram-se programas no MATLAB, utilizando o *toolbox* de redes neurais para testar o desempenho da rede MLP.

No estudo de casos de séries temporais, utilizou-se como ferramenta um grupo de técnicas estatísticas conhecida como análise multivariada de dados, onde em muitos casos, estas técnicas multivariadas são um meio de se executar em uma única análise aquilo que antes exigia múltiplas análises usando técnicas univariadas. Dentre as técnicas mais utilizadas, estão regressão múltipla e correlação múltipla. Foram consideradas séries temporais econômicas de periodicidade mensal.

1.3. ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação está organizada em 5 capítulos. O conteúdo dos capítulos seguintes é:

Capítulo 2: Apresenta uma revisão de Séries Temporais

Capítulo 3: Apresenta Redes Neurais. O capítulo termina mostrando o algoritmo de aprendizado de Retro-propagação.

Capítulo 4: Apresenta o trabalho realizado no estudo de casos de previsão de séries temporais.

Capítulo 5: Apresenta as conclusões do trabalho e propostas de trabalhos futuros.

2. SÉRIES TEMPORAIS

Séries temporais são uma sucessão de observações quantitativas de um fenômeno ordenadas no tempo [9]. Dados de séries temporais podem surgir de diferentes áreas de conhecimento como, por exemplo:

- Meteorologia: precipitação, temperatura, velocidade do vento, etc.;
- Economia: índice da Bolsa de Valores, inflação, PIB, crédito, etc.

Uma das características mais relevantes nestes tipos de dados é a dependência entre as observações vizinhas e, em geral, o interesse concentra-se na análise e modelagem desta dependência [9].

Os objetivos de se realizarem análises em séries temporais são variados, indo desde obter previsões sobre valores ou tendências futuras, ou se investigar a relação entre estruturas de duas ou mais séries. Estes objetivos, em geral, podem ser sintetizados da seguinte maneira:

- **Caracterização:** determinar as propriedades da série (padrões de tendência, existência de variação sazonal, regimes não-lineares, autocorrelação, etc.);
- **Previsão:** prever variações passadas e futuras em relação a uma seqüência de medidas;
- **Modelagem:** aplicar ferramentas para modelagem matemática da série.
- **Controle de processos:** controle estatístico de processos e qualidade dos dados;
- **Comparações:** usar a variação numa série para explicar a variação em outra.

Estacionariedade. Uma das principais suposições que se faz em séries temporais é a de que ela seja estacionária (que todas as distribuições unidimensionais são invariantes sob translações do tempo, logo a média e a variância são constantes), porém a maioria das séries temporais apresenta alguma forma de não-estacionariedade (tendência, sazonalidade, etc.). A maioria dos métodos de análise de séries temporais supõe estacionariedade, sendo necessário transformar através de diferenciação sucessiva a série original para se obter uma série estacionária.

Uma série temporal pode ser decomposta nas componentes de **tendência, ciclo, sazonalidade e erro.**

A **tendência** determina o comportamento de longo prazo da série, ou seja, se a série aumenta, diminui ou permanece estável, e qual a velocidade destas mudanças.

Os **ciclos** indicam as oscilações de picos e de vales nas séries, de forma suave e repetida, ao longo da componente de tendência, sendo geralmente percebido ao decorrer de longos períodos (vários anos).

A **sazonalidade** corresponde às oscilações de picos e de vales que ocorrem sempre em um mesmo período, sendo geralmente percebido ao decorrer de períodos de curta duração, menor que 1 anos.

O **erro ou ruído aleatório** representa movimentos ascendentes e descendentes da série após a ocorrência de um efeito de tendência, um efeito cíclico, ou de um efeito sazonal.

As séries temporais apresentam características de valores de forma contínua no tempo, no entanto normalmente os valores são observados a intervalos de tempo igualmente espaçados pré-determinados, configurando uma série temporal discreta.

2.1. TÉCNICAS DESCRITIVAS

Ao se analisar uma ou mais séries temporais a representação gráfica dos dados sequencialmente ao longo do tempo é fundamental e pode revelar padrões de comportamento importantes. Tendências de crescimento (ou decrescimento), padrões cíclicos, alterações estruturais, observações atípicas, são muitas vezes facilmente identificados. Sendo assim, o **gráfico temporal** deve ser sempre a primeira análise de uma série temporal.

2.1.1. Coeficiente de correlação de Pearson

Também chamado de “coeficiente de correlação produto-momento” ou “ r de Pearson” mede o grau de correlação e a direção dessa correlação (se positiva ou negativa) entre duas variáveis de escala métrica. Este coeficiente assume valores entre -1 e 1 e é representado por r .

- $r = 1$ significa uma correlação perfeita positiva entre duas variáveis.
- $r = -1$ significa uma correlação perfeita negativa entre as duas variáveis, isto é, se uma aumenta a outra diminui.
- $r = 0$ significa que as duas variáveis não dependem linearmente uma da outra.

O coeficiente de correlação de Pearson calcula-se utilizando a fórmula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.1)$$

Onde x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n são os valores medidos de ambas as variáveis.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2.2)$$

São as médias aritméticas de ambas as variáveis.

2.1.2. Autocorrelação

Uma ferramenta para se identificar as propriedades de uma série temporal consiste de uma série de quantidades chamadas coeficientes de autocorrelação amostral. Para medirmos a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo. Considerando x_1, \dots, x_{n-1} e x_2, \dots, x_n como duas variáveis, o coeficiente de correlação entre x_t e x_{t+1} é dado por

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)^2 \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_2)^2}} \quad (2.3)$$

onde as médias amostrais são

$$\bar{x}_1 = \sum_{t=1}^{n-1} \frac{x_t}{n-1} \quad \text{e} \quad \bar{x}_2 = \sum_{t=2}^n \frac{x_t}{n-1} \quad (2.4)$$

Como o coeficiente r_1 mede as correlações entre observações sucessivas ele é chamado de coeficiente de autocorrelação. Simplificando a equação utilizando-se a média de todas as observações temos

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(n-1) \sum_{t=1}^n (x_t - \bar{x})^2 / n} \quad (2.5)$$

A equação pode ser generalizada para calcular a correlação entre observações defasadas de k períodos de tempo

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (2.6)$$

Encontramos o coeficiente de correlação de ordem k . Assim como o coeficiente de correlação usual, as autocorrelações não possuem dimensão $-1 < r_k < 1$.

Os coeficientes de autocovariância $\{c_k\}$, definidos por analogia com a fórmula usual de covariância,

$$c_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{n} \quad (2.7)$$

Os coeficientes de autocorrelação são então obtidos como $r_k = c_k/c_0$.

O Correlograma. Um gráfico com os k primeiros coeficientes de autocorrelação como função de k é chamado de Correlograma, sendo uma ferramenta poderosa para identificar características da série temporal.

Séries aleatórias. O Correlograma nos permite determinar se uma série temporal é aleatória ou não. Para uma série completamente aleatória os valores defasados são não correlacionados e espera-se que $r_k = 0$.

Correlação de curto-prazo. Uma série temporal na qual uma observação acima da média “tende” a ser seguida por uma ou mais observações acima da média, similarmente para observações abaixo da média, é dita ter correlação de curto-prazo. Um Correlograma desta série deverá exibir um valor relativamente grande de r_1 seguido por valores que tendem a ficar sucessivamente menores. A partir de certa defasagem k os valores de r_k tendem a ser aproximadamente zero.

Correlação negativa. Se os valores de uma série temporal tendem a se alternar acima e abaixo de um valor médio, o Correlograma desta série também tende a se alternar. O valor de r_1

será negativo enquanto o valor de r_2 será positivo já que as observações defasadas de dois períodos tendem a estar do mesmo lado da média.

Séries não estacionárias. Para uma série temporal com tendência os valores de r_k não decairão para zero a não ser em defasagens grandes. Isto ocorre porque uma observação de um lado da média tende a ser seguida por um grande número de observações do mesmo lado (devido à tendência). Neste caso, pouca ou nenhuma informação pode ser extraída do Correlograma já que a tendência dominará outras características. Na verdade, a função de autocorrelação só tem um significado para séries estacionárias, sendo assim qualquer tendência deve ser removida antes do cálculo de $\{r_k\}$.

Variação sazonal. Um padrão sazonal é em geral facilmente identificado no Correlograma. Se uma série temporal tem flutuações sazonais o Correlograma irá exibir oscilações na mesma frequência.

Observações discrepantes. Se uma série temporal contém uma ou mais observações discrepantes (“*outliers*”) o Correlograma pode ser seriamente afetado. No caso de uma única observação discrepante o gráfico de x_t contra $x_t + k$ terá pontos extremos, podendo alterar os coeficientes de correlação para zero. Com dois valores discrepantes o efeito pode ser ainda pior, além de gerar uma correlação espúria quando k é igual à distância entre os valores.

2.2. PROGNÓSTICO

Os métodos de prognóstico podem ser classificados em três grupos [12]:

Subjetivo. É uma metodologia que utiliza dados qualitativos para obter resultados que permitam efetuar julgamentos intuitivos a partir do desenvolvimento de cenários capazes de indicar valores para as variáveis desconhecidas do problema.

Univariável. Processos que compreende a maior parte dos métodos de previsão de séries temporais, considera somente uma única série para a realização dos prognósticos. As previsões decorrentes da aplicação de métodos univariados podem estar relacionadas apenas com as informações contidas na série histórica de interesse (métodos baseados na estatística clássica) ou também, outras supostamente relevantes e que não estão contidas na série analisada (métodos baseados na estatística bayesiana).

Multivariável. É um processo quantitativo em que o prognóstico da variável em estudo depende do conhecimento do comportamento temporal de outras variáveis do problema. Para identificar o método que melhor se adapta ao problema em estudo, é comum considerar a aplicação de vários métodos de forma a gerar cenários que permitam estabelecer a modelagem matemática que melhor representa a série de valores.

2.3. MÉTODOS SIMPLES DE PREVISÃO DE SÉRIES TEMPORAIS

Os métodos simples efetuam a previsão do valor futuro da série temporal pelo alisamento das observações passadas da série de interesse [18]. Assumindo que os valores extremos da série representam flutuações aleatórias [9], o propósito desses métodos consiste em identificar o padrão básico presente nos dados históricos e, então, usar esse padrão para prever valores futuros.

Em [9] a grande popularidade desses métodos é associada à simplicidade, à eficiência computacional e à razoável previsão obtida. Entre os métodos simples de previsão destacam-se o da *Média Móvel*, o *Alisamento Exponencial Simples*, o *Alisamento Exponencial Linear* e o *Alisamento Exponencial Sazonal e Linear de Winter* [18], os quais são apresentados na seqüência.

2.3.1. Média móvel

Esse método considera como previsão para o período futuro a média das observações passadas recentemente [18].

Em [9] a média móvel para o período de tempo t é definida por

$$x_t = \frac{x_{t-1} + x_{t-2} + \dots + x_{t-n}}{n} \quad (2.8)$$

Onde n representa o número de observações incluídas na média x_t .

O termo *média móvel* é utilizado porque à medida que a próxima observação se torna disponível, a média das observações é recalculada, incluindo essa observação no conjunto de observações e desprezando a observação mais antiga [9].

Em [18] afirma-se que quanto maior o número de observações incluídas na média móvel, maior o efeito de alisamento na previsão. Assim, caso a série temporal apresente muita aleatoriedade ou pequenas mudanças nos padrões, um número maior de valores podem ser utilizados no cálculo da média móvel, obtendo-se uma previsão mais alisada. Entretanto, se houver pouca flutuação aleatória nos dados ou mudança significativa no padrão da série, um número me-

nor de observações deve ser incluído no conjunto de valores empregado na determinação da média para que se possa reagir a essas alterações mais rapidamente.

2.3.2. Alisamento exponencial simples

A princípio, o método conhecido como Alisamento Exponencial Simples se assemelha ao da Média Móvel por extrair das observações da série temporal o comportamento aleatório pelo alisamento dos dados históricos. Entretanto, a inovação introduzida pelo Alisamento Exponencial Simples advém do fato de este método atribuir pesos diferentes a cada observação da série. Enquanto que na Média Móvel as observações usadas para encontrar a previsão do valor futuro contribuem em igual proporção para o cálculo dessa previsão, no Alisamento Exponencial Simples as informações mais recentes são evidenciadas pela aplicação de um fator que determina essa importância [18].

Segundo Wheelwright [18], o argumento para o tratamento diferenciado das observações da série temporal é fundamentado na suposição de que as últimas observações contêm mais informações sobre o futuro e, portanto, são mais relevantes para a previsão.

Em [18] especifica-se o método Alisamento Exponencial Simples através da equação

$$F_{t+1} = \alpha x_t + (1 - \alpha)F_t \quad (2.9)$$

Onde F_{t+1} representa a previsão no tempo $t + 1$ e

α é o peso atribuído à observação x_t , $0 \leq \alpha \leq 1$.

De acordo com [9], quanto menor o valor da constante, mais estáveis serão as previsões, visto que a utilização de baixo valor implica na atribuição de peso maior às observações passadas e, conseqüentemente, qualquer flutuação aleatória no presente contribui com menor importância para a obtenção da previsão. Contudo, não há metodologia que oriente quanto à seleção de um valor apropriado, sendo normalmente encontrado por tentativa e erro [18].

