

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

DEPARTAMENTO DE ESTATÍSTICA

PREVISÃO DE SÉRIES TEMPORAIS:
REDES NEURAS ARTIFICIAIS VS. MODELOS ARIMA

AUTOR: GUSTAVO FUMAGALLI MANTOVANI

ORIENTADOR: FLÁVIO AUGUSTO ZIEGELMANN

TRABALHO DE CONCLUSÃO APRESENTADO PARA A
OBTENÇÃO DO GRAU DE BACHAREL EM ESTATÍSTICA

PORTO ALEGRE, DEZEMBRO DE 2004

SUMÁRIO

LISTA DE FIGURAS	3
LISTA DE TABELAS	5
INTRODUÇÃO	6
DESENVOLVIMENTO	7
1 MODELOS ARIMA	7
1.1 <i>Processos estocásticos</i>	7
1.2 <i>Processos estacionários</i>	8
1.3 <i>Função de autocorrelação</i>	8
1.4 <i>Modelos de médias móveis (MA)</i>	9
1.5 <i>Modelos auto-regressivos (AR)</i>	10
1.6 <i>Modelos auto-regressivos e de médias móveis (ARMA)</i>	11
1.7 <i>Processos não-estacionários</i>	11
1.8 <i>Modelos auto-regressivos integrados de médias móveis (ARIMA)</i>	11
1.9 <i>Modelos sazonais de Box-Jenkins</i>	12
1.10 <i>Metodologia de Box-Jenkins para previsão</i>	13
1.11 <i>Estimação e verificação</i>	14
1.12 <i>Previsão</i>	15
2 MODELOS DE REDES NEURAIS	16
2.1 <i>Introdução</i>	16

2.2	<i>Histórico</i>	17
2.3	<i>Neurônio Biológico</i>	19
2.4	<i>Neurônio Artificial</i>	20
2.5	<i>Representação</i>	21
2.5.1	Nomenclatura da rede	27
2.5.2	Interpretação dos parâmetros e reconhecimento de padrões.....	28
2.6	<i>Estimação</i>	28
2.6.1	Mínimos locais.....	31
2.6.2	Transformação de dados.....	31
2.6.3	Diminuição do peso (<i>weight decay</i>).....	31
2.7	<i>Seleção do modelo</i>	32
2.8	<i>Avaliação do modelo</i>	33
2.9	<i>Previsão</i>	34
3	IMPLEMENTAÇÃO NUMÉRICA: SIMULAÇÕES E ANÁLISE DE DADOS REAIS.....	36
3.1	<i>Comparação entre modelos</i>	36
3.1.1	Série de tempo nº1: meteorologia	36
3.1.2	Série de tempo nº2: saúde.....	40
3.1.3	Série de tempo nº3: finanças.....	42
3.2	<i>Identificação de outliers e quebras estruturais com RNAs</i>	45
	CONCLUSÃO	51
	REFERÊNCIAS	52
	ANEXOS	56

LISTA DE FIGURAS

FIGURA 1: NEURÔNIO BIOLÓGICO.....	19
FIGURA 2: NEURÔNIO ARTIFICIAL.....	20
FIGURA 3: “ESQUELETO” DA FUNÇÃO $F(x_t; \theta)$ DE UMA RNA (2.4) COM UMA ÚNICA ENTRADA E $Q=3$	24
FIGURA 4: PARTIÇÃO DO ESPAÇO (Y_{T-1}, Y_{T-2}) POR UM MODELO RNA (2.4) COM $Q=4$	25
FIGURA 5: REPRESENTAÇÃO DO MODELO DE REDE <i>SINGLE HIDDEN LAYER FEEDFORWARD</i> COM $K=4$ E $Q=2$..	27
FIGURA 6: TEMPERATURA MÉDIA, EM °C, NA CIDADE DE RECIFE.....	37
FIGURA 7: FUNÇÕES DE AUTOCORRELAÇÃO E DE AUTOCORRELAÇÃO PARCIAL DA SÉRIE Nº1.....	37
FIGURA 8: TAXA DE FERTILIDADE NA SUÉCIA, X1000.....	40
FIGURA 9: FUNÇÕES DE AUTOCORRELAÇÃO E DE AUTOCORRELAÇÃO PARCIAL DA SÉRIE Nº2.....	40
FIGURA 10: RETORNO DOS VALORES DO ÍNDICE SP500.....	43
FIGURA 11: FUNÇÕES DE AUTOCORRELAÇÃO E DE AUTOCORRELAÇÃO PARCIAL DA SÉRIE Nº3.....	43
FIGURA 12: SÉRIE SIMULADA “CONTAMINADA” COM UM <i>OUTLIER</i> ADITIVO	46
FIGURA 13: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> ADITIVO, $P=2$ E $Q=1$	47
FIGURA 14: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> ADITIVO, $P=2$ E $Q=2$	47
FIGURA 15: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> ADITIVO, $P=2$ E $Q=3$	47
FIGURA 16: SÉRIE SIMULADA “CONTAMINADA” COM UM <i>OUTLIER</i> INOVATIVO.....	48
FIGURA 17: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> INOVATIVO, $P=2$ E $Q=1$	48
FIGURA 18: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> INOVATIVO, $P=2$ E $Q=2$	48
FIGURA 19: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM <i>OUTLIER</i> INOVATIVO, $P=2$ E $Q=3$	49

FIGURA 20: SÉRIE SIMULADA “CONTAMINADA” COM MUDANÇA DE NÍVEL.....	49
FIGURA 21: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM MUDANÇA DE NÍVEL, $P=2$ E $Q=1$	49
FIGURA 22: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM MUDANÇA DE NÍVEL, $P=2$ E $Q=2$	50
FIGURA 23: SAÍDA DA CAMADA OCULTA PARA SÉRIE SIMULADA, COM MUDANÇA DE NÍVEL, $P=2$ E $Q=3$	50
FIGURA 24: RESÍDUOS DO MODELO SARIMA $(3,0,0) \times (0,1,1)$ APLICADO NA SÉRIE Nº1	58
FIGURA 25: RESÍDUOS DO MODELO ARIMA $(4,0,0)$ APLICADO NA SÉRIE Nº2.....	58
FIGURA 26: RESÍDUOS DO MODELO ARIMA $(3,0,0)$ APLICADO NA SÉRIE Nº3	59

LISTA DE TABELAS

TABELA 1: CRITÉRIO AIC PARA MODELOS ARIMA SUGERIDOS, SÉRIE Nº1.....	38
TABELA 2: PREVISÕES PELO MODELO ARIMA (3,0,0)x(0,1,1), SÉRIE Nº1.....	38
TABELA 3: PREVISÕES PELO MODELO RNA (15,3), SÉRIE Nº1.....	39
TABELA 4: COMPARAÇÃO ENTRE MÉTODOS, SÉRIE Nº1.....	39
TABELA 5: CRITÉRIO AIC PARA MODELOS ARIMA SUGERIDOS, SÉRIE Nº2.....	41
TABELA 6: PREVISÕES PELO MODELO ARIMA (4,0,0), SÉRIE Nº2.....	41
TABELA 7: PREVISÕES PELO MODELO RNA (4,2), SÉRIE Nº2.....	42
TABELA 8: COMPARAÇÃO ENTRE MÉTODOS, SÉRIE Nº2.....	42
TABELA 9: CRITÉRIO AIC PARA MODELOS ARIMA SUGERIDOS, SÉRIE Nº3.....	44
TABELA 10: PREVISÕES PELO MODELO ARIMA (3,0,0), SÉRIE Nº3.....	44
TABELA 11: PREVISÕES PELO MODELO RNA (3,2), SÉRIE Nº3.....	45
TABELA 12: COMPARAÇÃO ENTRE MÉTODOS, SÉRIE Nº3.....	45
TABELA 13: MÉDIA MENSAL DA TEMPERATURA NA CIDADE DE RECIFE ENTRE 1953 E 1962 (EM °C).....	56
TABELA 14: TAXA DE FERTILIDADE NA SUÉCIA, MEDIDA ANUALMENTE (1750-1849), x1000.....	56
TABELA 15: RETORNOS MENSAIS DOS VALORES DO ÍNDICE SP500.....	56

INTRODUÇÃO

A previsão de séries de tempo sempre foi um tema de grande interesse para profissionais de diversas áreas, como economia, meteorologia, saúde, agricultura, etc. Prever o comportamento de uma série de tempo que representa o preço de ações, por exemplo, é de suma importância para um economista, a fim de que este obtenha um melhor rendimento financeiro. Um agricultor necessita saber como a temperatura irá variar nos dias seguintes para saber se deve ou não plantar suas sementes. Enfim, pode-se perceber que as previsões de séries temporais são essenciais em diversos ramos de atividades.

Para realizar este tipo de previsão, existem diversas técnicas desenvolvidas. Neste trabalho, é dada ênfase a dois diferentes tipos de modelagem de séries: primeiramente, aos modelos ARIMA, implementados por Box e Jenkins na década de 70; após, aos modelos de redes neurais artificiais, que tiveram início na década de 40, e até hoje vêm tendo novos modelos propostos. Os modelos de redes neurais vêm sendo cada vez mais utilizados atualmente e têm como principal triunfo a capacidade de modelar quase que qualquer função não-linear.

Os objetivos principais deste trabalho são comparar a performance das previsões feitas pelas redes neurais artificiais e pelos modelos ARIMA, bem como explorar algumas das propriedades das redes neurais. Recentes trabalhos que usaram ambos os modelos para previsão indicaram que as redes neurais artificiais parecem produzir melhores resultados. Neste trabalho, as análises dos modelos ARIMA foram feitas com o *software* R, enquanto na análise dos modelos de redes neurais artificiais usou-se o *software* WinRats.

Nestas análises, foram utilizados dados reais, e os resultados são comunicados nas considerações finais do trabalho. Em tópico extra aqui abordado, é evidenciado como as redes neurais têm capacidade de modelar séries que possuem *outliers* e mudanças de nível. Neste caso, foram feitas simulações, novamente no *software* R.

Antes das análises apresentadas, é mostrada teoria sobre cada um dos modelos, com suas particularidades. Este trabalho não tem como objetivo cobrir completamente cada um dos modelos, mas sim mostrar o essencial, necessário para o entendimento e compreensão de tudo que foi feito.

DESENVOLVIMENTO

1 MODELOS ARIMA

Uma série temporal é um conjunto de pontos experimentais observados ao longo do tempo. Espera-se que uma série temporal tenha correlação seriada no tempo, devido ao fato de os dados terem sido coletados em seqüência. O que a metodologia de Box-Jenkins faz é justamente identificar o comportamento da correlação seriada ou autocorrelação entre os valores da série de tempo. Com base nisso, um modelo é sugerido e estimado, com o objetivo final de realizar previsões futuras. Esses modelos sugeridos são conhecidos genericamente por ARIMA (*Autoregressive Integrated Moving Averages*, ou Auto-regressivos Integrados de Médias Móveis).

Os modelos ARIMA são uma combinação de três componentes, que podem ser interpretados como “filtros”: o componente auto-regressivo (AR), o componente de médias móveis (MA) e o filtro de integração (I). Uma série temporal não precisa necessariamente incorporar os três componentes; um subconjunto deles já pode ser suficiente.

1.1 *Processos estocásticos*

Um *processo estocástico* pode ser definido como uma coleção de variáveis aleatórias ordenadas em algum índice – por exemplo, no tempo. A teoria de processos estocásticos é discutida aqui apenas nos aspectos relevantes particulares a análises de séries de tempo.

Nestas análises de uma série de tempo, é normalmente impossível fazer mais de uma observação no tempo t . Contudo, podemos considerar a série de tempo observada como apenas um exemplo de um infinito conjunto de séries que podem ser observadas. A série particular observada pode ser vista como uma realização particular, e é denotada por $X(t)$, $0 \leq t \leq T$, se as observações são contínuas, ou por X_t , $t = 1, 2, \dots, n$, se as observações são discretas.