2.3.3. Alisamento exponencial linear

Quando o método Alisamento Exponencial Simples é aplicado na previsão de séries temporais que apresentam tendência entre as observações passadas, os valores prognosticados su-

perestimam (ou subestimam) os valores reais [9]. Desta forma, a acuidade das previsões fica prejudicada.

Para evitar esse erro sistemático, o método Alisamento Exponencial Linear foi desenvolvido procurando reconhecer a presença de tendência na série de dados [18]. O valor da previsão obtido através deste método é alcançado pela aplicação da equação abaixo:

$$F_{t+m} = S_t + mT_t \quad (2.10)$$

Onde S_t corresponde à previsão no tempo t ;

T_t representa a componente de tendência e

m é o horizonte de previsão.

$$S_t = \alpha x_t + (1-\alpha)(S_{t-1} + T_{t-1}) \quad (2.11)$$

Onde α é o peso atribuído à observação x_t , $0 \leq \alpha \leq 1$.

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (2.12)$$

Onde β é o coeficiente de alisamento, análogo a α .

2.3.4. Alisamento Exponencial Sazonal e Linear de Winter

Este método produz resultados similares ao Alisamento Exponencial Linear, sendo, no entanto, capaz de manipular séries temporais que além de apresentarem tendência nos dados, apresentem também sazonalidade [18].

As equações abaixo definem o referido método de previsão.

$$S_t = \alpha \frac{x_t}{I_{t-l}} + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (2.13)$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (2.14)$$

$$I_t = \gamma \frac{x_t}{S_t} + (1 - \gamma)I_{t-l} \quad (2.15)$$

$$F_{t+m} = (S_t + mT_t)I_{t-l+m} \quad (2.16)$$

Onde I_t corresponde ao alisamento do fator de sazonalidade $\frac{x_t}{S_t}$;

l é o intervalo da sazonalidade e γ corresponde ao peso atribuído ao fator de sazonalidade.

2.4. MÉTODOS AVANÇADOS DE PREVISÃO DE SÉRIES TEMPORAIS

No universo dos métodos de previsão de séries temporais mais complexos encontram-se os modelos Autoregressivo e Médias Móveis (AR, MA e ARMA), modelos Autoregressivo Integrado de Médias Móveis (ARIMA), Filtros de Kalman e AEP, modelos ARARMA de Parzen, modelos ARMA Multivariáveis (MARMA), entre outros [18]. Os métodos assim classificados obtêm a previsão de algum valor futuro da série temporal pela combinação dos valores reais passados e/ou dos erros ocorridos.

Os modelos AR, MA, ARMA e ARIMA são descritos a seguir.

2.4.1. Modelos Autoregressivo e de Médias Móveis

Em[18] descreve-se três procedimentos capazes de representar as observações de uma série temporal estacionária: modelo *Autoregressivo* (AR), modelo de *Médias Móveis* (MA) e o modelo *Autoregressivo e de Médias Móveis* (ARMA).

a) Modelo Autoregressivo

A especificação de um modelo Autoregressivo (AR) é dada pela equação.

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + e_t \quad (2.17)$$

onde x_t corresponde à observação da série temporal no tempo t ;

ϕ_p corresponde ao parâmetro do modelo AR de ordem p e

e_t representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.

Caso as observações da série temporal possam ser representadas pela equação, a ordem do modelo puder ser determinada e os parâmetros estimados, é possível prever o valor futuro da série em análise [18].

b) Modelo de Médias Móveis

Um modelo de Médias Móveis (MA) fica definido conforme a equação.

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.18)$$

Onde e_t representa o erro de eventos aleatórios que não podem ser explicados pelo modelo e θ_q corresponde ao parâmetro do modelo MA de ordem q .

As equações acima são similares, exceto pelo fato de que o valor previsto para a observação de α depende dos valores dos erros observados em cada período passado, ao invés das observações propriamente ditas [18].

c) Modelo Autoregressivo e de Médias Móveis

Wheelwright e Makridakis [18] especificam o modelo misto Autoregressivo e de Médias Móveis (ARMA) através da equação 2.19, como sendo a combinação dos modelos AR e MA.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (2.19)$$

Analisando a equação é possível verificar que os modelos ARMA relacionam os valores futuros com as observações passadas, assim como também com os erros passados apurados entre os valores reais e os previstos.

2.4.2. Modelo de Box e Jenkins

O destaque atribuído ao modelo de George Box e Gwilyn Jenkins, que também pode ser incluído nesta classificação, é devido principalmente a sua fundamentação teórica, sendo a princípio capaz de manipular séries temporais de qualquer natureza.

O método de Box e Jenkins consiste na busca de um modelo ARIMA (*Autoregressive Integrate Moving Average*) que represente o processo estocástico gerador da série temporal, a partir de um modelo ARMA aplicável na descrição de séries temporais estacionárias, estendendo esse conceito para séries temporais não-estacionárias.

Genericamente, um processo ARIMA(p,d,q) pode ser representado pela equação.

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (2.20)$$

sendo $w_t = x_t - x_{t-d}$,

onde ϕ_p e θ_q são os parâmetros dos processos Autoregressivo e de Média Móvel de ordem p e q (ARMA(p,q));

e_t corresponde ao erro de eventos aleatórios que não podem ser explicados pelo modelo e d equivale ao grau de homogeneidade não-estacionária.

O objetivo da *identificação* é determinar os valores de p , d e q do modelo $ARIMA(p,d,q)$ [18]. Inicialmente, a série temporal X^T é diferenciada para se obter uma série estacionária. Com isso, o processo fica reduzido a um modelo $ARMA(p,q)$. Em seguida, a ordem do processo $ARMA$ é identificada pela análise dos coeficientes de autocorrelação e autocorrelação parcial. Ainda nesta etapa são efetuadas estimativas preliminares dos parâmetros do modelo identificado.

Após a identificação do modelo que seja uma representação adequada do mecanismo gerador da série, a *estimação* dos parâmetros desse modelo é efetuada. Os parâmetros do processo AR são estimados através de métodos de regressão; caso o processo MA esteja envolvido, a estimação dos parâmetros deste modelo é obtida pela aplicação de algum algoritmo de otimização não-linear.

Estimado o modelo, a *verificação* de sua habilidade em representar os fenômenos observáveis da série temporal é confirmada pela análise dos erros do modelo proposto. Caso a inadequação fique evidenciada, o ciclo de identificação, estimação e verificação é novamente aplicado, até que a representação apropriada seja encontrada.

Após a validação do modelo, a *previsão* dos valores futuros da série temporal modelada pode, enfim, ser obtida.

2.5. CONSIDERAÇÕES SOBRE PREVISÃO DE SÉRIES TEMPORAIS

Makridakis e Wheelwright [18] investigaram o poder preditivo de vários métodos comumente utilizados na previsão de séries temporais. Pela comparação dos resultados alcançados, identificaram situações onde esses métodos apresentam melhor desempenho e definiram alguns critérios para a seleção dos procedimentos de previsão, pela confrontação dos objetivos a atingir.

O estudo desenvolvido por tais pesquisadores constatou que o incremento da complexidade e da sofisticação estatística dos métodos de previsão de séries temporais não implica, necessariamente, em uma melhora na acuidade da previsão. "*Métodos simples de previsão podem apresentar desempenho extremamente satisfatório sob certas condições*" [18]. Além disso,

métodos de previsão menos complexos normalmente permitem alcançar total compreensão de suas suposições e limitações, e de interpretação de seus resultados. Assim, antes de se adotar um método de previsão mais complexo, é necessário avaliar os benefícios que um método dessa natureza pode gerar em relação ao custo de sua aplicação.

Ainda segundo os autores [18], como a acuidade de uma previsão fica determinada não apenas pelo horizonte de previsão especificado, mas também pelas características das observações da série temporal investigada, a otimização desse critério pode ser alcançada com a aplicação de mais de um método de previsão. A combinação de previsões ou a verificação da consistência desses valores permite aumentar a confiabilidade da previsão e reduzir a possibilidade de grandes desvios.

Mentzer e Cox também analisaram a frequência de aplicação de diferentes métodos de previsão de séries temporais considerando como critérios o horizonte de previsão, familiaridade com os procedimentos preditivos e nível de satisfação decorrente do emprego desses procedimentos. Em relação ao horizonte de previsão, o Alisamento Exponencial Simples e a Média Móvel são mais frequentemente usados para períodos curtos, menos para médio prazo e menos ainda para longo prazo, comprovando evidências empíricas que indicam melhor desempenho destes métodos para horizonte de previsão menores [18]. Independente do horizonte de previsão adotado, pouca aplicabilidade do método de Box e Jenkins pode ser observada.

Após identificar a ordem e estimar eficientemente os parâmetros de um modelo é necessário verificar sua adequação antes de utilizá-lo, por exemplo, para fazer previsões.

Podem-se fazer testes sobre ajustamentos, que consistem em incluir parâmetros extras no modelo e verificar sua significância estatística. No caso de modelos ARMA deve-se incluir um parâmetro de cada vez para evitar o problema de cancelamento de raízes.

2.6. ANÁLISES DOS RESÍDUOS

Após um modelo ter sido ajustado a uma série temporal deve-se verificar se ele fornece uma descrição adequada dos dados [11]. Assim como em outros modelos estatísticos a idéia é verificar o comportamento dos resíduos, onde resíduo é a parte de uma variável dependente não explicada por uma técnica multivariada. Associado a métodos de dependência que tentam prever a variável dependente, o resíduo representa a parte inexplicada da mesma. Os resíduos são uma medida do ajuste preditivo para uma única observação, calculada como a diferença

entre valores reais e previstos da variável dependente. Considera-se que os resíduos têm média zero e uma variância constante. Desempenham um papel chave para verificar se as suposições subjacentes da regressão foram atendidas, assim como uma ferramenta de diagnóstico na identificação de observações atípicas e influentes [11].

Para os modelos vistos aqui o valor ajustado é a previsão um passo a frente de modo que o resíduo fica definido como o erro de previsão um passo a frente. Por exemplo, em um modelo AR(1) se $\hat{\alpha}$ é a estimativa do coeficiente auto-regressivo então o valor ajustado no tempo t é $\hat{\alpha}x_{t-1}$ e o resíduo correspondente é $e_t = x_t - \hat{\alpha}x_{t-1}$.

Se o modelo tiver um bom ajuste espera-se que os resíduos se distribuam aleatoriamente em torno de zero com variância aproximadamente constante e sejam não correlacionados. Se a variância dos resíduos for crescente uma transformação logarítmica nos dados pode ser apropriada. O fenômeno de “não-constância” na variância é denominado de volatilidade na literatura de séries temporais e pode ser tratado através de transformações nos dados (por exemplo, transformações de Box-Cox). Além disso, em modelos de séries temporais os resíduos estão ordenados no tempo e é portanto natural tratá-los também como uma série temporal. É particularmente importante que os resíduos de um modelo estimado sejam serialmente (ou seja, ao longo do tempo) não correlacionados. Evidência de correlação serial nos resíduos é uma indicação de que uma ou mais características da série não foram adequadamente descritas pelo modelo. Conseqüentemente, duas maneiras óbvias de verificar a adequação do modelo consistem em representar graficamente os resíduos e o seu correlograma. O gráfico temporal poderá revelar a presença de dados discrepantes, efeitos de autocorrelação ou padrões cíclicos enquanto que o correlograma permite uma análise mais detalhada da estrutura de autocorrelação indicando possíveis termos faltantes no modelo. Ou seja, assim como em outros modelos estatísticos, a idéia é que os resíduos poderão identificar características que não foram adequadamente modeladas. Por exemplo, autocorrelações residuais significativas nas defasagens 1 ou 2, ou em defasagens sazonais (por exemplo, 12 para dados mensais) são uma indicação de que mais termos médias móveis devem ser incluídos no modelo. Por outro lado, um valor da equação r_k ligeiramente fora dos limites de confiança em defasagens sem significado óbvio (por exemplo, $k=5$) não é indicação suficiente para se rejeitar o modelo. O mesmo comentário vale para as autocorrelações parciais dos resíduos no que diz respeito a inclusão de termos auto-regressivos (sazonais e não sazonais).

Testes sobre os resíduos - Em vez de olhar para as autocorrelações residuais individualmente. Pode-se testar se um grupo de autocorrelações é significativamente diferente de zero por meio das chamadas estatísticas Q. A estatística Q testa correlação serial e indica se o quadrado do retorno possui uma autocorrelação relativamente maior do que a série sem o quadrado. Este fato pode ser encarado como indicativo de forte heterocedasticidade.¹

2.7. PREVISÃO

Uma das formas de utilização de um modelo ajustado é para fazer previsões de valores futuros. Assim, se t é o período corrente estamos interessados em prever os valores de X_{t+1}, X_{t+2}, \dots . A previsão de X_{t+k} , para $k = 1, 2, \dots$ será denotada por $\hat{x}_t(k)$ e é definida como a esperança condicional de X_{t+k} dados todos os valores passados, isto é $\hat{x}_t(k) = E(X_{t+k} | x_t, x_{t-1}, \dots)$. [11]

A equação acima é chamada de função de previsão e o inteiro k é chamado de horizonte de previsão. Pode-se mostrar que esta previsão tem o menor erro quadrático médio.

Se temos uma série temporal observada x_1, \dots, x_n as previsões podem ser feitas dentro do período amostral e comparadas com os valores observados. Esta é uma prática bastante comum para checar o desempenho preditivo do modelo. A diferença entre os valores previstos e observados, $\hat{x}_t(k) - x_{t+k}$, é chamada de erro de previsão k passos a frente e será denotado por e_{t+k} .

2.8. PERFORMANCE PREDITIVA

A idéia de verificar a adequação de um modelo em termos dos erros de previsão um passo a frente já foi apresentada anteriormente. Na prática é preciso verificar se os resíduos se comportam de maneira aleatória em torno de zero e com variância aproximadamente constante, além de serem não correlacionados ao longo do tempo [11].

Além disso, dois ou mais modelos podem ser comparados segundo a sua performance preditiva, ou seja construindo-se medidas baseadas nos erros de previsão. A maioria dos métodos de previsão baseia-se na idéia de minimizar somas de quadrados ou de valores absolutos dos erros de previsão e esta é também uma medida usada para comparar a adequação de modelos alternativos. A idéia então é comparar o erro quadrático médio P e $2t / (n - m)$ ou erro abso-

¹ Fenômeno estatístico que ocorre quando o modelo de hipótese matemático apresenta variâncias para Y e $X(X_1, X_2, X_3, \dots, X_n)$ diferentes para todas as observações.

luto médio $P |et|/(n - m)$ para diferentes modelos, onde m é o número de parâmetros a serem estimados.

Uma estratégia simples de se fazer previsões consiste em tomar a observação mais recente como a melhor previsão de um valor futuro da série, isto é $\hat{x}_t(1) = x_t$. Veja que esta é a previsão 1 passo a frente de um passeio aleatório. Assim, uma forma de medir a capacidade preditiva de um modelo consiste em comparar seus erros de previsão com aqueles do passeio Aleatório. Isto pode ser feito através da chamada estatística U de Theil definida como

$$U = \frac{\sum_{t=1}^{n-1} (x_{t+1} - \hat{x}_t(1))^2}{\sum_{t=1}^{n-1} (x_{t+1} - x_t)^2} \quad (2.21)$$

Os valores maiores do que 1 são uma indicação de que globalmente os erros de previsão tendem a ser grandes em relação aos erros de um passeio Aleatório. Esta não é uma boa característica, pois deseja-se que o valor de U fosse sempre menor do que 1. Vale notar também que neste caso os erros de previsão estão sendo avaliados independentemente da escala dos dados. Finalmente, vale notar que todas as medidas de capacidade preditiva citadas podem ser estendidas para erros de previsão k passos a frente. Outra prática comum em séries temporais consiste em estimar o modelo excluindo algumas observações finais e usar o modelo estimado para fazer previsões. Neste caso as previsões podem ser comparadas com os valores observados.

2.9. CRITÉRIOS DE INFORMAÇÃO

Em muitas aplicações vários modelos podem ser julgados adequados em termos do comportamento dos resíduos. Uma forma de discriminar entre estes modelos competidores é utilizar os chamados critérios de informação que levam em conta Previsões, limites inferior (li) e superior (ls) e semi-amplitudes dos intervalos de previsão.

Assim, um modelo com mais parâmetros pode ter um ajuste melhor mas não necessariamente será preferível em termos de critério de informação. A regra básica consiste em selecionar o modelo cujo critério de informação calculado seja mínimo.

A regra mais utilizada em séries temporais é o chamado critério de informação de Akaike, denotado por AIC. A definição mais comumente utilizada é $AIC = -2 \log \text{verossimilhança maximizada} + 2m$ onde m é o número de parâmetros (em modelos ARMA(p, q) $m = p + q + 1$). Para dados normalmente distribuídos e usando-se estimativas de máxima verossimilhança para os parâmetros pode-se mostrar que

$$AIC = n \log(\hat{\sigma}^2 \varphi) + 2m \text{ onde } \hat{\sigma}^2 \varphi = (1/n) P \hat{\varphi}^2 t. \quad (2.22)$$

Existem outros critérios de informação que são basicamente modificações do AIC na forma de penalizar a inclusão de parâmetros extras.

3. REDES NEURAIS ARTIFICIAIS

3.1. HISTÓRICO

O estudo científico de Redes Neurais iniciou-se com a publicação do artigo de McCulloch e Pitts em 1943 onde os mesmos descreveram um cálculo lógico das redes neurais que unificava os estudos de neurofisiologia e da lógica matemática. Eles assumiam que o seu modelo de neurônio seguia uma lei do “Tudo ou nada”. Mostraram que com um número suficiente de neurônios e com suas conexões sinápticas ajustadas e operando de forma síncrona, uma rede neural realizaria a computação de qualquer função computável [1].

O próximo desenvolvimento significativo veio em 1949 com a publicação do livro de Donald Hebb, *The Organization of Behavior*, no qual foi apresentada pela primeira vez uma regra explícita de aprendizagem fisiológica para a modificação sináptica. Hebb propôs que a conectividade do cérebro é continuamente modificada conforme vai aprendendo tarefas funcionais diferentes e que agrupamentos neurais são criados por tais modificações. Em seu postulado de aprendizagem, Hebb afirma que a eficiência de uma sinapse variável entre dois neurônios é aumentada pela ativação repetida de um neurônio causada pelo outro neurônio, através daquela sinapse. Acredita-se que o artigo de Rochester, Holland, Haibt e Duda (1956) seja a primeira tentativa de se usar simulação computacional para testar uma teoria neural com base no postulado de aprendizagem de Hebb, Os resultados de simulação relatados no referido artigo mostram que se deve adicionar inibição para que a teoria realmente funcione.

Minsky contribuiu de forma significativa para o desenvolvimento do estudo de redes neurais, primeiro em 1954 com uma tese de doutoramento em redes neurais na University of Princeton, intitulada “Theory of Neural-Analog Reinforcement Systems and Its Application to the Brain-Model Problem”, depois em 1961 com um excelente artigo sobre IA intitulado “Steps Toward Artificial Intelligence” e finalmente em 1967 com a publicação do livro *Computation: Finite and Infinite Machines*, onde de forma clara, estendeu os resultados de 1943 de McCulloch e Pitts e os colocou no contexto da teoria dos autômatos e da teoria da computação.

Em 1958 uma nova abordagem para o problema de reconhecimento de padrões foi introduzida por Rosenblatt em seu trabalho sobre o *perceptron*, com o método de aprendizagem supervisionada (teorema da convergência do perceptron). Logo após, em 1960, Widrow e Hoff introduziram o algoritmo do mínimo quadrado médio (LMS, *Least Mean-Square*) e o usaram para formular o Adaline (*adaptive linear element*). A diferença entre o perceptron e o Adaline está no procedimento de aprendizagem.

Na década de 70, a visão conexionista ficou “adormecida”. Em grande parte devido à repercussão do trabalho de Minsky e Papert, no qual se afirmava que o *perceptron* era limitado a resolver problemas linearmente separáveis, ou seja, problemas cuja solução pode ser obtida dividindo-se o espaço de entrada em duas regiões através de uma reta e que não valia a pena insistir nisso. Em 1982, John Hopfield publicou um artigo, mostrando a relação entre redes recorrentes auto-associativas e sistemas físicos, o que também abriu espaço para a utilização de teorias correntes da física para estudar tais modelos. Não obstante, a descrição do algoritmo de treinamento *backpropagation* ou retro-propagação alguns anos mais tarde, mostrou que a visão de Minsky e Papert sobre o *perceptron* era bastante pessimista. As redes neurais de múltiplas camadas são capazes de resolver problemas não-lineares, ou “difíceis de aprender”. A partir de meados da década de 80, houve nova explosão de interesse pelas redes neurais na comunidade internacional [1].

3.2. MODELO DE NEURÔNIO ARTIFICIAL

O neurônio artificial foi projetado para imitar as características de um neurônio biológico. A Figura 3.1 mostra um modelo de um neurônio k , identificando os três elementos básicos para o projeto de redes neurais: um conjunto de sinapses, cada qual caracterizada por um peso ou força própria (w_{ki}); um somador para somar os sinais de entrada, ponderado pelas respectivas sinapses do neurônio, e uma função de ativação (φ), possivelmente não linear.

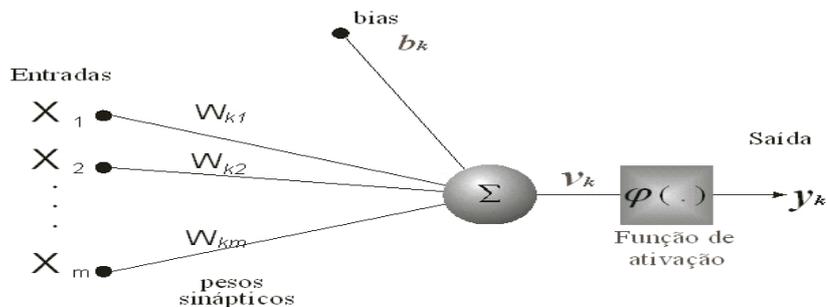


Figura 3-1 Modelo não-linear de um neurônio – Haykin 2001

Podem ser observados na figura (3.1), as entradas (x_i) do neurônio e o resultado do processamento neuronal, a saída (y_k), bem como o sinal de bias (b_k).

O neurônio pode ser representado matematicamente pelas seguintes expressões:

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (3.1)$$

$$y_k = \varphi(v_k) \quad (3.2)$$

As funções de ativação de cada neurônio podem ser, dentre outros, do tipo degrau, linear ou sigmoidal, e suas representações gráficas podem ser observadas na Figura 3.2.

3.3. FUNÇÕES DE ATIVAÇÃO

A função de ativação $\varphi(v)$ define a saída de um neurônio em termos do campo local induzido v . Esta função pode ter três tipos básicos:

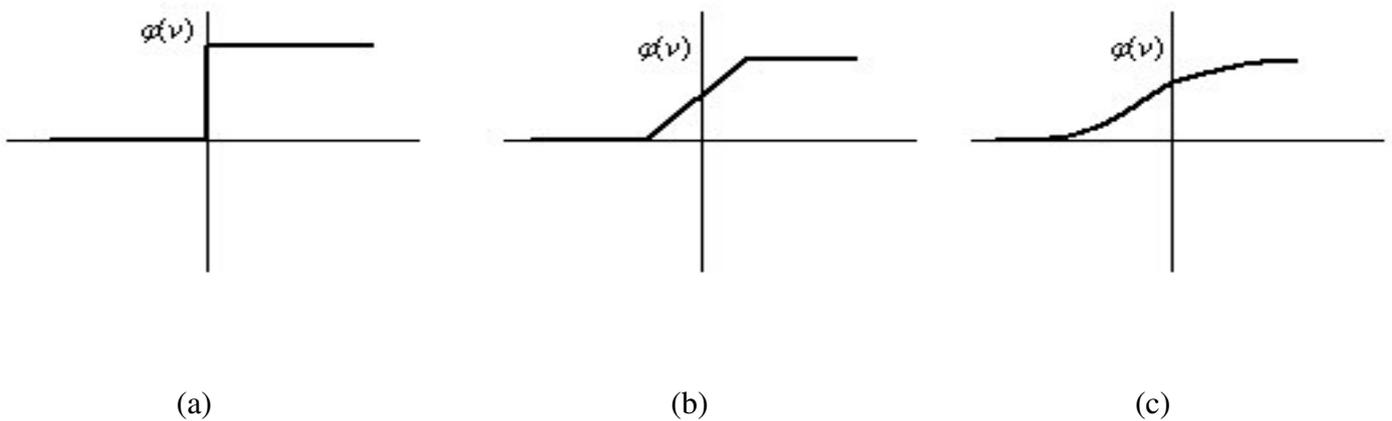


Figura 3-2 Exemplos de tipos de função de ativação

a) Função de Limiar. Para este tipo de função descrito na figura 3.2a temos

$\varphi(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases}$ Neste modelo, a saída de um neurônio assume o valor 1, se o campo local induzido daquele neurônio é não negativo, e 0 caso contrário. Esta definição descreve a propriedade tudo-ou-nada do modelo de McCulloch-Pitts.

b) Função Linear por partes. Para este tipo de função descrito na figura 3.2b temos

$$\varphi(v) = \begin{cases} 1 & \text{se } v \geq +\frac{1}{2} \\ v & \text{se } +\frac{1}{2} > v > -\frac{1}{2} \\ 0 & \text{se } v \leq -\frac{1}{2} \end{cases}$$

Esta forma de função de ativação pode ser vista como uma aproximação de um amplificador não-linear. As seguintes situações podem ser vistas como formas especiais da função linear por partes.

- Se a região linear da operação é mantida sem entrar em saturação, surge um combinador linear.
- A função linear por partes se reduz a função de limiar, se o fator de amplificação da região linear é feito infinitamente grande.

c) Função Sigmóide. Para este tipo de função descrito na figura 3.2c temos

$$\varphi(v) = \frac{1}{1 + \exp(-av)} \quad (3.3)$$

A função sigmóide, cujo gráfico tem a forma de S, é a função de ativação mais utilizada na construção de redes neurais, sendo definida como uma função estritamente crescente que exibe um balanceamento adequado entre comportamento linear e não-linear. Algumas vezes é desejável que a função de ativação se estenda de -1 a +1, assumindo uma forma anti-simétrica em relação à origem. Para a forma correspondente de uma função sigmóide logística, podemos utilizar a função tangente hiperbólica, definida por

$$\varphi(v) = \tanh(v) \quad (3.4)$$

3.4. ARQUITETURAS DE REDES NEURAIIS ARTIFICIAIS

As topologias de redes neurais estão divididas em não-recorrentes e recorrentes.