Uma maneira de descrever parcialmente um processo estocástico é através de seus *momentos*, mais

exatamente o primeiro e o segundo, conhecidos como a média, a variância e as funções de autocovariância. A média $\mu(t)$ é definida como $\mu(t) = E(X_t)$, a variância $\sigma^2(t)$ como $\sigma^2(t) = Var(X_t)$ e a função de autocovariância $\gamma(t_1, t_2)$ como $\gamma(t_1, t_2) = Cov(X_{t_1}, X_{t_2}) = E\{[X_{t_1} - \mu(t_1)][X_{t_2} - \mu(t_2)]\}$.

1.2 Processos estacionários

Uma classe importante dos processos estocásticos são os chamados estacionários. De modo informal, podemos dizer que um processo estocástico é estacionário quando se pode assumir que ele está em “equilíbrio”. Para um processo ser estritamente estacionário, necessita-se que a distribuição conjunta seja a mesma ao longo do tempo. Isto é, a distribuição conjunta de (X_1, X_2, X_4) deve ser a mesma de (X_7, X_8, X_{10}) , dependendo unicamente da distância entre os valores, e jamais da sua posição no tempo. Assim, é comum definir a estacionariedade de modo menos restrito, classificando os processos como *fracamente estacionários*. Considera-se um processo fracamente estacionário caso sua média e sua variância se mantenham constantes com o passar do tempo e caso a função de autocovariância dependa somente da defasagem entre os instantes de tempo. Ou seja, para um processo X_t ser estacionário, as seguintes condições devem estar satisfeitas:

- I) $E(X_t) = \mu$;
- II) $E(X_t^2) < \infty$;
- III) $Cov[X(t), X(t + \tau)] = \gamma(\tau)$.

1.3 Função de autocorrelação

A função de autocorrelação de um processo estocástico estacionário é uma importante ferramenta para o acesso a suas propriedades. Se um processo estocástico estacionário X_t possui média μ , variância σ^2 , função de autocovariância $\gamma(\tau)$ e função de autocorrelação $\rho(\tau)$, então

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2}.$$

A função de autocorrelação possui as seguintes propriedades:

- I) $\rho(\tau) = \rho(-\tau)$;
- II) $|\rho(\tau)| \leq 1$;
- III) Só há um particular processo estocástico normal possível para cada função de

autocorrelação. Jenkins e Watts (1968) mostram um exemplo de dois diferentes processos estocásticos com a mesma função de autocorrelação; porém, estes não são normalmente distribuídos.

1.4 Modelos de médias móveis (MA)

Um processo a_t será um ruído branco se satisfizer as seguintes condições:

$$I) E(a_t) = \mu;$$

$$II) Var(a_t) = \sigma_a^2;$$

$$III) Cov_a(s, t) = 0, \text{ para } s \neq t.$$

Suponha que a_t seja um ruído branco. Então, a série Y_t é um processo de médias móveis de ordem q (de notação MA(q)) se

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1.1)$$

e θ_i é uma constante, para $i = 1, 2, \dots, q$ e $\mu = E(Y_t)$.

Neste trabalho, será considerado $E(Y_t) = 0$. A variância do processo é dada por $Var(Y_t) = \sigma_a^2 \sum_{i=0}^q \theta_i^2$, desde que os a_i 's sejam independentes.

A função de autocovariância é dada por

$$\gamma(k) = Cov(Y_t, Y_{t+k}) = \sigma_a^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k},$$

para $k = 0, 1, \dots, q$, e assume valor zero se $k > q$. A função de autocorrelação de um processo MA(q) é dada por

$$\rho(k) = \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^q \theta_i^2},$$

para $k = 1, 2, \dots, q$, assume valor zero se $k > q$, e assume valor um, se $k = 0$.

Embora não haja restrições em θ_i para o processo MA ser estacionário, Box e Jenkins (1970) propõem restrições a θ_i , dando-lhes a chamada *condição de invertibilidade*. Esta condição pode ser expressa

através do operador defasagem, denotado por B , e definido por $B^j Y_t = Y_{t-j}$, para todo j . Assim, a equação em (1.1) pode ser reescrita como

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t = Y_t, \quad (1.2)$$

onde $\theta(B)$ é um polinômio de ordem q em B . Para um processo ser invertível, é necessário que as raízes da equação

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0$$

sejam maiores do que um.

1.5 Modelos auto-regressivos (AR)

Suponha que a_t seja um ruído branco. Então, a série Y_t é um processo auto-regressivo de ordem p (de notação $AR(p)$) se

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t \quad (1.3)$$

onde ϕ_i é uma constante, para $i = 1, 2, \dots, p$, e c é um valor constante, ao qual, neste trabalho, será atribuído o valor zero.

Percebe-se que este é um modelo de regressão múltipla. Contudo, Y_t é regredido em seus valores passados. Daí o termo *auto-regressivo*.

Novamente, podemos apresentar a equação escrita em (1.3) através da utilização do operador de defasagem B .

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \phi(B) Y_t = a_t \quad (1.4)$$

Considere o processo de primeira ordem, por questão de simplicidade. Encontra-se $E(Y_t) = 0$ e $Var(Y_t) = \sigma_a^2$. A função de autocovariância é dada por

$$\gamma(k) = \sigma_a^2 \sum_{i=0}^{\infty} \phi_i \phi_{i+k},$$

para $k \geq 0$, e a função de autocorrelação é dada por $\rho(k) = \phi^k$, para $k = 0, 1, \dots$. Para o modelo ser estacionário, é necessário que $|\phi| < 1$. Quando ϕ é positivo, a função de autocorrelação decai exponencialmente; quando é negativo, ocorre o mesmo, porém com alternância de sinais positivos e negativos (considerando um processo estacionário).

1.6 Modelos auto-regressivos e de médias móveis (ARMA)

Por vezes, pode ser necessário utilizar um alto número de parâmetros em modelos unicamente auto-regressivos ou unicamente de médias móveis. Nesses casos, é mais adequado usar uma mistura dos componentes de um modelo AR com os de um modelo MA, gerando, desta forma, um modelo ARMA. O modelo ARMA(p, q), assim, exigirá um número menor de termos e pode ser expresso por esta equação:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + c \quad (1.5)$$

A função de autocorrelação dos modelos ARMA(p, q) possui características semelhantes às da função MA(q) para as defasagens $k < q$, devido à “memória” do componente de médias móveis durar exatos q períodos. Para defasagens maiores de $k + 1$, as características são similares às de um modelo AR(p).

1.7 Processos não-estacionários

Uma série temporal não é estacionária se, por exemplo, não apresentar média ou variância constantes ao longo do tempo. O fato de a média depender do tempo é sugerido quando uma série apresenta tendência (determinística ou estocástica), ou seja, os valores das variáveis não permanecem ao redor de uma linha horizontal ao longo do tempo. Quando a variação dos dados não permanece constante, com flutuações que aumentam e diminuem, está caracterizada a mudança de variância ao longo do tempo.

Segundo Chatfield (1989), na prática, a grande maioria das séries de tempo é não-estacionária. A fim de detectar a não-estacionariedade de uma série, o seu comportamento pode ser analisado graficamente através da série original e da função de autocorrelação. Também podem ser utilizados testes estatísticos de raiz unitária (o mais comumente usado é o de Dickey-Fuller).

1.8 Modelos auto-regressivos integrados de médias móveis (ARIMA)

Uma vez que a maioria dos procedimentos de análise de séries temporais parte da condição que estas sejam estacionárias, é necessário transformá-las caso não sejam. A transformação mais corriqueira, sugerida por Box e Jenkins (1970), consiste em tomar sucessivas diferenças da série original até que esta se torne estacionária.

A primeira diferença é definida por

$$\Delta Y_t = Y_t - Y_{t-1}, \quad (1.6)$$

enquanto a segunda é definida por

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} - Y_{t-2}. \quad (1.7)$$

Normalmente, bastará tomar uma ou duas diferenças para que a série se torne estacionária. O número de diferenças necessárias para tornar a série estacionária é chamado ordem de integração. A sua inclusão no termo de ordem de integração viabiliza o uso do modelo ARIMA(p, d, q) dado a seguir:

$$W_t = \phi_1 W_{t-1} + K + \phi_p W_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - K - \theta_q \alpha_{t-q} \quad (1.8)$$

em que

$$W_t = \Delta^d Y_t.$$

Novamente, pode-se reescrever o modelo acima descrito com a utilização do operador de defasagem B .

$$(1 - \phi_1 B - K - \phi_p B^p) W_t = (1 - \theta_1 B - K - \theta_q B^q) \alpha_t,$$

com $W_t = (1 - B)^d Y_t$, ou ainda,

$$\phi(B)(1 - B)^d Y_t = \theta(B) \alpha_t. \quad (1.9)$$

1.9 Modelos sazonais de Box-Jenkins

Freqüentemente, há componentes periódicos em uma série de tempo que tendem a se repetir com um período de s unidades de tempo. Box e Jenkins (1970) generalizaram os modelos ARIMA para estes poderem lidar com a sazonalidade. Estes modelos são conhecidos como SARIMA (*Seasonal Autoregressive Integrated Moving Average*).

Os modelos SARIMA possuem uma parte não sazonal, com os parâmetros p , d e q , e uma sazonal, com parâmetros P , D e Q . O modelo mais geral é dado pela seguinte equação:

$$\begin{aligned} (1 - \phi_1 B - K - \phi_p B^p)(1 - \Phi_1 B^s - K - \Phi_P B^{Ps})(1 - B)^d (1 - B^s)^D Y_t = \\ (1 - \theta_1 B - K - \theta_q B^q)(1 - \Theta_1 B^s - K - \Theta_Q B^{Qs}) \alpha_t, \end{aligned} \quad (1.10)$$

em que

$(1 - \phi_1 B - K - \phi_p B^p)$ é o polinômio auto-regressivo não-sazonal de ordem p ;

$(1 - \Phi_1 B^s - K - \Phi_P B^{Ps})$ é o polinômio auto-regressivo sazonal de ordem P e período sazonal s ;

$(1 - B)^d$ é o polinômio de integração não-sazonal de ordem d ;

$(1 - B^s)^D$ é o polinômio de integração sazonal de ordem D e período sazonal s ;

$(1 - \theta_1 B - K - \theta_q B^q)$ é o polinômio não-sazonal de médias móveis de ordem q ;

$(1 - \Theta_1 B^s - K - \Theta_Q B^{Qs})$ é o polinômio sazonal de médias móveis de ordem Q e período sazonal s .

1.10 Metodologia de Box-Jenkins para previsão

Esta metodologia foi desenvolvida durante muitos anos por Box e Jenkins. Nesta seção, haverá apenas uma síntese dos principais aspectos relacionados a esta metodologia de previsão.

O procedimento é baseado no ajuste de modelos SARIMA. Segundo Box e Jenkins (1970), existem três etapas na construção do modelo:

I) Identificação: consiste na descoberta de qual modelo SARIMA descreve de forma adequada o comportamento da série. Para essa identificação, as técnicas de Box-Jenkins baseiam-se nas propriedades das funções de autocorrelação (ACF) e das funções de autocorrelação parciais (PACF).

II) Estimação: nesta etapa, estimam-se os parâmetros ϕ e Φ do componente auto-regressivo, θ e Θ do componente de médias móveis e a variância de a_t .

III) Verificação: avalia-se a adequabilidade do modelo estimado para a descrição do comportamento dos dados. A verificação será feita com a análise dos resíduos estimados do modelo.

O primeiro passo no procedimento de Box-Jenkins é diferenciar os dados até que estes se tornem estacionários. Isto é alcançado através do exame dos gráficos da função de autocorrelação de várias séries diferenciadas, até que se encontre um que decresça a zero quase que instantaneamente. Para dados não-sazonais, a diferenciação de primeira ordem provavelmente será suficiente. Para dados mensais, algumas vezes usa-se o operador diferença sazonal Δ^{12} . Diferenciações em excesso devem ser evitadas. Para dados trimestrais, deve-se usar o operador Δ^4 , e assim por diante.

Se os dados não são sazonais, pode-se ajustar um modelo ARMA de acordo com (1.8). Se forem sazonais, o modelo geral ARMA sazonal pode ser ajustado a partir da equação (1.10). Valores razoáveis de P , p , Q e q devem ser escolhidos a partir da observação do gráfico das funções de autocorrelação e autocorrelação parcial da série diferenciada.