As redes neurais *não-recorrentes* são aquelas que não possuem realimentação de suas saídas para suas entradas, isto é, "sem memória". A estrutura das redes neurais não-recorrentes é em camadas, podendo estas redes neurais ser formadas por uma (rede neural de camada única) ou mais camadas (rede neural multicamada). Perceptron multicamadas contém um conjunto de neurônios de entrada, uma camada de saída e uma ou mais camadas escondidas. A entrada não é considerada uma camada da rede, pelo fato de apenas distribuir os padrões. As camadas que não possuem ligações diretas nem com a entrada, nem com a saída são chamadas de ca-

mas escondidas. Não existem conexões ligando um neurônio de uma camada a outro de uma camada prévia, nem a um neurônio da mesma camada. As redes neurais multicamadas, são utilizadas neste trabalho e serão estudadas com mais detalhes na seção 3.6.

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

As redes neurais recorrentes, também chamadas de redes neurais com memória por possuírem realimentação, respondem a estímulos dinamicamente, isto é, após aplicar uma nova entrada, a saída é calculada e então realimentada para modificar a entrada. Para as redes estáveis, este processo é repetido várias vezes, produzindo pequenas mudanças nas saídas, até estas ficarem constantes. Todavia, as redes neurais recorrentes nem sempre são estáveis, mesmo com entradas constantes. O fato de não se conseguir prever quais redes seriam estáveis foi um problema que preocupou os pesquisadores até o início da década de 80, quando Cohen e Grossberg provaram um poderoso teorema para definir quando as redes neurais eram estáveis [13]. Este teorema determina que para as redes neurais recorrentes alcancem um estado estável é necessário que possuam conexões simétricas, melhor dizendo, $W_{ij}=W_{ji}$, sendo $W_{ii}=0$. Contribuições importantes também foram dadas por John Hopfield [14], tanto para a teoria como para a prática, sendo que algumas configurações passaram a ser chamadas de redes de Hopfield em sua homenagem. Hinton e Sejnowski [15] também tiveram seus méritos, introduzindo regras gerais de treinamento para redes, denominadas por eles de máquina de Boltzmann devido ao cálculo das saídas dos neurônios artificiais serem efetuado através de probabilidade com distribuição de Boltzmann.

3.5. TIPOS DE TREINAMENTO

De todas as propriedades interessantes das redes neurais artificiais, uma das que mais chamam a atenção é a habilidade de aprender. Ao invés de especificar todos os detalhes de uma computação tem-se a possibilidade de treinar uma rede para fazer esta computação. Isto significa que se podem tratar problemas onde regras apropriadas são muito difíceis de conhecer a priori.

O objetivo do treinamento de uma rede neural é fazer com que a aplicação de um conjunto de entradas produza um conjunto de saídas desejado ou no mínimo um conjunto de saídas consistentes. O treinamento é realizado pela aplicação sequencial ou não dos vetores de entradas (e em alguns casos também os de saída), enquanto os pesos da rede são ajustados de acordo com um procedimento de treinamento pré-determinado. Durante o treinamento, os pesos da rede geralmente convergem gradualmente para determinados valores, tal que a aplicação dos vetores de entrada produza as saídas necessárias.

Os procedimentos de treinamento que levam as redes neurais a aprender determinadas tarefas podem ser classificados em duas classes de treinamento: supervisionado e não-supervisionado.

O treinamento *supervisionado* necessita de um par de vetores composto do vetor de entrada e do vetor alvo que se deseja como saída. Juntos, estes vetores são chamados de par de treinamentos ou vetor de treinamento, salientando que a rede é treinada com vários vetores de treinamento.

O procedimento de treinamento funciona da seguinte forma: o vetor de entrada é aplicado. A saída da rede é calculada e comparada com o correspondente vetor alvo. O erro encontrado é então realimentado através da rede e os pesos são atualizados de acordo com um algoritmo determinado a fim de minimizar este erro. Este processo de treinamento é repetido até que o erro para os vetores de treinamento tenha alcançado níveis aceitáveis.

O treinamento *não supervisionado*, por sua vez, não requer vetor alvo para as saídas e, obviamente, não faz comparações para determinar a resposta ideal. O conjunto de treinamento modifica os pesos da rede de forma a produzir saídas que sejam consistentes, isto é, tanto a apresentação de um dos vetores de treinamento, como a apresentação de um vetor que é suficientemente similar, irá produzir o mesmo padrão nas saídas. O processo de treinamento extrai as propriedades estatísticas do conjunto de treinamento e agrupa os vetores similares em classes. A aplicação de um vetor de uma determinada classe à entrada da rede irá produzir um vetor de saída específico, mas não existe maneira de se determinar, antes do treinamento, qual o padrão que será produzido na saída para um vetor de entrada de uma determinada classe. Desta forma, a saída de algumas redes neurais deve ser transformada em uma forma interpre-

tável após o processo de treinamento, o que é um simples problema de identificação das relações entrada-saída estabelecido pela rede.

No que diz respeito aos algoritmos de treinamento usados, existe uma grande variedade, tanto para o treinamento supervisionado, como para o não supervisionado. Muitos deles baseiam-se no algoritmo proposto por D. O. Hebb em 1961, entretanto, os algoritmos mais eficientes foram desenvolvidos nos últimos trinta anos, baseados nos trabalhos de Rosenblatt (1962), Windrow (1959) e Windrow e Hoff (1960) [13]. Entre estes, o mais difundido com certeza foi o algoritmo de retro propagação ("*Backpropagation*") [16], [13], o qual é utilizado neste trabalho.

3.6. REDES NEURAIIS ARTIFICIAIS DE MÚLTIPLAS CAMADAS

Uma rede neural do tipo perceptron multicamada, com ligações unidirecionais entre os neurônios (na direção das entradas para as saídas), é comumente chamada de rede neural *feedforward*. As ligações entre neurônios, ou entre entradas e neurônios, têm, cada uma, um peso sináptico correspondente, denotado por w_{ij} (conexão do j -ésimo nó, ou j -ésima entrada, para o i -ésimo nó). A saída do j -ésimo neurônio, ou a j -ésima entrada, é multiplicada por w_{ij} e é uma entrada para o i -ésimo nó. A chamada excitação ou entrada resultante do i -ésimo nó é dada por $\text{excitação } i = \sum_j w_{ij} \text{ entrada } j + b_k$ onde b_k pode ser interpretado como um termo de intercepto. Vemos que uma rede *feedforward* com apenas um neurônio, com função de ativação linear, nada mais é do que uma regressão linear padrão.

No entanto, no caso geral, cada neurônio tem uma função não-linear correspondente. Ou seja, a saída do neurônio i é dada por $\text{saída } i = \varphi(\text{excitação } i)$ onde φ é uma função não-linear diferenciável e limitada. Entre outras, duas escolhas comuns para φ são

$$\varphi(v) = \frac{1}{1 + e^{-x}} \quad \text{e} \quad \varphi(v) = \tanh(x) \quad (3.5)$$

Para escrever a saída da rede como função de suas entradas, vamos considerar uma rede *feedforward* com uma camada intermediária e uma camada de saída. A saída do j -ésimo neurônio da camada intermediária é dada por $v_{pj} = \varphi(\sum_{i=1}^n w_{ij} x_{pi})$, onde n é o número total de entradas, w_{ij} é o peso que relaciona a entrada i ao neurônio j , x_{pi} é o valor da p -ésima observação da i -ésima entrada e φ é uma função não-linear diferenciável e limitada. Note que o intercepto

não está incluído na notação, mas esse termo pode ser acrescentado ao modelo com a inclusão de uma variável de entrada com valor 1.

O k -ésimo neurônio dos O neurônios da camada de saída tem como saída $\hat{y}_{pk} = r(\sum_{j=1}^m w_{kj} v_{pj})$ onde m é o número de neurônios na camada anterior, w_{kj} é o peso que relaciona a saída do j -ésimo neurônio da camada intermediária ao neurônio k da camada de saída e r é uma função de ativação. Em muitos casos, inclusive neste trabalho, onde o objetivo é predição, essa função de ativação é simplesmente a identidade, pois os valores alvo para a rede podem estar fora do intervalo $[0;1]$ ou $[-1;1]$. Portanto, daqui para frente, vamos considerar que $\hat{y}_{pk} = p_k$.

A fórmula geral para as saídas de uma rede *feedforward* com uma camada intermediária e uma camada de saída fica, então, para $k = 1; \dots; O$:

$$\hat{y}_{pk} = \sum_{j=1}^m w_{kj} \varphi(\sum_{i=1}^n w_{ji} x_{pi}) \quad (3.6)$$

O que pode ser visto simplesmente como uma regressão não-linear de parâmetros w_{ji} e w_{kj} desconhecidos.

3.7. O ALGORITMO DE RETRO PROPAGAÇÃO DE ERROS

O algoritmo de aprendizagem por retro propagação do erro ("*back-propagation*") é um método de estimação dos pesos sinápticos de uma rede *feedforward* multicamada. Foi desenvolvido por Rumelhart *et al.* (1986), apesar de ter sido explorado anos antes com resultados não reconhecidos e difundidos. O objetivo do algoritmo nada mais é do que minimizar a soma de quadrados de erros, o que representa estimação por máxima verossimilhança dos pesos da rede quando os erros têm distribuição normal.

Para efetuar essa aprendizagem, alguma medida de performance é necessária. A mais utilizada notadamente é a soma de quadrados de erros:

$$E = \frac{1}{2} \sum_p \sum_k (y_{pk} - \hat{y}_{pk})^2 \quad (3.7)$$

onde p é referente as observações, o índice k é referente aos nós de saída, y é a resposta desejada (alvo) e \hat{y} é a saída da rede (valor ajustado). Os pesos da rede são ajustados de forma a minimizar esse erro.

A cada passo do algoritmo, ou seja, quando a observação p é processada, o objetivo é minimizar:

$$E_p = \frac{1}{2} \sum_k (y_{pk} - \hat{y}_{pk})^2 \quad (3.8)$$

O que implica que $E = \sum_p E_p$.

Vamos nos ater ao caso de uma rede com duas camadas (intermediária e saída), tendo por objetivo encontrar os pesos w_{ji} e W_{kj} que minimizem a função E . Para tanto, usa-se o procedimento de minimização por descida de gradiente, percorrendo a rede da saída para a entrada, com a correção nos pesos definida pela “regra delta” [1]:

$$\Delta W_{kj} = -n \frac{\partial E_p}{\partial W_{kj}} \quad (3.9)$$

Onde n é a taxa de aprendizagem. O sinal negativo na equação é necessário para alterar o valor dos pesos na direção de uma redução no valor de E .

A fórmula para atualização dos pesos da camada de saída fica:

$$W_{kj}^{p+1} = W_{kj}^p + \Delta W_{kj} \quad (3.10)$$

Para os pesos da camada intermediária,

$$\Delta W_{ji} = -n \frac{\partial E_p}{\partial W_{ji}} \quad (3.11)$$

Algoritmo de retro-propagação

Inicialização: $w_{ji} = w_{ji}^1, w_{kj} = w_{kj}^1$

Para $p = 1, 2, 3, \dots$:

1. Com w_{ij}^p e W_{kj}^p , calcular a saídas dos neurônios da rede, isto é, v_{pj} e \hat{y}_{pk} ;
2. Calcular os erros da camada de saída: $\delta_{pk} = (y_{pk} - \hat{y}_{pk})$;
3. Calcular os erros da camada intermediária: $\psi_{pj} = \sum_{k=1}^0 \delta_{pk} W_{kj} v_{pj} (1 - v_{pj})$;
4. Atualize os pesos:

$$w_{ji}^{p+1} = w_{ji}^p + n \psi_{pj} x_{pi} \quad (3.12)$$

$$W_{kj}^{p+1} = W_{kj}^p + n \delta_{pk} v_{pj} \quad (3.13)$$

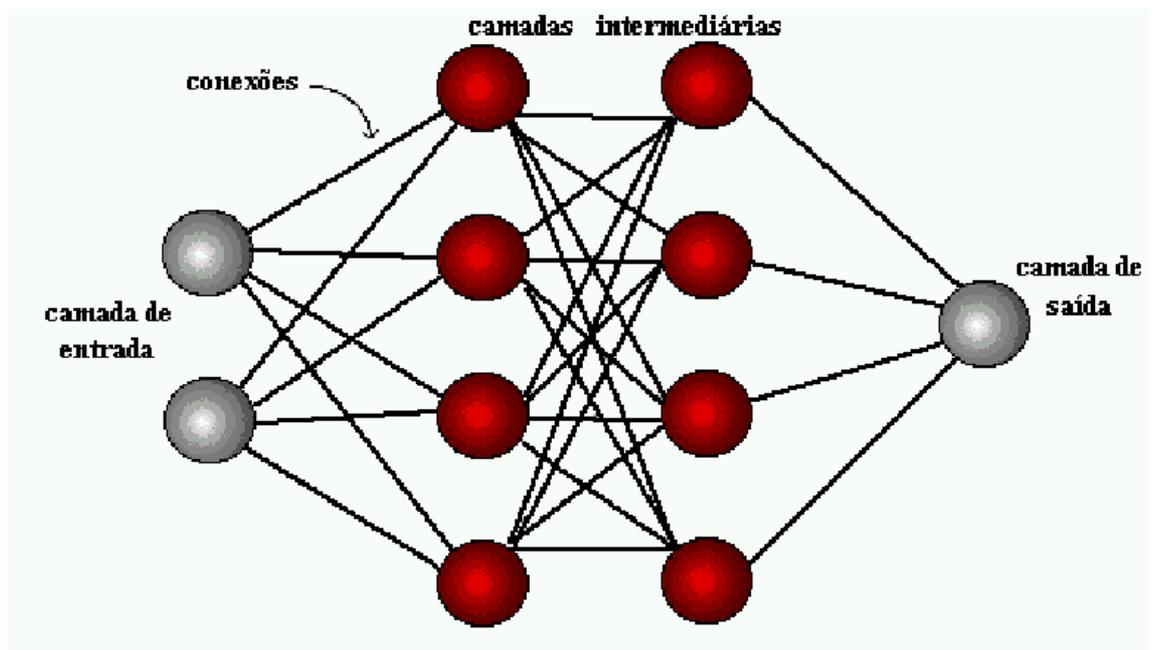


Figura 3-3 Camadas de uma rede – Fontes (1998)