Após identificar o modelo que aparenta ser o mais adequado, estimativas dos parâmetros do modelo podem ser obtidas pela máxima verossimilhança condicional de um modo similar ao proposto para os modelos ARMA usuais.

A adequação do modelo ajustado deve ser validada pelo que Box e Jenkins chamam de *verificação de diagnóstico*. Isto consiste, essencialmente, em examinar os resíduos do modelo ajustado para ver se há

indícios de que estes não são um ruído branco.

Se o modelo não for considerado adequado, volta-se à fase de identificação e repetem-se as três etapas. É comum identificar mais de um modelo para serem estimados e verificados. No momento em que se obtém um resultado satisfatório, é hora da etapa final da metodologia de Box-Jenkins: realizar previsões. Uma aplicação prática desse processo ocorrerá com a aplicação da análise estatística sobre um conjunto de dados na seção 3.1.

1.11 Estimação e verificação

Após identificar um modelo provisório para a série temporal, o passo seguinte é estimar seus parâmetros. Em dado momento, será necessário usar um procedimento iterativo de estimação de mínimos quadrados não-linear, e as estimativas preliminares encontradas na fase de identificação serão usadas como valores iniciais neste procedimento.

Considere um modelo ARIMA (p,d,q) e coloque seus $p+q+1$ parâmetros no vetor $\underline{\xi} = (\underline{\phi}, \underline{\theta}, \sigma_a^2)$, onde $\underline{\phi} = (\phi_1, K, \phi_p)$ e $\underline{\theta} = (\theta_1, K, \theta_q)$. Aqui, quando $d > 0$, está sendo suposto que $\mu_W = 0$. Caso contrário, μ é incluído como mais um parâmetro a ser estimado, e tem-se $p+q+2$ parâmetros.

Para estimar $\underline{\xi}$, emprega-se o método de máxima verossimilhança. Dadas as observações X_1, \dots, X_n , considere a função de verossimilhança $L(\underline{\xi} | X_1, K, X_n)$, encarada como função de $\underline{\xi}$. Os estimadores de máxima verossimilhança (EMV) de $\underline{\xi}$ serão os valores que maximizam L ou $l = \log L$.

Tomando-se d diferenças para alcançar estacionariedade, fica-se com $n = N - d$ observações W_1, \dots, W_n , onde $W_t = \Delta^d Y_t$. Como o modelo ARMA (p,q) resultante é estacionário e invertível, tem-se a equação

$$a_t = \tilde{W}_t - \phi_1 \tilde{W}_{t-1} - K - \phi_p \tilde{W}_{t-p} + \theta_1 a_{t-1} + K + \theta_q a_{t-q}, \quad (1.11)$$

onde $\tilde{W}_t = W_t - \mu_W$.

Para calcular os a_t 's através de (1.11), é necessário obter valores iniciais para os \tilde{W} 's e para os a 's. Esta questão de valores iniciais pode ser resolvida através de dois procedimentos: um, condicional, em que os valores iniciais desconhecidos são substituídos por valores que supõe-se serem razoáveis; outro,

incondicional, em que os valores iniciais são estimados de uma amostra de dados.¹

Após a estimação do modelo, é necessário verificar se ele representa bem os dados. A verificação é feita através da análise dos resíduos ajustados. Suponha que o modelo ajustado seja $\phi(B)Y_t = \theta(B)a_t$. Assim, os resíduos estimados são dados por

$$\hat{a}_t = \frac{\hat{\phi}(B)}{\hat{\theta}(B)} Y_t. \quad (1.12)$$

Portanto, pode-se utilizar testes para as funções de autocorrelação e autocorrelação parcial dos resíduos estimados. Para o modelo ser confiável, os resíduos devem “parecer” com um ruído branco.

Box e Pierce (1970) sugerem um teste global baseado na estatística $Q = n \sum_{k=1}^K (\rho_Y^T(k))^2$, onde rejeita-se a hipótese nula para valores altos de Q , e a hipótese nula sugere que as funções de autocorrelação dos resíduos estimados de um ajuste de um modelo ARIMA (p,q) são nulas até o *lag* j .

Caso os resíduos estimados pareçam um ruído branco, então o modelo ajustado é um candidato a ser escolhido como modelo final. Caso haja outras escolhas, há critérios que comparam os modelos escolhidos. Por outro lado, caso os resíduos estimados não se adequem a hipótese de ruído branco, modificações devem ser realizadas neste modelo a fim de que os resíduos “pareçam” um ruído branco. Os critérios mais comumente utilizados são o AIC (Critério de Informação de Akaike) e o BIC (Critério de Informação Bayesiano de Schwarz). Estes critérios avaliam quantitativamente os modelos escolhidos e indicam o mais adequado.

1.12 Previsão

Finalmente, pode-se utilizar o modelo identificado e estimado para realizar previsões. Suponha que temos as observações $(\dots, X_{t-2}, X_{t-1}, X_t)$, até o instante t , $t=1, 2, \dots, T$, que é chamado *origem das previsões*. O interesse é prever um valor para X_{t+h} , $h \geq 1$. A previsão de origem t e horizonte h é denotada por $\hat{x}_t(h)$.

Em um modelo ARIMA, é razoável supor que $\hat{x}_t(h)$ é uma função das observações até o instante t , conseqüentemente, uma função de a_t, a_{t-1}, \dots . Se a melhor previsão é indicada por

$$\hat{x}_t(h) = \Psi_h^* a_t + \Psi_{h+1}^* a_{t-1} + \Psi_{h+2}^* a_{t-2} + \dots + K \quad (1.13)$$

¹ Para mais detalhes, ver Morettin e Toloi (1981).

os pesos Ψ_j^* devem ser determinados a fim de minimizar o erro quadrático médio (EQM), que é dado por

$$E(X_{t+h} - \hat{x}_t(h))^2 = E\left(\sum_{j=0}^{\infty} \Psi_j a_{t+h+j} - \sum_{j=0}^{\infty} \Psi_{h+j}^* a_{t-j}\right)^2 \quad (1.14)$$

O erro de previsão é dado por

$$\begin{aligned} e_t(h) &= X_{t+h} - \hat{x}_t(h) \\ &= \Psi_0 a_{t+h} + \Psi_1 a_{t+h-1} + K + \Psi_{h-1} a_{t+1} - \sum_{j=0}^{\infty} (\Psi_{h+j} - \Psi_{h+j}^*) a_{t-j} \end{aligned} \quad (1.15)$$

Logo, o EQM pode também ser escrito como

$$E(e_t(h))^2 = (1 + \Psi_1^2 + K + \Psi_{h-1}^2) \sigma_a^2 + \sum_{j=0}^{\infty} (\Psi_{h+j} - \Psi_{h+j}^*)^2 \sigma_a^2, \quad (1.16)$$

uma vez que os a_t 's não são autocorrelacionados e têm média zero.

Assim, a previsão de EQM mínimo é dada por

$$\hat{x}_t(h) = \Psi_h a_t + \Psi_{h+1} a_{t-1} + \Psi_{h+2} a_{t-2} + K. \quad (1.17)$$

2 MODELOS DE REDES NEURAIIS

2.1 Introdução

Um ser humano possui circuitos neurais cerebrais complexos, que são formados por diversas conexões entre seus neurônios (sinapses). Estes interagem entre si, fazendo emergir comportamento inteligente. Desta forma, nasce a idéia de que é possível fazer surgir comportamento igualmente inteligente em máquinas, caso conseguíssemos modelar computacionalmente estas conexões neurais. É neste contexto que surgem as redes neurais artificiais (RNAs), inspiradas nas redes de neurônios e sinapses biológicas. Esta proposta de modelagem cerebral forma o ramo da Inteligência Artificial denominado Inteligência Artificial Conexionista.

A existência de classes de problemas cuja complexidade algorítmica é imensa, de modo a se tornar praticamente impossível sua resolução via algoritmos convencionais, é outro ponto importante para a utilização das redes neurais artificiais. Todavia, o ser humano tem mecanismos próprios que solucionam problemas providos de uma certa complexidade, como o reconhecimento da fala ou a segmentação de imagens de maneira inata.

Além disso, as RNAs são toleráveis a ruídos e falhas, conseqüência da capacidade de generalização das redes, capazes de manter o resultado desejado ao suportar estes ruídos e imperfeições do ambiente.

Apesar do pouco conhecimento que se tem do cérebro humano, além da dificuldade que é modelar mesmo o que já sabemos, as pesquisas sobre redes neurais vêm se mostrando bastante promissoras. Possuem grande destaque em áreas como engenharia, computação e até mesmo nas neurociências, sendo de grande utilidade para problemas como reconhecimento de padrões, agrupamentos, previsão de séries temporais, etc.

Segundo Franses e van Dijk (2000), entre 20 e 30 artigos sobre redes neurais lidando com modelagem e previsão de séries temporais são publicados a cada ano².

A principal razão para o crescimento da popularidade das RNAs no ramo da previsão de séries temporais é que seus modelos podem aproximar quase qualquer função não-linear. Conseqüentemente, quando aplicadas em séries de tempo caracterizadas por relações não-lineares, as RNAs detectam este comportamento e geram um ajuste melhor do que os modelos de séries temporais tradicionais.

Uma desvantagem das RNAs é o fato de os parâmetros serem de difícil, senão impossível, interpretação. Uma RNA estimada não nos dá informação sobre quais tipos de modelos paramétricos de séries de tempo podem ser adequados para descrever os padrões não-lineares que são detectados. Por isso, considera-se que as RNAs contêm uma “caixa preta” e são construídas principalmente com a proposta de reconhecimento de padrões e previsões futuras.

Contudo, um ajuste de forma muito exata dentro da amostra pode dificultar a previsão da RNA fora da amostra. O perigoso ajuste “em excesso” (*overfitting*) pode se tornar um problema. Aumentando a flexibilidade do modelo, é possível obter um ajuste quase perfeito dentro da amostra, com medidas como R^2 chegando a valores próximos da unidade. Neste caso, todavia, pode-se estar modelando o erro aleatório, prejudicando a capacidade do modelo para previsões fora da amostra.

2.2 Histórico

Os estudos iniciais sobre redes neurais artificiais, assim como propostas de modelos para estas redes, surgiram nos anos 40. Os primeiros modelos evoluíram bastante - alguns deles se destacaram e tornaram-se famosos. Mesmo assim, até hoje são propostos novos modelos de redes neurais.

No início dos anos 40, surgiram as primeiras pesquisas sobre redes neurais. Idealizando

² Para alguns exemplos, ver Gençay (1996, 1999), Häfke e Helmenstein (1996a, 1996b), Gençay e Stengos (1998), Qi e Maddala (1995, 1999), Kuan e Liu (1995), Franses e van Griensven (1998), Franses e van Homelen (1998), Swanson e White (1995) e Hutchinson, Lo e Poggio (1994).

procedimentos matemáticos análogos à forma de funcionamento dos neurônios humanos, McCulloch e Pitts (1943) estabeleceram as bases da neurocomputação. Na literatura, considera-se esta idéia meramente ilustrativa, uma vez que os autores não desenvolveram aplicações práticas para seu trabalho. Inclusive, os sistemas por eles propostos não tinham habilidade nem sequer capacidade de aprender.

O primeiro grande passo na história das redes neurais foi dado por Hebb (1949), que sugeriu uma maneira de gerar capacidade de aprendizado às redes neurais artificiais. De um certo modo, sua sugestão veio a viabilizar a criação de redes neurais eficazes.

Na década de 50, diversas implementações de redes neurais já haviam sido desenvolvidas. Acreditava-se que havia sido descoberto o caminho para o entendimento da inteligência humana. A construção de uma rede neural razoavelmente grande seria suficiente para descrever o comportamento do cérebro humano. Frank Rosenblatt (1959) criou o *perceptron*, que deu origem aos modelos baseados no aprendizado por correção de erros. Outros modelos similares ao *perceptron* também foram desenvolvidos nesta época.