Modos de treinamento: seqüencial e por batelada (*batch*)

Na utilização do algoritmo de retro-propagação, o modo de treinamento pode variar de acordo com a maneira com que o conjunto de treinamento (amostra utilizada para a estimação dos pesos) é apresentado. Uma apresentação de todo o conjunto de treinamento durante o processo de aprendizagem é chamado de época e os pesos são alterados depois que o algoritmo processa uma observação ou todo o conjunto de treinamento. Assim, o algoritmo de retro-propagação pode ser implementado de duas formas [1]:

1. Modo seqüencial: também referido como modo *on-line*. A atualização dos pesos é feita após a apresentação de cada uma das observações do conjunto de treinamento. Quando o conjunto de treinamento inteiro é processado, o treinamento é recommçado e o processo continua, até que seja interrompido através de um critério de parada. Requer um menor armazenamento para cada conexão, e apresenta menos possibilidade de convergência para um mínimo local.

2. Modo por batelada (*batch*): a atualização dos pesos é feita após a apresentação de todas as observações do conjunto de treinamento, que constituem uma época. A cada passo do algoritmo, a função E é minimizada. O conjunto de treinamento é apresentado ao algoritmo quantas vezes forem necessárias para a estimação dos pesos (o treinamento é interrompido através de um critério de parada). Fornece uma melhor estimativa do vetor gradiente.

Critério de parada:

Os métodos para encontrar os pesos são iterativos. Isso significa que, em um número grande de iterações, o valor da função E pode ser tão pequeno quanto se queira. Isso normalmente representa um bom ajuste dentro da amostra, mas em geral o modelo não conseguirá ajustar valores fora do plano amostral, pois neste caso o modelo estaria incluindo o erro em sua predição (memorização x generalização), o que não interessa. Para evitar esse inconveniente, usa-se em redes neurais, entre outras, a técnica do “*earlystopping*”.

O conjunto de dados é separado em uma amostra de treinamento, que será utilizada para estimar os pesos, e em uma amostra de validação. Quando os pesos são estimados pela amostra de treinamento, ao mesmo tempo avaliam-se os erros nas amostras, ou seja, verifica-se se o valor de E continua pequeno mesmo fora da amostra de treinamento. Desta forma, tomam-se os pesos da iteração que corresponde ao valor mínimo de E na amostra de validação (mas com

os pesos calculados baseados na amostra de treinamento), evitando-se assim o que se chama de sobre-ajuste. Pode ser mostrado que este procedimento corresponde a um método de regularização [6].

É comum separar o conjunto de dados em mais uma amostra, chamada de amostra de teste, para eventuais cálculos de performance preditiva e comparações entre diferentes modelos.

4. APLICAÇÃO DE REDES NEURAIS NA PREDIÇÃO DE DEMANDA DE CRÉDITO NO SISTEMA FINANCEIRO NACIONAL

Este capítulo aborda a aplicação de redes neurais na previsão de séries temporais e mostra a integração de redes neurais com técnicas estatísticas. A proposta do trabalho é prever uma variável dependente a partir de várias variáveis independentes, mostrando com isto que num espaço de alta dimensão as informações contidas nas variáveis independentes podem explicar o comportamento da variável dependente. A série utilizada para este exemplo consiste em observações dos recursos livres no SFN e sua previsão através das diversas variáveis correlacionadas, tais como taxa de câmbio, SELIC, produção, emprego e renda etc.

Para efeito de comparação dos resultados, consideraram-se, para cada uma das séries, os dados compreendidos de Janeiro de 2001 até Dezembro de 2006, utilizados como conjunto de treinamento e teste.

4.1. TRATAMENTOS ESTATÍSTICOS

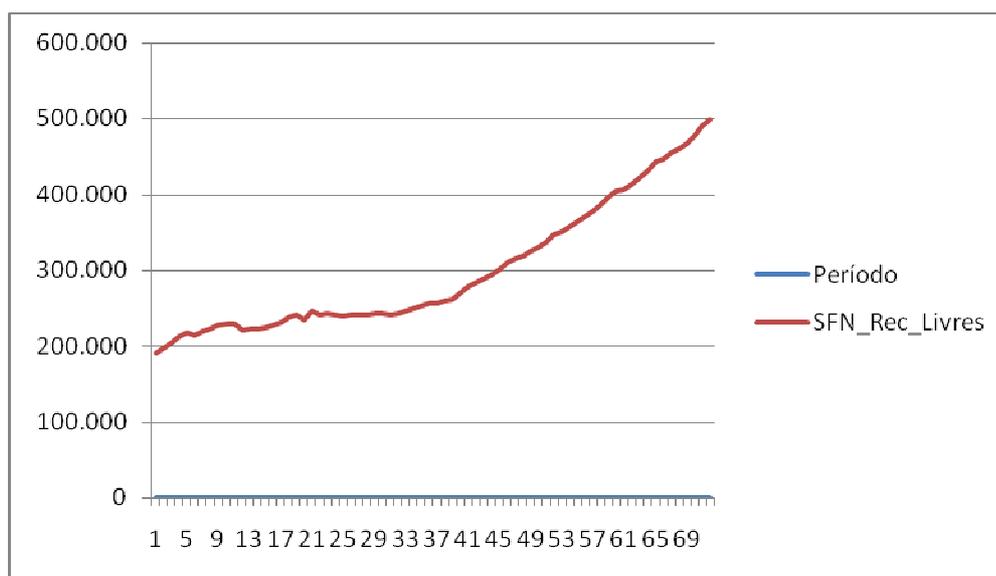


Figura 4-1 Recursos Livres no Sistema Financeiro Nacional

Primeiramente foi efetuada uma visualização gráfica da série recursos livre (Fig. 4.1), com periodicidade mensal e valores em bilhões de reais, a qual logo em seguida foi verificada suas propriedades tais como o padrão de tendência, variação sazonal, ciclo e resíduos, etc. A análise iniciou-se com a decomposição clássica da série, na qual as propriedades de uma série temporal X_t podem ser observadas assumindo a seguinte forma de decomposição:

$$X_t = T_t + C_t + R_t \quad (4.1)$$

Onde T_t é uma componente de tendência, C_t é uma componente cíclica ou sazonal e R_t é um componente aleatório ou ruído. A série tem um total de 72 entradas, correspondentes aos meses de janeiro de 2001 até dezembro de 2006.

4.1.1. Análise de correlação e regressão

Nos testes de correlação foi inicialmente visualizado o diagrama de dispersão que mostra a localização dos pontos (x,y) em um sistema de coordenadas retangulares. Foi utilizada a função Proc Plot do software SAS. Para o tratamento quantitativo do problema da dispersão dos dados amostrais é necessário estabelecerem-se medidas de correlação. O coeficiente de correlação de Pearson foi utilizado para medir os relacionamentos entre as variáveis. Sendo selecionadas as variáveis que apresentaram correlação positiva ou negativa superior a 0,60. Foi utilizada a função Proc Corr do software SAS.

Tabela 4-1 Avaliação qualitativa de r (correlação) quanto à intensidade

$ r $	Avaliação
0,0	Nula
0,0 – 0,3	Fraca
0,3 0,6	Regular
0,6 0,9	Forte
0,9 1,0	Muito forte
1,0	Perfeita

A análise de correlação nos mostra apenas a existência ou não de algum relacionamento entre as variáveis. Para investigar como estes relacionamentos ocorrem, um método apropriado é a

análise de regressão, na qual estas relações podem ser descritas matematicamente. O diagrama de dispersão permite visualizar uma curva que se aproxime dos dados (curva de ajustamento) tendo sido verificado que a Função Exponencial foi a que teve melhor aderência à série de recursos livres, pelo método dos mínimos quadrados.

4.2. REGRESSÃO LINEAR MÚLTIPLA

A Regressão Linear Múltipla foi a técnica estatística usada para estudar a relação entre a variável dependente (critério) e várias variáveis independentes (preditoras) [4]. O objetivo principal da análise de regressão múltipla é utilizar as variáveis independentes para prever os valores da variável dependente. Sua fórmula básica é:

$Y_1 = X_1 + X_2 \dots + X_n$ Onde Y_1 é a variável dependente e $X_1 + X_2 \dots + X_n$ são as variáveis independentes ou preditoras; Após a análise multivariada de dados foram encontrados os seguintes coeficientes:

$$\text{SFN_Rec_Livres} = 36313 + 771.63xEP6 - 0.1352xFP13 - 0.5514xFN13 + 1.1125xPD12 + 1.6421xPD13 \quad (4.2)$$

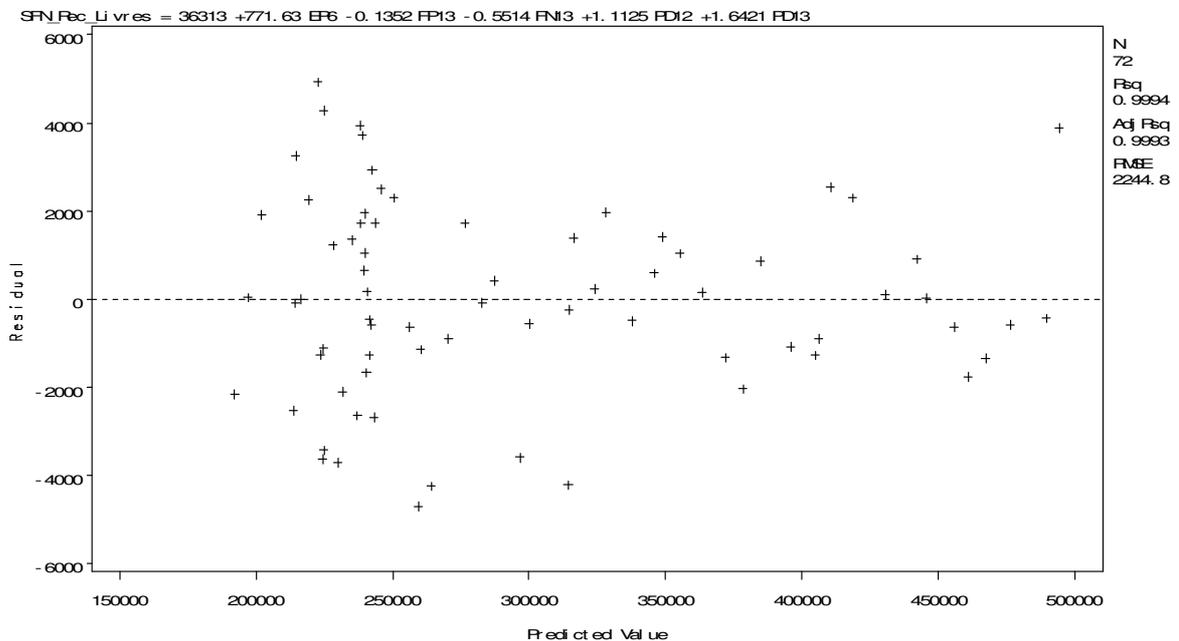


Figura 4-2 Análise resíduos

O objetivo da regressão múltipla neste estudo é prever o volume de Crédito com Recursos Livres no Sistema Financeiro Nacional (variável dependente) através de variáveis econômicas.

4.2.1. Suposições em análise de regressão múltipla

O modelo de regressão linear clássico depende de algumas suposições que descrevem as formas do modelo e também ditam os procedimentos adequados de estimação e inferência, conforme descritos a seguir [4].

Normalidade. Nos testes de normalidade foi comprovado que os períodos jan/2001 a dez/2006 das séries utilizadas neste estudo provêm de uma distribuição normal. Através destas amostras uma estatística foi calculada e checada para comprovar esta idéia. O teste de *Kolmogorov-Smirnov* consiste em verificar se todas as variáveis consideradas apresentam distribuição normal conforme a curva de Gauss. Em caso afirmativo, ficará assegurada a distribuição normal multivariada das variáveis. A não normalidade das variáveis não invalida a utilização dos métodos estatísticos, porém os resultados são mais consistentes quando se trabalham com amostras com variáveis que apresentem distribuição normal de média e desvio padrão. Foi utilizada a ferramenta *Proc Univariate* do software SAS sendo o teste baseado na estatística de SHAPIRO_WILK ($N < 2000$).