Os modelos baseados no *perceptron* receberam uma forte crítica de Minsky e Papert (1969). Nesta publicação, os autores demonstraram que estes modelos de redes neurais baseados nos *perceptrons* não eram capazes de aprender uma função lógica trivial do tipo “ou-exclusivo” (*Exclusive Or*). Esta função possui um padrão de valores de entrada e saída incompreensível pelos modelos de redes neurais baseados em *Perceptrons* disponíveis naquela época. O impacto desta publicação abalou os pesquisadores desta área de estudo. Após esta rígida análise matemática, quando ficou comprovado o baixo poder computacional desses modelos, as pesquisas neste campo ficaram esquecidas até meados dos anos 80.

No início da década de 80, retomou-se o interesse pelas redes neurais, principalmente devido aos novos modelos neurais desenvolvidos. Hopfield (1982) criou um modelo importante de rede, sendo esta uma rede com conexões recorrentes e com comportamento baseado na competição entre os neurônios, onde o aprendizado era não-supervisionado, diferentemente dos modelos baseados nos *perceptrons*.

A aparição de computadores mais potentes, junto com o desenvolvimento de modelos matemáticos que permitiam a solução do problema apontado por Minsky e Papert, contribuiu para o ressurgimento de modelos baseados em *perceptrons* na década de 80. O modelo que permitiu esse ressurgimento foi o das redes multi-nível, onde o algoritmo de aprendizado, chamado de *backpropagation* (retro-propagação), solucionou quase todos os problemas de aprendizado existentes até então. Esse algoritmo tornou-se famoso em uma obra de Rumelhart, Hinton e Williams (1986).

Além dos modelos de Hopfield e do modelo de redes multi-nível utilizando o algoritmo de retro-propagação, outro modelo importante surgido nesta época foi o de Kohonen (1987). Este modelo permite o aprendizado competitivo com uma auto-organização da rede neural, gerando os chamados “mapas de

atributos auto-organizáveis” (*self-organizing feature maps*).

Finalmente, o último modelo de destaque neste período foi o ART (*Adaptive Resonance Theory*), proposto por Carpenter e Grossberg (1987). Este modelo possui um aprendizado do tipo não-supervisionado, criando protótipos (*clusters*) dos padrões aprendidos.

Os estudos sobre redes neurais passaram por uma revolução na década de 80, conforme demonstrado acima. Desde então, cada vez mais esta área vem se destacando, seja pelas promissoras características dos modelos propostos, seja pelas condições tecnológicas atuais de implementação, que obtêm ótimas performances dos sistemas.

Na década de 90, as redes neurais tiveram uma explosão exponencial de aplicações e desenvolvimento de modelos. A cada ano, são centenas de propostas de novos ou aperfeiçoados modelos, tal o interesse pela área. A partir daí, as redes neurais consolidaram-se como parte integrante do estudo da Inteligência Artificial.

2.3 Neurônio Biológico

Um neurônio é composto por um corpo celular, ou *soma*, um axônio tubular e várias ramificações, conhecidas como dendritos. Os dendritos formam uma malha de filamentos muito fina ao redor do neurônio. Já o axônio é um tubo longo e fino que, em um de seus extremos, se divide em ramos. Nestes, formam-se os pequenos bulbos, que quase tocam os dendritos dos outros neurônios. O espaço entre o fim do bulbo e o dendrito é conhecido como sinapse, através da qual as informações se propagam. A quantidade de sinapses recebidas por cada neurônio varia entre 100 e 100.000. Na Figura 1, é mostrada uma ilustração de um neurônio biológico.

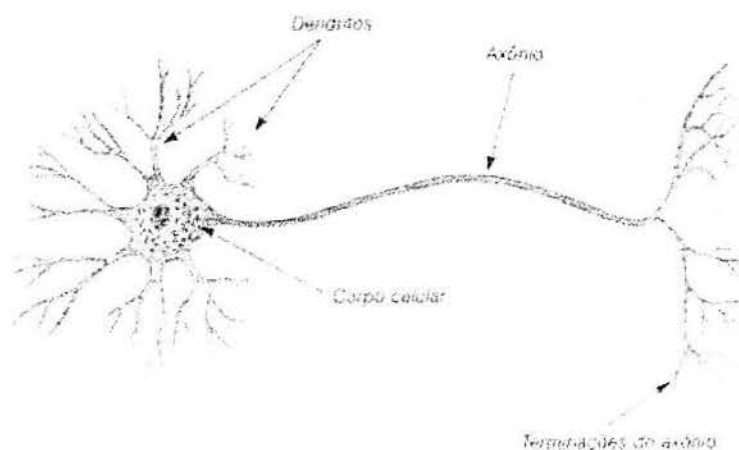


Figura 1: Neurônio biológico

Fonte: www.isurp.com.br/aula/ciencia/Marcio/osistema.htm

A concentração diferente dos íons Na^+ e K^+ dentro e fora das células dá à célula nervosa um potencial

de repouso. Devido a isso, qualquer perturbação na membrana do neurônio gera uma série de mudanças durante um curto período de tempo.

A alteração dessa concentração dos íons gera o chamado *trem de pulso*, que se expande localmente nas zonas próximas aos dendritos. Dependendo da intensidade do estímulo, o trem de pulso pode exceder a um certo limiar no corpo da célula e produzir um sinal de amplitude constante ao longo do axônio. No momento de disparo do neurônio, é gerado um potencial de ação que impulsiona para as outras células o fluxo do sinal formado pelo corpo celular.

O pulso elétrico provido desse potencial libera substâncias químicas contidas nos bulbos do axônio, chamadas neurotransmissores. Estes são repassados para os dendritos no neurônio seguinte. Desta forma, quando o conjunto de neurotransmissores que chegam aos dendritos atinge um certo limiar, eles novamente lançam um potencial de ação, que vai repetir todo o processo.

É interessante ressaltar que as sinapses podem ser excitatórias – facilitando o fluxo dos sinais elétricos – ou inibitórias, que dificultam a passagem dessa corrente³.

2.4 Neurônio Artificial

Mesmo com todos os esforços existentes para se modelar os neurônios biológicos, tudo o que se conseguiu até hoje foi uma aproximação elementar. A Figura 2 ilustra o modelo:

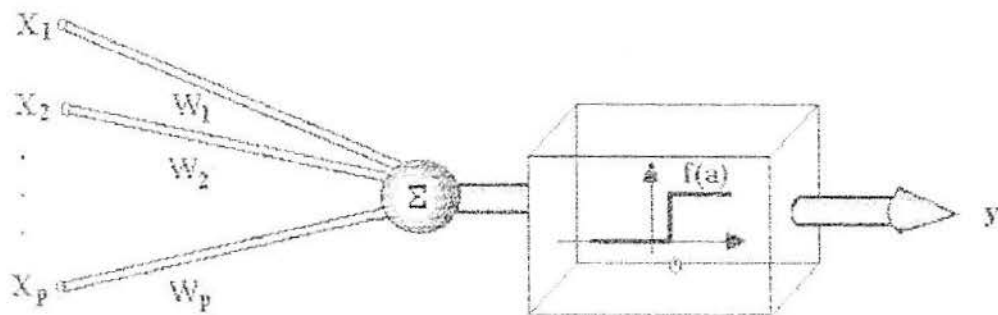


Figura 2: Neurônio artificial

Fonte: desconhecida

³ Para mais informações acerca do funcionamento do sistema neural, ver Nicholls (2000) e Lavine (1983).

Neste modelo, as entradas do neurônio X_j podem ser saídas de outros neurônios, entradas externas, uma constante (*bias*), ou qualquer combinação destes elementos. Estas entradas são ponderadas pelos pesos W_{ij} , que são inspirados na força da conexão sináptica. Assim, temos o chamado *net* do neurônio, que é a soma de todas as entradas multiplicadas por seus respectivos pesos.

$$net_i = \sum_{j=1}^p X_j W_{ij}$$

Também temos a função de ativação (ou função de transferência), que produz a saída do neurônio. O principal objetivo da função de ativação é introduzir uma não-linearidade no modelo. Sem esta função, o poder computacional do neurônio, ou mesmo de uma rede neural inteira, seria equivalente ao de um sistema linear, o que não possibilitaria a solução de classes importantes de problemas. Normalmente, a função de ativação tem forma contínua e crescente, de tal sorte que seu domínio encontra-se normalmente no âmbito dos números reais. Pode-se utilizar como função de ativação as seguintes funções:

1) *Função de limiar*: representa uma função de decisão abrupta. Descreve a propriedade “tudo ou nada” de um neurônio de McCulloch-Pitts;

2) *Função linear por partes*: resulta em uma combinação linear, se a operação for restrita à região linear;

3) *Função sigmóide*: é uma função em forma de “s”, monotonamente crescente, que exhibe um balanceamento adequado entre comportamento linear e não-linear. Exemplos deste tipo de função são a função tangente hiperbólica e a função logística, esta quase sempre indicada como a ideal⁴.

2.5 Representação

Considere o seguinte modelo para uma série de tempo

$$y_t = \phi_0 + \beta_1 G(\gamma[y_{t-1} - c]) + a_t, \quad (2.1)$$

onde $G(\cdot)$ é a função logística

$$G(z) = \frac{1}{1 + \exp(-z)}. \quad (2.2)$$

O modelo (2.1) descreve a situação onde a média condicional de y_t depende do valor de y_{t-1} relativo

⁴ Mais detalhes em Dayhoff (1998), Fausett (1994), Bishop (1995) e Bosque (2002).

ao limiar c . Para $y_{t-l} \ll c$, a média condicional de y_t é aproximadamente ϕ_0 , enquanto muda gradualmente para $\phi_0 + \beta_1$ à medida que y_{t-l} aumenta. Uma rede neural artificial pode ser obtida ao se assumir que a média condicional de y_t depende do valor de uma combinação linear de p valores defasados ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) relativos ao limiar c . O modelo (2.1) então se torna $y_t = \phi_0 + \beta_1 G(\gamma[\tilde{x}_t' \delta - c]) + a_t$, onde $\tilde{x}_t = (y_{t-1}, K, y_{t-p})'$. Após alguma manipulação, podemos escrever:

$$y_t = \phi_0 + \beta_1 G(x_t' \gamma_1) + a_t, \quad (2.3)$$

onde $x_t = (1, \tilde{x}_t)'$ e os elementos individuais do vetor de parâmetros $\gamma_1 = (\gamma_{0,1}, \gamma_{1,1}, K, \gamma_{p,1})'$ podem ser facilmente obtidos por γ, δ e c .

Esta rede neural artificial pode ser interpretada como um modelo de *switching-regression* (regressão com mudança de regime), onde a mudança é determinada por uma particular combinação linear entre as p variáveis defasadas no vetor x_t . Em aplicações de redes neurais não se costuma focar em uma interpretação do regime. Ao invés disso, o objetivo é modelar a relação entre y_t e x_t , possivelmente não-linear. Um modo de alcançar isto é incluindo componentes logísticos adicionais no modelo, chegando a

$$y_t = \phi_0 + \sum_{j=1}^q \beta_j G(x_t' \gamma_j) + a_t. \quad (2.4)$$

Suponha que uma relação subjacente apropriada entre y_t e x_t é dada por

$$y_t = g(x_t; \xi) + \eta_t,$$

onde $g(x_t; \xi)$ é uma função contínua. É possível mostrar que as RNAs da forma (2.4) podem aproximar qualquer função $g(x_t; \xi)$ com qualquer grau de precisão desejado, contanto que o número q de componentes não-lineares seja suficientemente grande. Tecnicamente, escrevendo (2.4) como

$$y_t = F(x_t; \theta) + a_t,$$

onde

$$F(x_t; \theta) = \phi_0 + \sum_{j=1}^q \beta_j G(x_t' \gamma_j),$$

pode-se mostrar que para qualquer função contínua $g(x_t; \xi)$, cada subconjunto K de \mathfrak{R}^K , e cada $\delta > 0$, existe uma rede neural artificial $F(x_t; \theta)$ tal que

$$\sup_{x \in K} |F(x; \theta) - g(x; \xi)| < \delta. ^5$$

Intuitivamente, este resultado pode ser entendido com o exemplo seguinte. Considere o modelo (2.4) e assumamos que apenas y_{t-1} age como uma entrada – isto é, $x_t = (1, y_{t-1})'$. A Figura 3 mostra o “esqueleto” da função $F(x_t; \theta)$ de uma rede com $\phi_0 = 2$, $q = 3$, $\beta_1 = 8$, $\beta_2 = -12$, $\beta_3 = 6$, $G(x'_t, \gamma_1) = 1/(1 + \exp[-40 - 10y_{t-1}])$, $G(x'_t, \gamma_2) = 1/(1 + \exp[-y_{t-1}])$ e $G(x'_t, \gamma_3) = 1/(1 + \exp[20 - 20y_{t-1}])$. Os valores das funções logísticas $G(x'_t, \gamma_j)$, para $j = 1, 2, 3$, estão dispostos ao longo do eixo horizontal. Para valores muito negativos de y_{t-1} , as três funções são iguais a zero, e $F(x_t; \theta) = \phi_0$. Por volta do ponto $y_{t-1} = -4$, a função $G(x'_t, \gamma_1)$ cresce de 0 para 1 rapidamente, e, devido a esse crescimento, o valor da função $F(x_t; \theta)$ também aumenta. Ao mesmo tempo, a função $G(x'_t, \gamma_2)$ é ativada e começa a crescer vagarosamente. A função $F(x_t; \theta)$ decresce suavemente até o ponto $y_{t-1} = 1$. Nesse momento, a função $G(x'_t, \gamma_3)$ vai de 0 a 1 quase que instantaneamente, causando um crescimento agudo no valor de $F(x_t; \theta)$. Após este ponto, a função resume-se em um gradual declínio, devido ao contínuo crescimento de $G(x'_t, \gamma_2)$. Resumidamente, o valor do “esqueleto” da rede muda como uma função da entrada y_{t-1} à medida que as funções $G(x'_t, \gamma_j)$ mudam de valor. Fica bem claro que, aumentando o número de componentes logísticos em (2.4), a função resultante $F(x_t; \theta)$ se torna mais flexível, podendo então aproximar praticamente todas as relações não-lineares que possam existir entre y_t e y_{t-1} .⁶

⁵ Ver Cybenko (1989), Carroll e Dickinson (1989), Funabashi (1989), Hornik, Stinchcombe e White (1989, 1990) para mais detalhes.

⁶ Ver Donaldson e Kamstra (1996) para um exemplo similar.

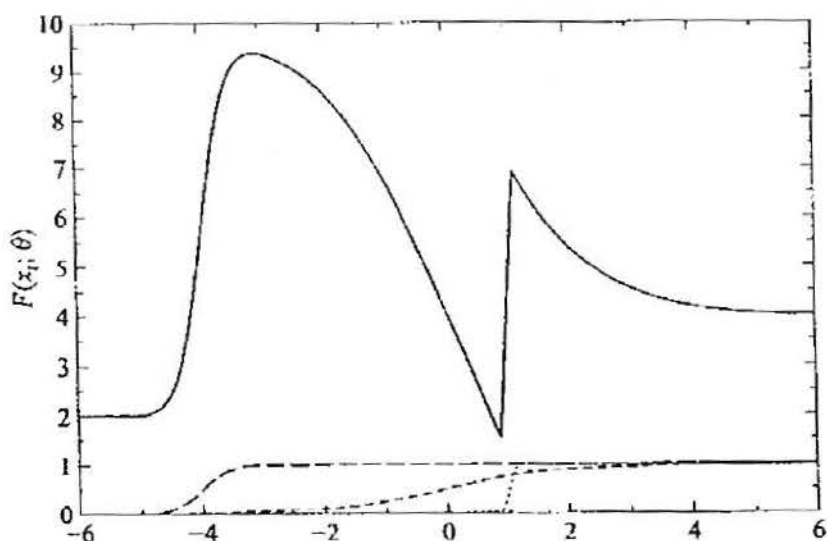


Figura 3: “Esqueleto” da função $F(x_t; \theta)$ de uma RNA (2.4) com uma única entrada e $q=3$ (linha contínua); os valores das funções de ativação $G(x_t, \gamma_j)$, para $j = 1, 2, 3$, são mostrados no eixo horizontal (linhas pontilhadas de traços maiores para os menores, respectivamente)

Fonte: Franses e van Dijk (2000)

Para ver como o exemplo acima pode ser generalizado para redes mais complicadas, que envolvem mais de uma entrada, considere uma rede em que \tilde{x}_t contenha os primeiros dois lags de y_t – ou seja, $\tilde{x}_t = (y_{t-1}, y_{t-2})'$. Então, a combinação linear $x_t' \gamma_1$ divide em dois regimes o espaço dos valores que \tilde{x}_t pode assumir: um regime para $\gamma_{1,1}y_{t-1} + \gamma_{2,1}y_{t-2} < \gamma_{0,1}$, outro para $\gamma_{1,1}y_{t-1} + \gamma_{2,1}y_{t-2} > \gamma_{0,1}$. No primeiro regime, a média condicional de y_t é igual a ϕ_0 , enquanto no segundo é igual a $\phi_0 + \beta_1$ (assumindo que a transição da função logística $G(x_t' \gamma_1)$ de 0 a 1 seja instantânea). Aumentando o número de componentes logísticos, ainda em (2.4), o número de partições do espaço de (y_{t-1}, y_{t-2}) também é aumentado, fazendo crescer efetivamente o número de regimes (cada regime é caracterizado por uma particular combinação de valores 0 e 1 da função $G(x_t' \gamma_j)$) e o número de valores distintos que a média condicional de y_t pode assumir. A Figura 4 mostra um exemplo de resultado da partição de quatro combinações lineares $x_t' \gamma_1, K, x_t' \gamma_4$. Também estão indicados os valores das funções logísticas $G(x_t' \gamma_1), K, G(x_t' \gamma_4)$ associadas, em cada um dos regimes envolvidos. Como o número de componentes logísticos em (2.4) cresce sem saltos, o particionamento do espaço (y_{t-1}, y_{t-2}) torna-se tão denso que a média condicional de y_t pode “assumir” um valor distinto para cada valor de (y_{t-1}, y_{t-2}) . Assim, qualquer tipo de relação não-linear que possa existir entre y_t, y_{t-1} e y_{t-2} pode ser aproximada deste jeito.

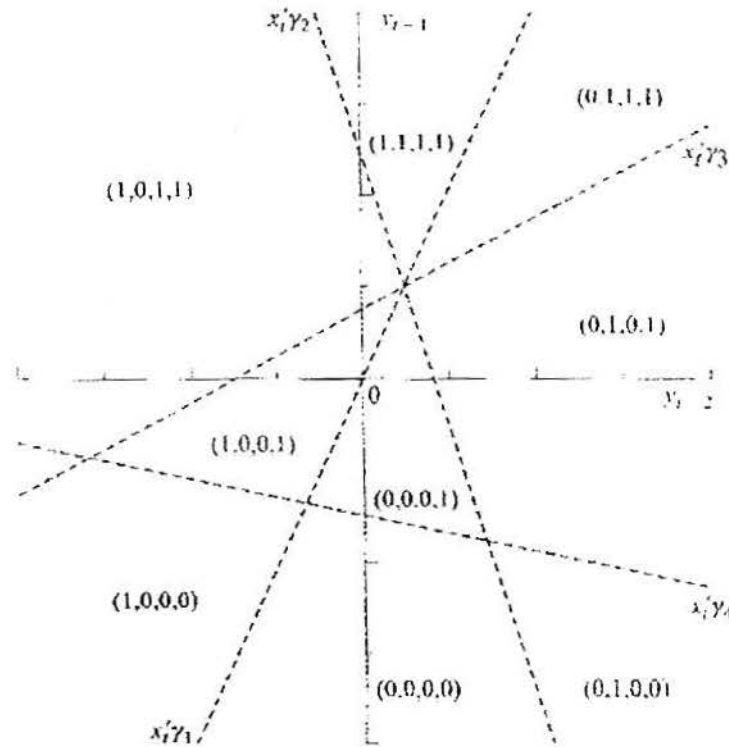


Figura 4: Partição do espaço (y_{t-1}, y_t) por um modelo RNA (2.4) com $q=4$

Fonte: Franses e van Dijk (2000)

O que foi descrito acima deixa claro que as RNAs são fortemente relacionadas com os métodos não-paramétricos, que é um termo coletivo para um grande conjunto de técnicas que, por exemplo, podem ser aplicadas para estimar uma relação entre y_t e y_{t-1} sem especificar uma forma paramétrica particular. Para verificarmos isto, consideremos o modelo auto-regressivo não-linear de primeira ordem geral

$$y_t = g(y_{t-1}) + a_t \quad (2.5)$$

e suponha que nosso interesse seja estimar o valor de $g(\cdot)$ em um ponto particular $y_{t-1} = y^*$, usando a série de tempo observada y_1, y_2, \dots, y_n . Alguns métodos não-paramétricos fazem uso do que podemos chamar de 'média local'. Por exemplo, o estimador da média condicional local é dado pela média de todas as observações y_t cujo correspondente y_{t-1} esteja dentro do intervalo $(y^* - h, y^* + h)$ para algum $h > 0$. Isto é,

$$\hat{g}(y^*) = \frac{\sum_{t=1}^n I[|y^* - y_{t-1}| < h] y_t}{\sum_{t=1}^n I[|y^* - y_{t-1}| < h]}, \quad (2.6)$$

onde $I[\cdot]$ é a função indicadora. Métodos mais sofisticados usam uma média ponderada dos y_t observados, onde os pesos normalmente dependem da distância entre y_{t-1} e y^* . Por exemplo, o estimador Kernel de

Nadaraya-Watson é dado por

$$\hat{g}(y^*) = \frac{\sum_{t=1}^n K((y^* - y_{t-1})/h)y_t}{\sum_{t=1}^n K((y^* - y_{t-1})/h)}, \quad (2.7)$$

onde $K(z)$ é uma função Kernel que satisfaz $\int K(z)dz = 1$. Usualmente, $K(z)$ é tido como uma função densidade de probabilidade unimodal, como por exemplo o Kernel gaussiano

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right).$$

Um outro método para estimar $g(y^*)$ é através da *regressão linear local*. Estes estimadores baseiam-se no fato de que a função $g(y)$ pode ser aproximada por uma função linear em uma região de vizinhança em volta de y , isto é:

$$g(y) \approx g(y^*) + g'(y^*)(y - y^*) = \beta_0 + \beta_1(y - y^*), \quad \text{para } y \text{ próximo de } y^*.$$

Os valores de $g(y^*)$ então são estimados por $\hat{\beta}_0$, onde $\hat{\beta}_0$ é obtido pela minimização da soma dos quadrados dos resíduos ponderada

$$Q_n = \frac{1}{nh} \sum_{t=1}^n (y_t - \beta_0 - \beta_1(y_{t-1} - y^*))^2 K((y_{t-1} - y^*)/h). \quad (2.8)$$

Métodos não-paramétricos constituem um campo de pesquisa amplo e de rápido desenvolvimento, sendo apenas muito brevemente mencionados aqui.⁷

Nas aplicações econométricas de RNAs, costuma-se incorporar as p variáveis defasadas em \tilde{x}_t diretamente como regressões lineares. Isto é, estende-se (2.4) para se obter a rede neural aumentada

$$\begin{aligned} y_t &= \phi_0 + \tilde{x}_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j) + \alpha_t \\ &= \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} \\ &\quad + \sum_{j=1}^q \beta_j G(\gamma_{0,j} + \gamma_{1,j} y_{t-1} + \gamma_{2,j} y_{t-2} + \dots + \gamma_{p,j} y_{t-p}) + \alpha_t. \end{aligned} \quad (2.9)$$

⁷ Ver Härdle, Lütkepohl e Chen (1997), Ziegelmann (2003) e Fan e Gijbels (1996) para mais informações sobre técnicas disponíveis.

Embora esta generalização não seja necessária, ela facilita a interpretação do modelo. O termo $\tilde{x}'_t \phi$ pode ser visto como o representante da parte linear da relação entre y_t e \tilde{x}_t , enquanto os componentes logísticos medem a porção não-linear que está presente.