Linearidade. A suposição de linearidade inclui o distúrbio estocástico aditivo e pode envolver uma transformação adequada das variáveis. Portanto, no contexto da regressão, a linearidade se refere à forma pela qual os parâmetros e o distúrbio entram na equação e não necessariamente à relação entre as variáveis envolvidas.

Heteroscedasticidade. Quando os termos de erro têm variância crescente ou flutuante. É uma das violações mais comuns de suposições, sendo o seu diagnóstico feito com gráfico de resíduos ou testes estatísticos simples. Quando verificada a sua ocorrência, foi aplicada, como ação corretiva uma transformação de estabilização da variância.

Independência dos termos de erro.

Foi verificado que cada valor previsto é independente, através da análise do gráfico de resíduos das variáveis e quando ocorreram violações desta suposição, foram efetuadas transformações nos dados (primeira diferença).

4.2.2. Estimação do modelo de regressão e avaliação do ajuste geral do modelo

Foram utilizados métodos de busca seqüencial que têm em comum a abordagem geral de estimar a equação de regressão com um conjunto de variáveis incrementando ou decrementando de forma seletiva até que os critérios sejam alcançados. Há dois tipos de abordagens de busca seqüencial: a estimação *Stepwise* e a adição *Forward* e eliminação *Backward*. Foram selecionadas 5 variáveis que juntas explicam mais de 99% da variância da série recursos livres:

Tabela 4-2 Variáveis selecionadas regressão múltipla

EP 6 – Grupamento Emprego (nível emprego – indústria - índice FIESP).
FN 13 – Grupamento Financeiro (divida externas - estatais – líquida).
FP 13 – Grupamento Finanças-Públicas (divida fiscal - setor público – líquida).
PD 12 – Grupamento Produção (operações de crédito do sistema financeiro – setor privado – pessoas físicas).
PD 13 – Grupamento Produção (operações de crédito do sistema financeiro – setor privado – outros serviços).

Tabela 4-3 Demais variáveis econômicas iniciais

Arrecadação das receitas federais - receita bruta - R\$(milhões) - Min. Fazenda/SRF - SRF12_TOTGER12
Balança comercial - (FOB) - saldo - US\$(milhões) - BCB Boletim/BP - BPN12_SBC12
Bônus global República (12) - spread - (p.p.) - Valor Econômico - VALOR12_GLOBAL1212
Cheques sem fundo - (%) - Serasa - SERASA12_CHSF12
Cide - combustíveis - receita bruta - R\$(milhões) - Min. Fazenda/SRF - SRF12_CIDE12
Commodities - carnes - preço - índice (jan. 2002 = 100) - IPEA - GAC12_CARN12

Conta capital e financeira - conta capital - US\$(milhões) - BCB Boletim/BP - BPN12_STKK12
Conta capital e financeira - conta financeira - US\$(milhões) - BCB Boletim/BP - BPN12_STF12
Conta financeira - investimentos diretos - brasileiros no ext. - emprést. intercompanhias - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDBE12
Conta financeira - investimentos diretos - brasileiros no ext. - retornos - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDBREC12
Conta financeira - investimentos diretos - brasileiros no ext. - saídas - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDBDES12
Conta financeira - investimentos diretos - brasileiros no exterior - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDB12
Conta financeira - investimentos diretos - estrang. no país - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDE12
Conta financeira - investimentos diretos - estrang. no país - emprést. intercomp. - amortizações - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEEDES12
Conta financeira - investimentos diretos - estrang. no país - emprést. intercomp. - desembolsos - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEEREC12
Conta financeira - investimentos diretos - estrang. no país - emprést. intercompanhias - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEE12
Conta financeira - investimentos diretos - estrang. no país - ingressos - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEREC12
Conta financeira - investimentos diretos - estrang. no país - saídas - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEDES12

Conta financeira - investimentos estrang. em carteira - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDEC12
Conta financeira - investimentos estrang. em carteira - despesas - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDECDES12
Conta financeira - investimentos estrang. em carteira - receitas - US\$(milhões) - BCB Boletim/BP - BPN12_FINIDECREC12
Demais receitas - receita bruta - R\$(milhões) - Min. Fazenda/SRF - SRF12_DEMAIS12
Dívida - setor público - ajuste patrimonial - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DIAJPY12
Dívida - setor público - ajuste patrimonial - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DIAJPN12
Dívida - total - estatais - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DTEEY12
Dívida - total - estatais - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DTEEN12
Dívida externa - estados e municípios - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DEXEMY12
Dívida externa - estados e municípios - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DEXEMN12
Dívida externa - estatais - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DEXEEY12
Dívida externa - estatais - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DEXEEN12
Dívida externa - gov. federal e Banco Central - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DEXGFY12
Dívida externa - gov. federal e Banco Central - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DEXGFN12

Dívida externa - setor público - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DEXSPN12
Dívida fiscal - setor público - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DIFISPY12
Dívida fiscal - setor público - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DIFISPN12
Dívida interna - estados e municípios - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DINEMY12
Dívida interna - estados e municípios - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DINEMN12
Dívida interna - estatais - líquida - (% PIB) - BCB Boletim/F. Públ. - BM12_DINEEY12
Dívida interna - estatais - líquida - R\$(milhões) - BCB Boletim/F. Públ. - BM12_DINEEN12
Estados Unidos - taxa de juros - T-Bond (30 anos) - (% a.a.) - Valor Econômico - VALOR12_TBOND3012
Estados Unidos - taxa de juros - T-Note (10 anos) - (% a.a.) - Valor Econômico - VALOR12_TNOTES1012
Estados Unidos - taxa de juros - T-Note (2 anos) - (% a.a.) - Valor Econômico - VALOR12_TNOTES212
Estados Unidos - taxa de juros - T-Note (5 anos) - (% a.a.) - Valor Econômico - VALOR12_TNOTES512
Exportações - (FOB) - US\$(milhões) - BCB Boletim/BP - BPN12_XTV12
Importações - (FOB) - US\$(milhões) - BCB Boletim/BP - BPN12_MTV12
Nível de emprego - indústria - índice (dez. 1999 = 100) - SP - Fiesp - FIESP12_NEINSPX12
Núcleo - FIPE - (% a.m.) - IPEA - GAC12_NUCLEOFIPE

Núcleo - INPC - (% a.m.) - IPEA - GAC12_NUCLEOINPC
Núcleo - IPA-OG - (% a.m.) - IPEA - GAC12_NUCLEOIPA
Núcleo - IPCA - (% a.m.) - IPEA - GAC12_NUCLEOIPCA
Operações de crédito do sistema financeiro - setor privado - outros serviços - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPRIOUT12
Operações de crédito do sistema financeiro - setor privado - pessoas físicas - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPRIPF12
Operações de crédito do sistema financeiro - setor privado - rural - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPRIRUR12
Operações de crédito do sistema financeiro - setor público - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPUBL12
Operações de crédito do sistema financeiro - setor público - governo federal - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPUBLGF12
Operações de crédito do sistema financeiro - setor público - governos estaduais e municipais - R\$(milhões) - BCB Boletim/Moeda - BM12_SFPUBLEM12
Operações de crédito do sistema financeiro - total - R\$(milhões) - BCB Boletim/Moeda - BM12_SFTOT12
Reservas internacionais - liquidez internacional - US\$(milhões) - BCB Boletim/BP - BM12_RESLIQ12
Resultado global do balanço de pagamentos - US\$(milhões) - BCB Boletim/BP - BPN12_SBP12
Taxa de câmbio - efetiva real - INPC - exportações - manufaturados - índice (média 2000 = 100) - IPEA - GAC12_TCERXMINPC12

Taxa de câmbio - efetiva real - IPA-IT - exportações - manufaturados - índice (média 2000 = 100) - IPEA - GAC12_TCERXM12
Taxa de câmbio - efetiva real - IPA-OG - exportações - índice (média 2000 = 100) - IPEA - GAC12_TCERXT12
Taxa de câmbio - efetiva real - setor: agropecuária - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP1S4
Taxa de câmbio - efetiva real - setor: artigos de vestuário - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP23S4
Taxa de câmbio - efetiva real - setor: borracha - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP16S4
Taxa de câmbio - efetiva real - setor: café - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP25S4
Taxa de câmbio - efetiva real - setor: calçados - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP24S4
Taxa de câmbio - efetiva real - setor: celulose, papel e gráfica - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP15S4
Taxa de câmbio - efetiva real - setor: elementos químicos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP17S4
Taxa de câmbio - efetiva real - setor: equipamentos eletrônicos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP11S4
Taxa de câmbio - efetiva real - setor: extrativa mineral - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP2S4
Taxa de câmbio - efetiva real - setor: madeira e mobiliário - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP14S4

Taxa de câmbio - efetiva real - setor: material elétrico - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP10S4
Taxa de câmbio - efetiva real - setor: metalurgia não ferrosos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP6S4
Taxa de câmbio - efetiva real - setor: minerais não metálicos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP4S4
Taxa de câmbio - efetiva real - setor: máquinas e tratores - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP8S4
Taxa de câmbio - efetiva real - setor: outros produtos metalúrgicos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP7S4
Taxa de câmbio - efetiva real - setor: peças e outros veículos - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP13S4
Taxa de câmbio - efetiva real - setor: plástica - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP21S4
Taxa de câmbio - efetiva real - setor: siderurgia - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP5S4
Taxa de câmbio - efetiva real - setor: têxtil - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP22S4
Taxa de câmbio - efetiva real - setor: veículos automotores - índice (média 2000 = 100) - IPEA - GAC12_COEFEXP12S4
Taxa de juros pré fixada - estrutura a termo - LTN - 1 mês - (% a.a.) - Andima - ANDIMA12_TJTLN112
Taxa de juros pré fixada - estrutura a termo - LTN - 12 meses - (% a.a.) - Andima - ANDIMA12_TJTLN1212

Taxa de juros pré fixada - estrutura a termo - LTN - 3 meses - (% a.a.) - Andima - ANDIMA12_TJTLN312
Taxa de juros pré fixada - estrutura a termo - LTN - 6 meses - (% a.a.) - Andima - ANDIMA12_TJTLN612
Taxa referencial - swaps - DI pré-fixada - 180 dias - fim de período - (% a.a.) - BM&F - BMF12_SWAPDI180F12
Taxa referencial - swaps - DI pré-fixada - 180 dias - média do período - (% a.a.) - BM&F - BMF12_SWAPDI18012
Taxa referencial - swaps - DI pré-fixada - 360 dias - fim de período - (% a.a.) - BM&F - BMF12_SWAPDI360F12
Taxa referencial - swaps - DI pré-fixada - 360 dias - média do período - (% a.a.) - BM&F - BMF12_SWAPDI36012
Índice de ações - Ibovespa - fechamento - (% a.m.) - Andima - ANDIMA12_IBVSP12

4.2.3. Avaliação da significância estatística do modelo na previsão da variável dependente

Após as variáveis independentes serem selecionadas e os coeficientes de regressão estimados, foi avaliado se o modelo estimado atendia às suposições de regressão múltipla. Foram verificadas as variáveis individuais e a variável estatística quanto à linearidade, variância constante, independência e normalidade e efetuado testes de significância estatística:

a) Teste da variação explicada (coeficiente de determinação)

Efetuoado para testar a hipótese de que a quantidade de variação explicada pelo modelo de regressão é maior que a variação explicada pela média (R^2 é maior que zero). R^2 é conhecido como coeficiente de correlação múltipla ao quadrado ou coeficiente de determinação. Quando o ajuste é bom o modelo explica boa parte da variação total e conseqüentemente o valor de R^2 é próximo de 1. O coeficiente de determinação é uma medida da qualidade do ajuste. Foi utilizado o teste estatístico da razão F.

b) Teste de coeficientes

Testar a significância estatística para os coeficientes estimados em análise de regressão é necessário quando a análise é baseada em uma amostra da população, para se verificar o quanto os resultados são generalizáveis para a população. Testes de significância dos coeficientes de regressão fornecem uma estimativa estatística baseada na probabilidade de que os coeficientes estimados em muitas amostras de um dado tamanho sejam de fato diferentes de zero.

4.2.4. Interpretação da variável estatística de regressão

Os coeficientes de regressão estimados foram utilizados para calcular os valores previstos de cada observação e para expressar a variação esperada na variável dependente para cada variação unitária nas variáveis independentes, além de avaliar o impacto de cada variável independente na previsão da dependente.

Avaliação da multicolinearidade - Os efeitos da multicolinearidade podem ser classificados em termos de explicação e estimação. Os efeitos na explicação se referem principalmente à habilidade do procedimento de regressão em representar os efeitos de cada variável independente na variável estatística de regressão. Quando ocorre multicolinearidade, mesmo em níveis relativamente baixos (abaixo de 0,30), o processo de separação de efeitos individuais se torna mais difícil. Primeiro, ela limita o tamanho do coeficiente de determinação e fica cada vez mais difícil acrescentar uma única previsão explanatória a partir de variáveis adicionais. Segundo, ela torna a determinação da contribuição de cada variável independente mais complicada porque os efeitos das variáveis independentes estão misturados. A multicolinearidade resulta em partes maiores de variância compartilhada e níveis mais baixos de variância única da qual os efeitos das variáveis independentes individuais podem ser explicadas. Foi utilizado o fator de inflação da variância (VIF) para identificar e eliminar as variáveis que apresentavam altos fatores de multicolinearidade (maior que 10) .

Validação dos resultados - Para validação dos resultados do modelo utilizou-se a amostra original, onde os modelos foram estimados.

4.3. REDES NEURAIAS

Este trabalho foi realizado utilizando-se para as simulações de redes neurais a *Neural Network Toolbox* do software Matlab 7.0. A topologia utilizada foi a MLP (*Multilayer Perceptron feedforward*) com uma camada de entrada, com 5 nós, conforme as variáveis obtidas na Análise das Regressão múltipla, uma camada escondida, e uma camada de saída com um neurônio representando o valor dos recursos livres no SFN.

O teorema da Aproximação Universal [1], afirma que uma única camada oculta é suficiente para um *perceptron* de múltiplas camadas (MLP) realizarem uma aproximação efetiva de um conjunto de treinamento representada pelo conjunto de entradas e o alvo. Testes foram feitos com o número de neurônios na camada oculta variando de 5 a 50 em passos de 5.

Foi utilizada uma função de ativação na camada oculta contínua e diferenciável. Utilizou-se na camada oculta a função de ativação tangente hiperbólica (*tansig*), que assume um intervalo contínuo de valores entre -1 e 1 e na camada de saída a função de ativação não-linear do tipo sigmoidal (*logsig*), que assume um intervalo contínuo de valores entre 0 e 1 e é definida como uma função estritamente crescente, que exibe um balanceamento adequado entre comportamento linear e não-linear.

Para o treinamento, foi utilizado o algoritmo de *Scaled Conjugate Gradient* (Trainscg), com a soma dos quadrados dos erros (SSE) em 0,0001, sendo as redes projetadas para calcular os pesos sinápticos do *perceptron* de múltiplas camadas, isto é, aprendizagem por retro propagação do erro.

Para a avaliação do desempenho da rede neural, utilizou-se MAPE, RMSE e U de Theil, tanto do conjunto de treinamento, quanto do conjunto de testes. Para se determinar o quanto uma previsão é precisa, é necessário haver uma forma de estimar o quanto se está errando, em média. Este erro é mensurado de três formas distintas neste trabalho:

O MAPE (*Mean Absolute Percentage Error*): Representa o erro percentual absoluto médio.

$$MAPE = \frac{\sum_{k=1}^n \left| \frac{a_k - y_k}{a_k} \right|}{N} \times 100\% \quad (4.3)$$

Onde

- N é o número de previsões realizadas
- ak é a saída desejada para a previsão índice “ k ”.
- yk é a saída prevista para a previsão índice “ k ”.

O RMSE (*Root Mean Squared Error*): Representa a raiz do erro quadrático médio, sendo uma medida de erro que penaliza muito mais os erros de maior magnitude.

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (a_k - y_k)^2}{N}} \quad (4.4)$$

Onde

- N é o número de previsões realizadas
- ak é a saída desejada para a previsão índice “ k ”.
- yk é a saída prevista para a previsão índice “ k ”.

O U de Theil: Esta métrica representa o quanto os resultados estão melhores que uma previsão ingênua ou trivial.

$$U = \frac{\sqrt{\sum_{k=1}^n (a_k - y_k)^2}}{\sqrt{\sum_{k=1}^n (a_k - a_{k-1})^2}} \quad (4.5)$$

- ak é a saída desejada para a previsão índice “ k ”.
- yk é a saída prevista para a previsão índice “ k ”.

Através deste coeficiente pode-se analisar a qualidade de uma previsão da seguinte maneira:

- quando $U \geq 1$, o erro do modelo é maior que o erro ingênuo.
- quando $U < 1$, o erro do modelo é menor que o erro ingênuo (boa previsão).

4.3.1. Pré-processamento das séries

Antes de treinar a rede neural foram feitas algumas transformações nos dados:

Inicialmente foram preparados os pontos de entrada da rede, para que estivessem em uma faixa entre -1 e 1 . Isto porque a função de ativação dos neurônios utilizada foi a tangente hiperbólica, que possui seu intervalo mais significativo de variação entre -1 e 1 .

4.3.2. Pós-processamento das séries

Depois de treinada a rede com o conjunto de treinamento apropriado, e feitas às previsões com o conjunto de teste também transformado (pré-processado), foram feitas as transformações inversas ao pré-processamento para se obter a série prevista final \hat{s} .

4.4. ESTUDOS DE CASOS

Em vista de avaliar o desempenho das redes neurais temporais em aplicações de previsão de dados reais, foi considerada a demanda de crédito com recursos livres no sistema financeiro nacional. Foram escolhidas séries disponíveis por entidades como o IPEA. Foram feitos os tratamentos de pré-processamento, antes de apresentar os dados às redes neurais, e seus respectivos pós-processamento para apresentar e analisar os resultados.

Durante o estudo de caso foram feitos vários testes. Nestes testes foram manipulados vários parâmetros das Redes Neurais em busca de resultados cada vez melhores. Estas variações foram feitas de maneira sistemática para que fosse possível analisar se as mudanças tinham sido vantajosas e quais os fatores responsáveis por esta melhora. Percebeu-se que a integração de séries submetidas à análise multivariada de dados proporcionou resultados muito superiores aos resultados de séries escolhidas apenas por forte correlação. Esta constatação foi confirmada após sucessivas simulações das redes MLP.

4.5. TESTES E RESULTADOS

Com o objetivo de avaliar os resultados das Redes Neurais de forma consistente, foram realizadas medidas diferentes de erro, onde cada medida observou os resultados por um prisma diferente.