A RNA em (2.9) pode ser generalizada com a inclusão de variáveis exógenas $z_{1,t}, z_{2,t}, \dots, z_{m,t}$ no vetor de entradas \tilde{x}_t . Refere-se aos elementos de \tilde{x}_t como $x_{1,t}, x_{2,t}, \dots, x_{k,t}$, $k = p + m$, de modo que $x_{i,t}$, $i = 1, 2, \dots, k$, possa ser uma variável endógena defasada y_{t-j} , $j = 1, 2, \dots, p$, ou uma variável exógena $z_{j,t}$, $j = 1, 2, \dots, m$. Denomina-se a RNA (2.9) como RNA($k, q, 1$), identificando que a rede tem k variáveis de entrada $y_{t-1}, y_{t-2}, \dots, y_{t-p}, z_{1,t}, z_{2,t}, \dots, z_{m,t}$, q componentes logísticos $G(x'_t \gamma_j)$, e uma variável de saída y_t . Devido a estar-se considerando casos com somente uma saída, pode-se abreviar para RNA(k, q).

2.5.1 Nomenclatura da rede

Considere a representação gráfica do modelo de RNA(k, q) com $k = 4$ e $q = 2$ na Figura 5. Vê-se que a rede consiste em três diferentes *camadas*. Na base, está a *camada de entrada*, que consiste nas variáveis explicativas em x_t , chamadas de *entradas*. Estas entradas são multiplicadas pelas chamadas *forças de conexão* $\gamma_{i,j}$ à medida que entram na *camada oculta*, composta por q *unidades ocultas*, isto é, as funções logísticas $G(\cdot)$. O nome ‘camada oculta’ surge do fato de não podemos observá-la diretamente. Na camada oculta, as combinações lineares $x'_t \gamma_j$ são formadas e transformadas em um valor entre 0 e 1 pela *função de ativação* $G(\cdot)$. Finalmente, estes são multiplicados pelos pesos β_j , produzindo a *saída* y_t .

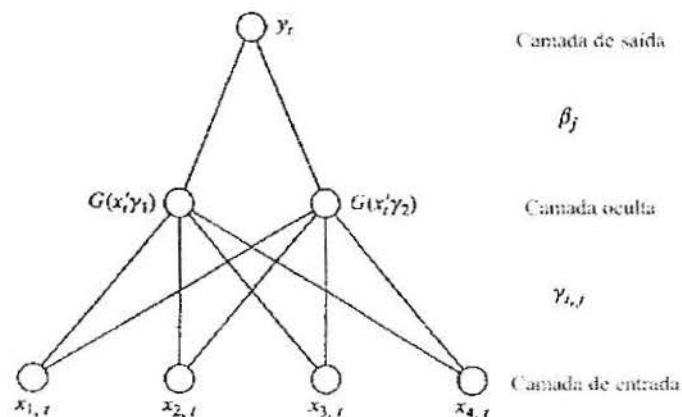


Figura 5: Representação do modelo de rede *single hidden layer feedforward* com $k=4$ e $q=2$

Fonte: Franses e van Dijk (2000)

Este tipo de rede neural artificial é comumente chamado de modelo *single hidden layer feedforward network*, pois contém apenas uma camada oculta e a informação flui em somente um sentido – da entrada para a saída. Extensões do modelo que permitem mais de uma camada oculta ou em que a informação faz o caminho inverso também são possíveis. Entretanto, o modelo de rede *single hidden layer feedforward* vem sendo o mais popular em séries temporais, e restringiremos atenção a esta particular RNA⁸.

2.5.2 Interpretação dos parâmetros e reconhecimento de padrões

Dado que o ajuste por uma rede neural artificial pode alcançar a unidade, em medidas como o R^2 , e dado que conjuntos de unidades da camada oculta podem ajustar um ruído branco, fica claro que os parâmetros individuais não podem ser interpretados. Por essa razão, não é possível inferir sobre o tipo de não-linearidade que é capturada pela RNA nos parâmetros individuais. Para explorar melhor as propriedades da parte não-linear do modelo, é recomendado investigar a contribuição conjunta dos componentes não-lineares, isto é,

$$\sum_{j=1}^q \beta_j G(x'_t \gamma_j), \quad (2.10)$$

ou então a contribuição individual de cada unidade da camada oculta

$$\beta_1 G(x'_t \gamma_1), \quad \beta_2 G(x'_t \gamma_2), \quad \dots, \quad \beta_q G(x'_t \gamma_q). \quad (2.11)$$

Na Seção 3.2, será ilustrada a flexibilidade das RNAs, demonstrando que elas são capazes de reconhecer anomalias na série, como *outliers* de diferentes tipos e mudanças de nível.

2.6 Estimação

Os parâmetros no modelo RNA(k, q)

$$y_t = x'_t \phi + \sum_{j=1}^q \beta_j G(x'_t \gamma_j) + a_t \quad (2.12)$$

podem ser estimados pela minimização da função soma dos quadrados dos resíduos

$$Q_n(\theta) = \sum_{t=1}^n [y_t - F(x_t; \theta)]^2, \quad (2.13)$$

onde

⁸ Para discussões mais aprofundadas, ver Kuan e White (1994) e Warner e Misra (1996).

$$F(x_t; \theta) = x_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j),$$

com θ sendo o vetor que consiste dos $k + 1 + q(k + 2)$ parâmetros em $\phi, \beta_1, K, \beta_q, \gamma_1, K, \gamma_q$. Devido ao modelo de redes neurais artificiais ser visto mais como um modelo de aproximação do que um processo subjacente que gera os dados, pode acabar sendo incrementalmente mal especificado. A teoria da estimação por mínimos quadrados em modelos mal especificados é atualmente bem desenvolvida e pode ser aplicada para obter-se as propriedades do estimador de mínimos quadrados não-linear de θ em (2.13), denotado $\hat{\theta}_n$.⁹ Sob condições gerais, $\hat{\theta}_n$ converge para θ^* , definido como $\theta^* = \arg \min_{\theta} E([y_t - F(x_t; \theta)]^2)$, na medida em que o tamanho da amostra n cresce. Além disso, o estimador normalizado $\sqrt{n}(\hat{\theta}_n - \theta^*)$ converge para uma distribuição normal multivariada com média zero e uma matriz de covariâncias que pode ser estimada com consistência.

Qualquer algoritmo de mínimos quadrados não-linear convencional pode ser aplicado para se obter a estimativa de $\hat{\theta}_n$. Dadas estimativas da r -ésima iteração, $\hat{\theta}_n^{(r)}$, pode-se computar a soma dos quadrados dos resíduos $Q_n(\hat{\theta}_n^{(r)})$ e o gradiente

$$\nabla Q_n(\hat{\theta}_n^{(r)}) = \frac{\partial Q_n(\hat{\theta}_n^{(r)})}{\partial \theta} \quad (2.14)$$

para obter a nova estimativa $\hat{\theta}_n^{(r+1)}$ como

$$\hat{\theta}_n^{(r+1)} = \hat{\theta}_n^{(r)} - \lambda A(\hat{\theta}_n^{(r)})^{-1} \nabla Q_n(\hat{\theta}_n^{(r)}), \quad (2.15)$$

onde λ é a duração de um passo e $A(\hat{\theta}_n^{(r)})$ uma matriz que é uma função dos parâmetros θ . Por exemplo, no algoritmo de Newton-Raphson, $A(\hat{\theta}_n^{(r)})$ é tida como a matriz Hessiana. Um método que também vem sendo popular na literatura das RNAs é o da *queda mais íngreme (steepest descent)*, que equivale a designar $A(\hat{\theta}_n^{(r)})$ igual a matriz identidade. Este é normalmente referido como o método de *backpropagation*, que provém do fato de que os diferentes elementos no vetor gradiente $\nabla Q_n(\hat{\theta}_n^{(r)})$ podem ser computados recursivamente. Por exemplo, a derivada parcial de $Q_n(\theta)$ em relação a β_j é dada por

⁹ Pötscher e Prucha (1997) para maiores detalhes.

$$\frac{\partial Q_n(\theta)}{\partial \beta_j} = -2 \sum_{i=1}^n [y_i - F(x_i; \theta)] G(x_i' \gamma_j), \quad (2.16)$$

enquanto a derivada parcial de $Q_n(\theta)$ em relação a $\gamma_{i,j}$ é igual a

$$\begin{aligned} \frac{\partial Q_n(\theta)}{\partial \gamma_{i,j}} &= -2 \sum_{i=1}^n [y_i - F(x_i; \theta)] \beta_j G(x_i' \gamma_j) [1 - G(x_i' \gamma_j)] x_{i,t} \\ &= -2 \sum_{i=1}^n \frac{\partial q_i(\theta)}{\partial \beta_j} \beta_j [1 - G(x_i' \gamma_j)] x_{i,t}, \end{aligned} \quad (2.17)$$

onde $q_i(\theta) = [y_i - F(x_i; \theta)]^2$ denota os resíduos ao quadrado da i -ésima observação. Dadas as estimativas de $\hat{\theta}_n^{(r)}$, pode-se calcular o valor das unidades ocultas $\hat{\beta}_j^{(r)} G(x_i' \hat{\gamma}_j^{(r)})$, o valor ajustado $F(x_i; \hat{\theta}_n^{(r)})$ e o resíduo $\hat{a}_i^{(r)} = y_i - F(x_i; \hat{\theta}_n^{(r)})$. Após, os resíduos $\hat{a}_i^{(r)}$ são usados para estimar o valor das derivadas parciais (2.16) sobre $\hat{\theta}_n^{(r)}$, que no caso são usadas para obter as derivadas parciais (2.17). Finalmente, as estimativas são atualizadas para $\hat{\theta}_n^{(r+1)} = \hat{\theta}_n^{(r)} - \lambda \nabla_{\theta} q_i(\hat{\theta}_n^{(r)})$. Conseqüentemente, cada iteração no algoritmo da *queda mais íngreme* consiste em um *loop*, primeiro avançando através da rede, e então “retropropagando” novamente.

Acima, as estimativas dos parâmetros são atualizadas usando a informação de todas n observações da amostra. Uma alternativa muito popular nas aplicações de RNAs é basear a atualização em apenas uma observação – isto é, utilizar *estimação recursiva*. Dados valores iniciais $\hat{\theta}_0$, o valor do gradiente para a primeira observação, $\nabla q_1(\hat{\theta}_0)$, é calculado e usado para obter um novo conjunto de estimativas $\hat{\theta}_1$, que são usadas para se obter o gradiente para a segunda observação, $\nabla q_2(\hat{\theta}_1)$, e assim por diante. Este processo é repetido até que os parâmetros estimados tenham convergido. Se a convergência ainda não tenha sido alcançada quando todas as observações já tiverem sido processadas, é possível continuar o processo começando com a primeira observação novamente. Assim como $\nabla q_i(\theta) = -2 \nabla F(x_i; \theta) [y_i - F(x_i; \theta)]$, a fórmula de atualização das estimativas dos parâmetros agora toma a forma

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \lambda \nabla F(x_{t+1}; \hat{\theta}_t) [y_{t+1} - F(x_{t+1}; \hat{\theta}_t)],$$

onde $\hat{\theta}_t$ representa as estimativas dos parâmetros baseadas nas t primeiras observações. Este método de estimação dos parâmetros recursivamente é referido como *aprendizado (learning)* e foi popularizado por Rumelhart, Hinton e Williams (1986).

2.6.1 Mínimos locais

A soma de quadrados $Q_n(\theta)$ em (2.13) é conhecida por possuir muitos mínimos locais. Assim, se o algoritmo de estimação converge, não está garantido que se obtém o mínimo *global*. Um método bastante aplicado para aumentar as chances de encontrar o mínimo global é estimar a RNA diversas vezes, usando diferentes valores iniciais $\hat{\theta}_n^{(0)}$, e escolher as estimativas que alcançaram um menor valor para $Q_n(\theta)$.

2.6.2 Transformação de dados

Uma outra forma capaz de melhorar as propriedades dos métodos de mínimos quadrados não-lineares numéricos é transformar as variáveis y_t e $x_{i,t}$, $i = 1, K, k$, de modo que todas estejam em uma mesma escala, permitindo comparação. É comum transformar as variáveis para o intervalo $[0,1]$ aplicando

$$z_t^* = \frac{z_t - \min(z_t)}{\max(z_t) - \min(z_t)}, \quad (2.18)$$

ou transformar as variáveis de modo que tenham média zero e variância unitária aplicando

$$z_t^* = \frac{z_t - \bar{z}_t}{\sigma(z_t)}, \quad (2.19)$$

onde \bar{z}_t e $\sigma(z_t)$ representam a média e o desvio padrão de z_t , respectivamente, e z_t é a saída y_t ou uma das entradas $x_{i,t}$.

2.6.3 Diminuição do peso (*weight decay*)

Por fim, a estimação dos parâmetros em RNAs pode ser benéfica por prevenir que as estimativas de parâmetros individuais se tornem indevidamente grande. Isto pode ser efetivamente alcançado alterando a função de soma de quadrados (2.13), de modo a incluir um termo de penalidade, ao qual costuma-se referir como '*diminuidor*' do peso. Isto é, a função objetiva a ser minimizada passa a ser

$$Q_n(\theta) = \sum_{t=1}^n [y_t - F(x_t; \theta)]^2 + r_\phi \sum_{i=0}^k \phi_i^2 + r_\beta \sum_{j=1}^q \beta_j^2 + r_\gamma \sum_{j=1}^q \sum_{i=0}^k \gamma_{ij}^2, \quad (2.20)$$

onde r_ϕ , r_β e r_γ devem ser especificados. Ainda, é necessário transformar as variáveis de entrada de modo que elas estejam em magnitude comparável, para assim a *diminuição de peso* ser validada.

2.7 Seleção do modelo

Considere novamente o modelo (2.9) para y_t , que contém $p + 1 + q(p + 2)$ parâmetros a serem estimados. Implementar um modelo RNA(p, q) requer algumas decisões a serem tomadas:

- escolher a função de ativação $G(\cdot)$
- escolher o número de unidades ocultas q
- escolher o número de *lags* p a serem usados como variáveis de entrada.

Normalmente, a escolha da função de ativação não é considerada um problema de decisão. Em quase todos os casos, usa-se a função logística. Todavia, outras escolhas, como a função tangente hiperbólica $G(z) = \tanh(z)$, por vezes também são utilizadas.

Existem diversas estratégias para a escolha do número de p e q em (2.9). Uma delas é estimar todos modelos de RNAs possíveis com $p \in \{1, 2, K, p^*\}$ e $q \in \{1, 2, K, q^*\}$, para valores pré-determinados de p^* e q^* , e usar critérios de seleção de modelo como o AIC

$$AIC = n \ln(\hat{\sigma}^2) + 2k, \quad (2.21)$$

com $k = p + 1 + q(p + 2)$, ou BIC

$$BIC = n \ln(\hat{\sigma}^2) + k \ln n, \quad (2.22)$$

onde $\hat{\sigma}^2$ representa a estimativa da variância dos resíduos. Nessa estratégia, escolhe-se o número de unidades ocultas e o número de *lags* de entrada simultaneamente.

Uma estratégia alternativa é decidir primeiramente somente o valor de p , especificando um modelo linear AR(p) para y_t , usando os critérios AIC, BIC ou algum outro adequado para definir a ordem apropriada. Em um segundo estágio, adiciona-se unidades ocultas ao modelo, mantendo o valor de p fixo. Novamente, AIC e BIC podem ser usados para decidir sobre o valor apropriado de q . Contudo, esta estratégia está sujeita a erros, uma vez que a ordem de p escolhida no primeiro estágio pode ser um valor muito pequeno ou muito grande.

Na implementação de modelos de RNAs que podem incluir variáveis exógenas entre as k entradas, decisões similares às citadas acima hão de ser tomadas. De fato, a situação é ainda mais complicada, pois se precisa não apenas decidir sobre k , o número de entradas a incluir, mas também sobre quais variáveis, devem ser incluídas. Isto pode ser visto como uma escolha de k entradas de um conjunto de m candidatas $x_{1,t}, K, x_{m,t}$. Se m é um valor muito grande, a primeira estratégia supracitada provavelmente tomaria muito

tempo, pois o número de modelos a serem estimados se tornaria igualmente grande. Parece ser necessário usar conhecimentos externos para decidir sobre as entradas apropriadas, ou então confiar na segunda estratégia e selecionar as entradas através de uma regressão linear entre y_t e as várias seleções das m candidatas.

Swanson e White (1995, 1997a, 1997b) adotam uma estratégia similar. Passo a passo, adiciona-se ao modelo entradas e unidades ocultas até que não haja mais melhoras no critério de seleção escolhido – o AIC, por exemplo. A cada passo, a entrada adicionada é aquela cuja melhora no critério escolhido é maior.

2.8 Avaliação do modelo

Considere novamente o modelo RNA(k, q) em (2.12). Não existem normas formais sobre como determinar a significância das variáveis de entrada $x_{i,t}$ individualmente. Assim como os parâmetros individuais nas redes neurais artificiais não têm um significado bem definido, as forças de conexão $\gamma_{i,j}$ também não podem ser usadas para este propósito. White (1989) e Kuan & White (1994) discutem uma estatística para testar a significância de todos os parâmetros relacionados a $x_{i,t}$ de forma conjunta e argumentam que isto pode ser útil para se testar a significância da variável de entrada. Se S_i é a matriz de seleção $(q+1) \times (k+1) + q(k+2)$ que seleciona ϕ_i e $\gamma_{i,j}$, $j=1, K, p$, do vetor de parâmetros completo θ , então a hipótese nula de que $x_{i,t}$ não é relevante pode ser expressa como $H_0 : S_i \theta = 0$. Dada uma estimativa $\hat{\theta}_n$ de θ , esta hipótese pode ser testada usando-se a estatística de Wald

$$W = n \hat{\theta}_n' S_i' (S_i \hat{C}_n S_i')^{-1} S_i \hat{\theta}_n \quad (2.23)$$

onde $\hat{C}_n = \hat{A}_n^{-1} \hat{B}_n \hat{A}_n^{-1}$ com

$$\hat{A}_n = \frac{1}{n} \sum_{t=1}^n \frac{\partial^2 Q_n(\hat{\theta}_n)}{\partial \theta \partial \theta'}$$

com $Q_n(\theta)$ sendo a função de soma de quadrados dada em (2.13), e

$$\hat{B}_n = \frac{1}{n} \sum_{t=1}^n \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta'}$$

Sob a hipótese nula, a estatística W de Wald tem distribuição assintótica qui-quadrado com $q+1$ graus de liberdade.

Outro método de se avaliar a influência da entrada $x_{i,t}$ na saída y_t é efetuar uma análise de sensibilidade. Nesta abordagem, as outras entradas $x_{1,t}, x_{2,t}, \dots, x_{i-1,t}, x_{i+1,t}, \dots, x_{k,t}$ são fixadas em, por exemplo, seu valor médio, enquanto a entrada de interesse $x_{i,t}$ é variada em uma amplitude de diferentes valores. A mudança na saída devido a mudanças em $x_{i,t}$ nos mostra, superficialmente, a importância desta variável de entrada.

2.9 Previsão

Como já foi mencionado, as redes neurais artificiais são consideradas modelos de ‘caixa preta’ e são construídas principalmente com o propósito de fazerem previsões. Fazer previsões com redes neurais é análogo a fazer previsões com outros modelos não-lineares. As principais características serão mostradas a seguir.

Primeiramente, uma previsão de y_{n+1} um passo a frente pode ser calculada diretamente de um modelo RNA(p,q) como

$$\hat{y}_{t+1|t} = x_t' \phi + \sum_{j=1}^q \beta_j G(x_t' \gamma_j), \quad (2.24)$$

onde $x_t = (1, y_t, K, y_{t-p+1})'$. Sob a suposição adicional de que os erros α_t em (2.9) são normalmente distribuídos, o erro de previsão um passo a frente $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$ é normalmente distribuído também (desde que, por definição, $e_{t+1|t} = \alpha_{t+1}$) e os intervalos de confiança podem ser construídos da maneira usual.

Para previsões diversos passos à frente, tudo se torna muito mais complicado. Não existem expressões prontas para $\hat{y}_{t+h|t}$, onde $h > 1$, e deve-se confiar em métodos de simulação para a obtenção das previsões. Para construção de intervalos de confiança para previsões diversos passos à frente e avaliação destas previsões por RNAs, usa-se métodos sugeridos por Franses e van Dijk (2000).

Aqui é necessário salientar novamente o perigo dos ajustes “em excesso” nas redes neurais artificiais. Aumentando o número de unidades ocultas q , a RNA pode se tornar bastante flexível, de modo que um ajuste quase perfeito possa ser obtido dentro da amostra. Porém, isso não implica necessariamente uma melhora na performance da previsão fora da amostra. Se o número de unidades ocultas for muito grande, algumas destas unidades poderão estar sendo usadas para modelar ruído e/ou *outliers*.

O equilíbrio entre o ajuste dentro da amostra e a performance da previsão é também influenciado pela quantidade de *diminuição de peso* utilizada. Quando o termo de penalidade na função objetiva é

diminuído, permite-se que os parâmetros individuais se tornem maiores, o que novamente faz com que algumas unidades ocultas ajustem *outliers* ou outros eventos atípicos. Por outro lado, reduzir a quantidade de *diminuição de peso* nos leva a um melhor ajuste dentro da amostra, mas a piores previsões fora da amostra.

É difícil alcançar um equilíbrio “ótimo” entre o ajuste dentro da amostra e a performance da previsão fora da amostra. Obviamente, o número de unidades ocultas q e os valores da *diminuição de peso* são importantes. Um método para se limitar o perigo do ajuste em excesso frequentemente aplicado é chamado de *validação cruzada*. Neste caso, as observações disponíveis são divididas entre amostra de estimação e amostra de teste. Os parâmetros da RNA são estimados usando somente as observações da amostra de estimação, mas é a soma dos erros quadráticos para as observações na amostra de teste que é usada, durante esta otimização iterativa. A estimação é parada quando esta soma de erros quadráticos começa a crescer. Assim, acredita-se que os parâmetros possam estar descrevendo os padrões não-lineares presentes nos dados, enquanto qualquer melhoria a mais seria entendida como o ajuste de *outliers* ou outros eventos atípicos¹⁰.

Uma outra possibilidade sensata para se proteger contra a previsão de *outliers* não existentes, ou de outros eventos que ocorrem uma única vez, é examinar as unidades ocultas para verificar se elas estão ativadas para apenas uma ou algumas poucas observações. Se estas unidades ocultas espúrias estiverem realmente presentes, omiti-las do modelo na etapa da previsão pode ser uma boa idéia. Isto é, suponhamos que foi estimado um modelo $\text{RNA}(k,q)$

$$y_t = x_t' \phi + \beta_1 G(x_t' \gamma_1) + \Lambda + \beta_q G(x_t' \gamma_q) + a_t,$$

onde $x_t = (1, y_{t-1}, K, y_{t-p}, z_{1t}, K, z_{mt})'$ e $k = p + m$, e que após uma inspeção individual das funções de ativação encontrou-se que somente o primeiro $q_1 < q$ está ajudando a descrever a relação não-linear na série de tempo, enquanto as demais unidades descrevem apenas eventos de ocorrência única. Pode-se, então, usar o modelo reduzido

$$y_t = x_t' \theta + \beta_1 G(x_t' \gamma_1) + \Lambda + \beta_{q_1} G(x_t' \gamma_{q_1}) + \sum_{j=q_1+1}^q \beta_j \bar{G}(x_t' \gamma_j) + \eta_t$$

para previsões fora da amostra, onde $\bar{G}(x_t' \gamma_j)$ é o valor médio da j -ésima função de ativação, excluindo a(s) observação(ões) para as quais está(ão) sendo ativada(s). Nota-se que este procedimento requer um julgamento subjetivo para decidir-se se certa unidade irá ou não contribuir para explicar padrões regulares nos dados ou apenas eventos esporádicos. Não está claro se regras ou normas objetivas podem ser

¹⁰ LeBaron e Weigend (1998) discutem esta abordagem e os efeitos das diferentes divisões dos dados em amostras de estimação e de teste.

desenvolvidas para formalizar este processo.