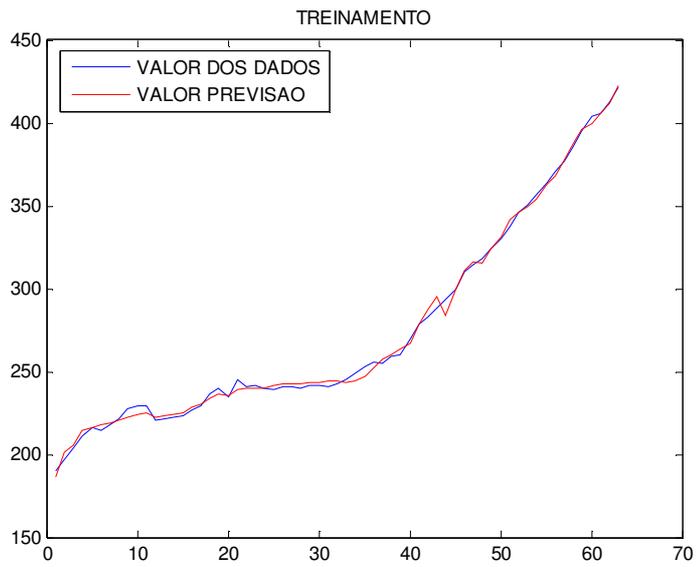


Figura 4-3 Utilização de séries sem tratamento estatísticos - treinamento

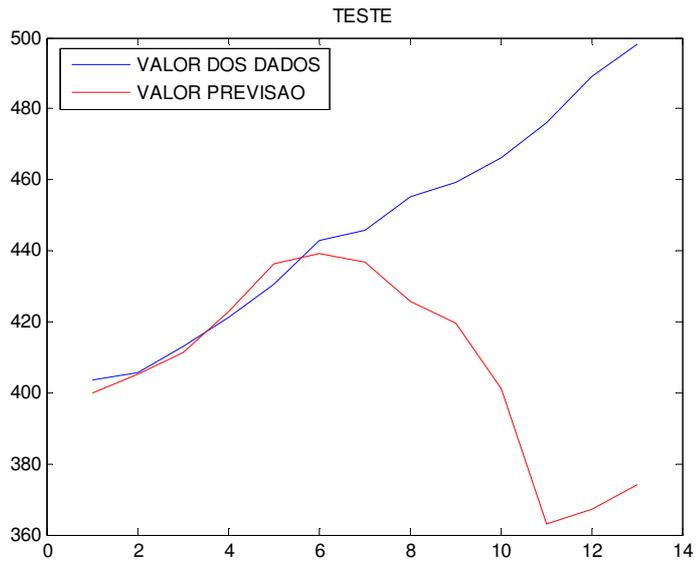


Figura 4-4 Utilização de séries sem tratamento estatísticos - teste

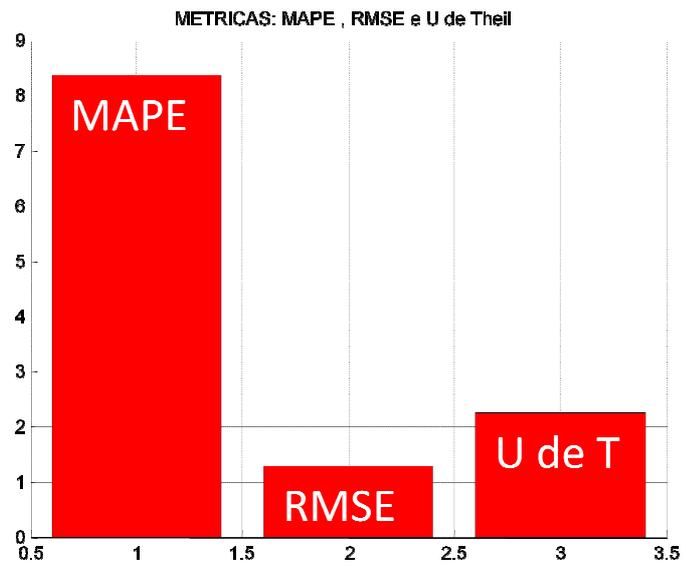


Figura 4-5 Utilização de séries sem tratamento estatísticos - métricas

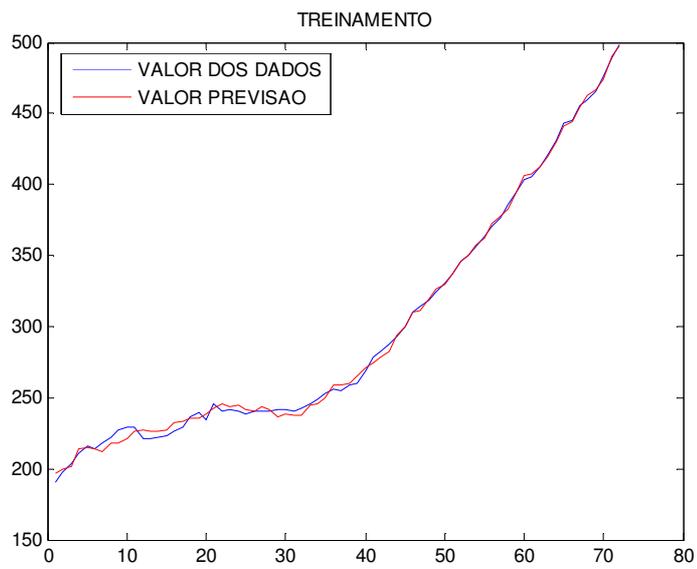


Figura 4-6 Utilização de séries com tratamento estatísticos - treinamento

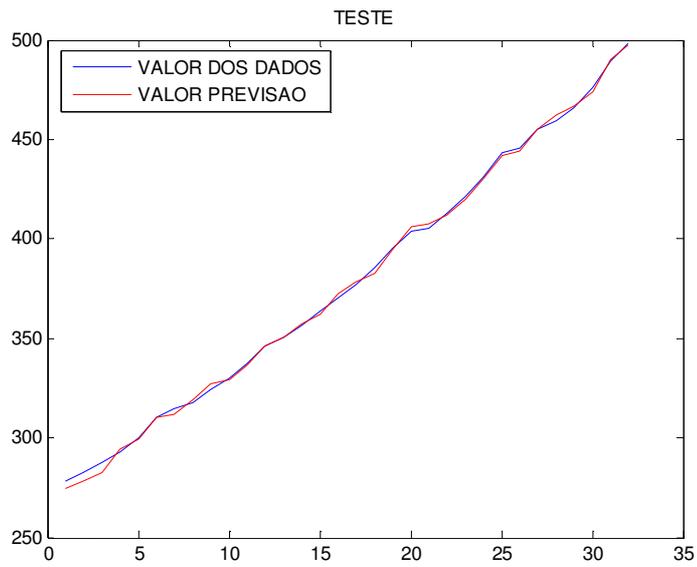


Figura 4-7 Utilização de séries com tratamento estatísticos - teste

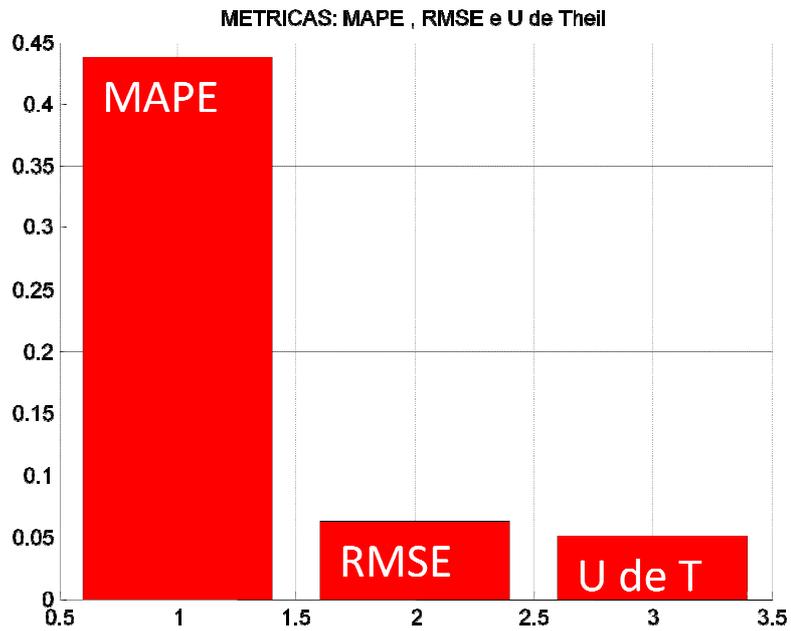


Figura 4-8 Utilização de séries com tratamento estatísticos - métricas

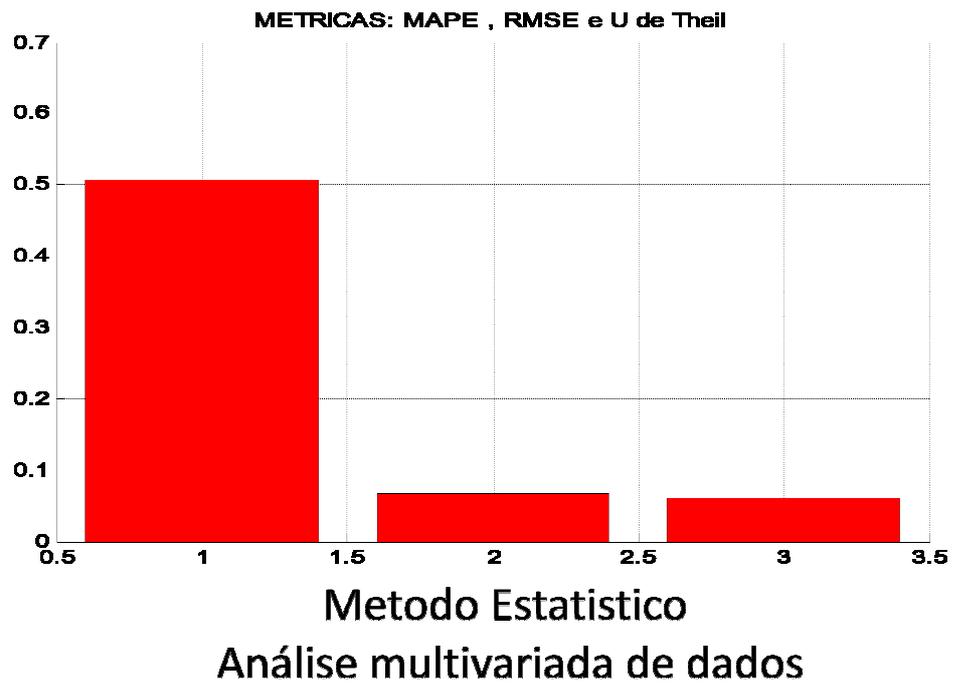


Figura 4-9 Método estatístico análise multivariada de dados – métrica

5. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Os objetivos apresentados para o tema do trabalho, foram atingidos, abordando-se análise multivariada de dados, previsão de séries temporais e analisando a teoria de regressão, podendo-se concluir sobre o estudo de caso:

1- A rede neural, aplicada ao caso em estudo, foi capaz de prognosticar demandas de crédito utilizando séries econômicas.

2- Os melhores resultados encontrados, na determinação do prognóstico de demanda de crédito no curto prazo, foram obtidos com séries submetidas a análise multivariada de dados e com a seguinte topologia:

Rede MLP, épocas = 2000 e 40 neurônios na camada escondida, obtendo MAPE = 0,44 e RMSE = 0,06. Resultado melhor do que o obtido com métodos estatísticos MAPE= 0,64 e RMSE= 0,05.

3- A escolha das características estáveis da série em estudo, com rede neural, apresentou resultados de prognóstico, e, portanto, capacidade de generalização, igual ou melhor do que o método estatístico.

4- As dificuldades encontradas no processamento de regressão com redes neurais sugerem, a continuidade dos trabalhos de pesquisa, na integração de técnicas estatísticas com inteligência computacional voltados principalmente para o estudo de relações entre as séries econômicas e o volume de recursos livres no Sistema Financeiro Nacional.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Haykin, S. (2001). *Redes Neurais – Princípios e prática*. 2.ed. Bookman, Porto Alegre.
- [2] Kovács, Z. L. (1996). *Redes Neurais Artificiais: Fundamentos e Aplicações*. 2a ed. Edição Acadêmica/USP.
- [3] Hair, Jr, JF. (2005). *Análise Multivariada de Dados*. 5.ed. Bookman, Porto Alegre.
- [4] Mingoti, Sueli Aparecida. (2005). *Análise de Dados Através de Estatística Multivariada – Uma abordagem aplicada*. 1.ed. Editora UFMG, Belo Horizonte.
- [5] Hanselman, Duane. (2003). *Matlab 6: Curso completo*. 1.ed. Prentice Hall, São Paulo.
- [6] Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- [7] McCulloch, C. E., Searle, S. R. (2001). *Generalized, Linear and Mixed Models*. Wiley.
- [8] McNelis, P. D., McAdam, P. (2004). *Forecasting Inflation with Thick Models and Neural Networks*. ECB Working Paper n. 352.
- [9] Morettin, P. A., Toloi, M. C. (2004). *Análise de Series Temporais*. Edgard Blucher, São Paulo.
- [10] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). *Learning Internal Representations by Error Propagation*. In: *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1: Foundations, ed. Rumelhart, D. E., McClelland, J. L. MIT Press, Cambridge.
- [11] Ehlers, R.S. (2005) *Análise de Series Temporais*. Departamento de Estatística, UFPR. Disponível em <http://leg.est.ufpr.br/~ehlers/notas>. Acesso em: 16 dez. 2007.
- [12] C. Chatfield. *“The analysis of time series. An introduction”*. 6ª Edition. Chapman & Hall /CRC Press, 2003.

[13]-WASSERMAN, P. D. Neural computing: theory and practice. New York: Van Nostrand Reinhold, 1989.

[14]-HOPFIELD, J. J. Neurons with graded responses have collective computational properties like those of two-states neurons. In: National Academy of science. E.U.A., 1981. /Anais/E.U.A., p. 3088-3092, 1984.

[15]-HINTON, G. E., e SEJNOWSKI, T. J. Learning and relearning in Boltzmann machines. In: Parallel distributed processing. /Anais /. v. 1, cap 7, 1986.

[16]-HERTZ, J., KROGH, A. e PALMER, R. G. Introduction to the theory of neural computation. Redwood City: Addison-Wesley, 1991.

[17] Simon Haykin. "Neural Networks-A comprehensive foundation". Prentice Hall, 1994.

[18] WHEELWRIGHT, Steven C.; MAKRIDAKIS, Spyros. Forecasting Methods for Management. 4th edition. New York : John Wiley & Sons Inc, 1985.

APÊNDICES

The UNIVARIATE Procedure

Variable: SFN_REC_LIVRES (SFN_REC_LIVRES)

Moments

N	72	Sum Weights	72
Mean	298372.178	Sum Observations	21482796.8
Std Deviation	85652.4599	Variance	7336343885
Skewness	0.87128232	Kurtosis	-0.5345559
Uncorrected SS	6.93075E12	Corrected SS	5.2088E11
Coeff Variation	28.706584	Std Error Mean	10094.2392

Basic Statistical Measures

Location		Variability	
Mean	298372.2	Std Deviation	85652
Median	255292.4	Variance	7336343885
Mode	.	Range	308386
		Interquartile Range	124540

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----
Student's t	t 29.55866	Pr > t <.0001
Sign	M 36	Pr >= M <.0001
Signed Rank	S 1314	Pr >= S <.0001

Tests for Normality

Test	--Statistic---	-----p Value-----
Shapiro-Wilk	W 0.865735	Pr < W <0.0001
Kolmogorov-Smirnov	D 0.215039	Pr > D <0.0100
Cramer-von Mises	W-Sq 0.687946	Pr > W-Sq <0.0050
Anderson-Darling	A-Sq 3.848619	Pr > A-Sq <0.0050

The UNIVARIATE Procedure

Variable: EP6 (EP6)

Moments

N	72	Sum Weights	72
Mean	113.361111	Sum Observations	8162
Std Deviation	6.62752595	Variance	43.9241002
Skewness	0.309841	Kurtosis	-1.4798811
Uncorrected SS	928372	Corrected SS	3118.61111
Coeff Variation	5.84638407	Std Error Mean	0.78106142

Basic Statistical Measures

Location		Variability	
Mean	113.3611	Std Deviation	6.62753
Median	111.0000	Variance	43.92410
Mode	107.0000	Range	21.00000
		Interquartile Range	13.00000

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 145.1373	Pr > t	<.0001
Sign	M 36	Pr >= M	<.0001
Signed Rank	S 1314	Pr >= S	<.0001

Tests for Normality

Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.886579	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.166954	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.556678	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 3.171768	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure

Variable: FP13 (FP13)

Moments

N	72	Sum Weights	72
Mean	637542.222	Sum Observations	45903040
Std Deviation	108885.429	Variance	1.1856E10
Skewness	0.00131259	Kurtosis	-1.2341852
Uncorrected SS	3.01069E13	Corrected SS	8.41779E11
Coeff Variation	17.0789361	Std Error Mean	12832.2708

Basic Statistical Measures

Location		Variability	
Mean	637542.2	Std Deviation	108885
Median	654571.0	Variance	1.1856E10
Mode	.	Range	361395
		Interquartile Range	191644

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 49.68273	Pr > t	<.0001
Sign	M 36	Pr >= M	<.0001
Signed Rank	S 1314	Pr >= S	<.0001

Tests for Normality

Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.94224	Pr < W	0.0025
Kolmogorov-Smirnov	D 0.114373	Pr > D	0.0202
Cramer-von Mises	W-Sq 0.150157	Pr > W-Sq	0.0234
Anderson-Darling	A-Sq 1.108192	Pr > A-Sq	0.0066

The UNIVARIATE Procedure

Variable: FN13 (FN13)

Moments

N	72	Sum Weights	72
Mean	2012.25	Sum Observations	144882
Std Deviation	13063.0935	Variance	170644411
Skewness	-0.663164	Kurtosis	-0.591043
Uncorrected SS	1.24073E10	Corrected SS	1.21158E10
Coeff Variation	649.178456	Std Error Mean	1539.50033

Basic Statistical Measures

Location		Variability	
Mean	2012.250	Std Deviation	13063
Median	7146.500	Variance	170644411
Mode	.	Range	49631
		Interquartile Range	23651

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 1.30708	Pr > t	0.1954
Sign	M 7	Pr >= M	0.1249
Signed Rank	S 267	Pr >= S	0.1351

Tests for Normality

Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.913148	Pr < W	0.0001
Kolmogorov-Smirnov	D 0.192312	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.413211	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 2.24326	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure

Variable: PD12 (PD12)

Moments

N	72	Sum Weights	72
Mean	122955.792	Sum Observations	8852817
Std Deviation	52331.2618	Variance	2738560957
Skewness	0.81708968	Kurtosis	-0.7633491
Uncorrected SS	1.28294E12	Corrected SS	1.94438E11
Coeff Variation	42.5610384	Std Error Mean	6167.29834

Basic Statistical Measures

Location		Variability	
Mean	122955.8	Std Deviation	52331
Median	99650.0	Variance	2738560957
Mode	.	Range	172385
		Interquartile Range	82028

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 19.93673	Pr > t	<.0001
Sign	M 36	Pr >= M	<.0001
Signed Rank	S 1314	Pr >= S	<.0001

Tests for Normality

Test	--Statistic---		-----p Value-----	
Shapiro-Wilk	W	0.848451	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.203881	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.750561	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.298832	Pr > A-Sq	<0.0050

The UNIVARIATE Procedure

Variable: PD13 (PD13)

Moments

N	72	Sum Weights	72
Mean	76198.3194	Sum Observations	5486279
Std Deviation	18159.4693	Variance	329766324
Skewness	0.72642116	Kurtosis	-0.2309653
Uncorrected SS	4.41459E11	Corrected SS	2.34134E10
Coeff Variation	23.8318501	Std Error Mean	2140.11398

Basic Statistical Measures

Location		Variability	
Mean	76198.32	Std Deviation	18159
Median	71963.50	Variance	329766324
Mode	.	Range	74346
		Interquartile Range	22310

Tests for Location: $\mu_0=0$

Test	-Statistic-	-----p Value-----	
Student's t	t 35.6048	Pr > t	<.0001
Sign	M 36	Pr >= M	<.0001
Signed Rank	S 1314	Pr >= S	<.0001

Tests for Normality

Test	--Statistic---	-----p Value-----	
Shapiro-Wilk	W 0.940007	Pr < W	0.0019
Kolmogorov-Smirnov	D 0.121753	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.243059	Pr > W-Sq	<0.0050
Anderson-DarLing	A-Sq 1.466172	Pr > A-Sq	<0.0050