Para avaliar a performance de previsões através de redes neurais artificiais, existem ainda diversos outros critérios, como as estatísticas MSPE (*Mean Squared Prediction Errors*) e MAPE (*Mean Absolute Prediction Errors*), o teste de exatidão direcional de Pesaran e Timmermann (1992) ou a estatística de comparação de previsões de Diebold e Mariano (1995). Em aplicações financeiras de RNAs, é usual considerar outras medidas da performance. Por exemplo, as previsões freqüentemente são usadas em estratégias de troca envolvendo compra de ações e vendas de apólices. Neste caso, uma medida de avaliação da performance é comparar lucros de uma estratégia ativa com os de uma estratégia “*buy-and-hold*”¹¹.

3 IMPLEMENTAÇÃO NUMÉRICA: SIMULAÇÕES E ANÁLISE DE DADOS REAIS

3.1 Comparação entre modelos

Nesta seção, são apresentados os resultados obtidos após a comparação dos modelos ARIMA e de redes neurais artificiais com três séries de tempo diferentes. O primeiro conjunto de dados, relacionado com o estudo da meteorologia, é referente à temperatura média mensal na cidade de Recife nas décadas de 50 e 60. O segundo, relacionado à área médica, é referente à taxa de fertilidade na Suécia, medido anualmente entre 1750 e 1849. O terceiro, finalmente, é relacionado à área de finanças e mercado, e consiste em dados do retorno financeiro dos valores do índice SP500, medidos mensalmente entre a década de 20 e a década passada.

Para realizar esta comparação entre os dois diferentes modelos, foram feitos dois tipos de previsão. Nos dois casos, as últimas 12 observações foram deixadas em aberto, para posterior comparação. No primeiro modo de previsão, previu-se as 12 observações “excluídas” de uma vez só, enquanto no segundo modo previu-se “passo a passo”, isto é, após uma observação ser prevista, atualizava-se o banco para o passo seguinte.

3.1.1 Série de tempo nº1: meteorologia

Para explorar a área de estudos da meteorologia, foi escolhida uma série de tempo cujos dados, retirados de Chatfield (1989), representam a temperatura média (em °C) na cidade de Recife, Pernambuco, coletados mensalmente entre janeiro de 1953 e dezembro de 1962. O período de estimação da amostra usado

¹¹ Ver Gençay (1998) para exemplo.

foi entre janeiro de 1953 e dezembro de 1961, deixando as 12 últimas observações em aberto, a fim de serem usadas para posterior comparação com os dados reais. Conforme explicado anteriormente, foram gerados dois tipos de previsões: na primeira, previu-se para os doze anos seguintes de uma vez só; na segunda, efetuou-se a previsão passo a passo (o mesmo procedimento é adotado nas previsões por modelos ARIMA e por RNAs). O gráfico abaixo mostra o comportamento da série durante o período citado.

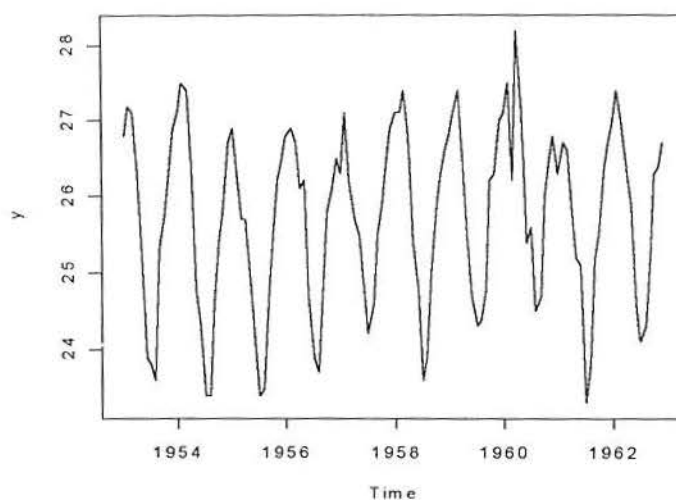


Figura 6: Temperatura média, em °C, na cidade de Recife

A primeira metodologia trabalhada foi a de Box-Jenkins, através dos modelos ARIMA.

Primeiramente, são observadas as funções de autocorrelação e autocorrelação parcial da série para ver que modelo deve ser sugerido.

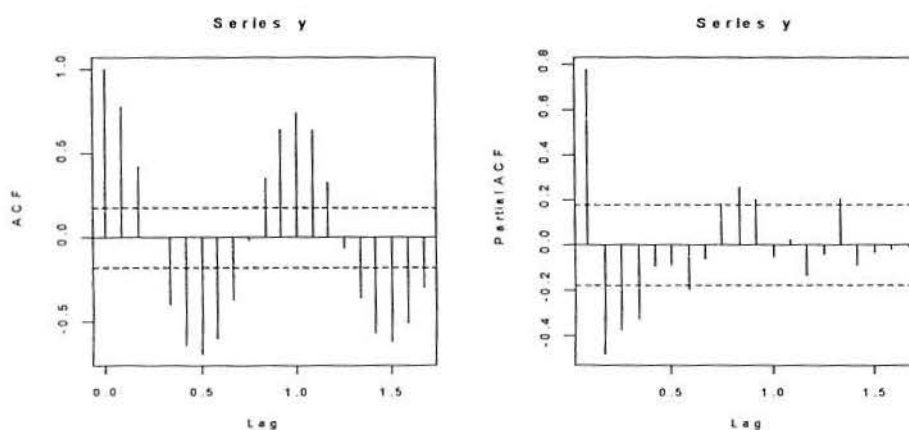


Figura 7: Funções de autocorrelação e de autocorrelação parcial da série nº1

Percebe-se claramente que há sazonalidade presente nos dados. Diversos modelos ARIMA foram testados. Dentre todos, o melhor modelo ARIMA encontrado, pelo critério AIC, é mostrado abaixo. O modelo estimado possui três parâmetros auto-regressivos e um sazonal de média móvel, além da primeira diferença sazonal, sob a notação SARIMA(3,0,0)x(0,1,1). Com os coeficientes estimados, o modelo pode ser escrito como

$$(1 - 0,3616B - 0,1068B^2 - 0,1914B^3)(1 - B^{12})Y_t = (1 + B^{12})a_t.$$

A Tabela 1 mostra o valor do critério AIC para alguns dos modelos sugeridos, justificando a escolha do modelo SARIMA(3,0,0)x(0,1,1), que possui um valor para AIC mais reduzido. Em anexo estão os gráficos de autocorrelação e autocorrelação parcial dos resíduos do modelo, comprovando ser este apropriado para modelar esta série.

Tabela 1: Critério AIC para modelos ARIMA sugeridos, série nº1

Modelo sugerido	AIC
ARIMA (0,0,0)x(0,1,1) ₁₂	176,41
ARIMA (1,0,0)x(0,1,1) ₁₂	150,03
ARIMA (0,0,1)x(0,1,1) ₁₂	158,13
ARIMA (2,0,0)x(0,1,1) ₁₂	148,40
<u>ARIMA (3,0,0)x(0,1,1)₁₂</u>	<u>146,39</u>
ARIMA (4,0,0)x(0,1,1) ₁₂	146,86
ARIMA (3,0,1)x(0,1,1) ₁₂	147,53
ARIMA (3,0,0)x(0,1,2) ₁₂	147,51

Tabela 2: Previsões pelo modelo SARIMA (3,0,0)x(0,1,1), série nº1

Período	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (k passos à frente, k=1..12)	Erro	Erro percentual
1962:1	27,0	26,736	-0,264	-0,98	26,736	-0,264	-0,98
1962:2	27,4	27,140	-0,260	-0,95	27,043	-0,357	-1,30
1962:3	27,0	26,887	-0,113	-0,42	26,727	-0,273	-1,01
1962:4	26,3	26,547	0,247	+0,94	26,364	0,064	+0,24
1962:5	25,9	25,609	-0,291	-1,12	25,541	-0,359	-1,39
1962:6	24,6	24,816	0,216	+0,88	24,645	0,045	+0,18
1962:7	24,1	23,952	-0,148	-0,61	23,943	-0,157	-0,65
1962:8	24,3	24,053	-0,247	-1,02	23,935	-0,365	-1,50
1962:9	25,2	25,097	-0,103	-0,41	24,955	-0,245	-0,97
1962:10	26,3	25,947	-0,353	-1,34	25,790	-0,510	-1,94
1962:11	26,4	26,553	0,153	+0,58	26,273	-0,127	-0,48
1962:12	26,7	26,889	0,189	+0,71	26,742	0,042	+0,16

Após esta modelagem, trabalhou-se com os modelos de redes neurais artificiais. Conforme já visto,

um dos principais problemas das redes neurais artificiais é determinar o número de unidades ocultas que deve ser usado na camada oculta. Assim, foram utilizados resultados do modelo ARIMA para termos uma idéia da arquitetura ideal da rede neural. Uma vez que os dados são sazonais e auto-regressivos de ordem três, foram usados 12 variáveis de entrada para representar os fatores sazonais e três para representar os períodos defasados da série (y_{t-1} , y_{t-2} e y_{t-3}). As “variáveis sazonais” consistem em uma sequência binária que atribui valor um para o mês em questão e zero para os demais. Isto é, o neurônio de entrada de janeiro é um vetor (1 0 0 0 0 0 0 0 0 0 0 0), o de fevereiro é um vetor (0 1 0 0 0 0 0 0 0 0 0 0), e assim por diante. Além disso, a rede possui três neurônios na camada oculta, definidos a partir do critério AIC. Usando a notação anteriormente mostrada, temos uma RNA (15,3,1) (15 variáveis de entrada, 3 unidades ocultas e uma variável de saída), ou simplesmente RNA (15,3).

Tabela 3: Previsões pelo modelo RNA (15,3), série nº1

Período	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (k passos à frente, $k=1..12$)	Erro	Erro percentual
1962:1	27,0	26,774	-0,226	-0,84	26,774	-0,226	-0,84
1962:2	27,4	27,079	-0,321	-1,17	27,083	-0,317	-1,16
1962:3	27,0	26,920	-0,080	-0,30	26,925	-0,075	-0,28
1962:4	26,3	26,612	0,312	+1,19	26,590	0,290	+1,10
1962:5	25,9	25,553	-0,347	-1,34	25,514	-0,386	-1,49
1962:6	24,6	24,886	0,286	+1,16	24,854	0,254	+1,03
1962:7	24,1	24,032	-0,068	-0,28	24,001	-0,099	-0,41
1962:8	24,3	24,023	-0,277	-1,14	24,032	-0,268	-1,10
1962:9	25,2	25,214	0,014	+0,06	25,229	0,029	+0,12
1962:10	26,3	25,895	-0,405	-1,54	25,908	-0,392	-1,49
1962:11	26,4	26,490	0,090	+0,34	26,482	0,082	+0,31
1962:12	26,7	26,717	0,017	+0,06	26,745	0,045	+0,17

Para ajudar na comparação dos dois métodos de previsão abordados, foi calculada a soma dos quadrado dos erros (SQE) para cada caso, mostrados na Tabela 4.

Tabela 4: Comparação entre métodos, série nº1

	ARIMA (passo a passo)	ARIMA (k passos à frente)	RNA (passo a passo)	RNA (k passos à frente)
SQE	0,62	0,90	0,71	0,69

Os resultados nos mostram que a previsão feita passo a passo pelo modelo ARIMA é a que nos dá melhores resultados. Contudo, ao comparar as duas formas de previsão dentro de cada modelo, obtém-se

resultados opostos: enquanto no modelo ARIMA a previsão passo a passo gera melhor resultados, no modelo de redes neurais é a previsão k passos à frente que fornece melhores resultados.

3.1.2 Série de tempo nº2: saúde

A segunda área de estudos explorada foi a da saúde. Foi escolhida uma série de tempo que representa a taxa de fertilidade na Suécia, medida anualmente entre 1750 e 1849. O período de estimação da amostra usado foi entre 1750 e 1837, deixando, novamente, as 12 últimas observações em aberto. O gráfico abaixo mostra o comportamento da série durante os 100 anos.

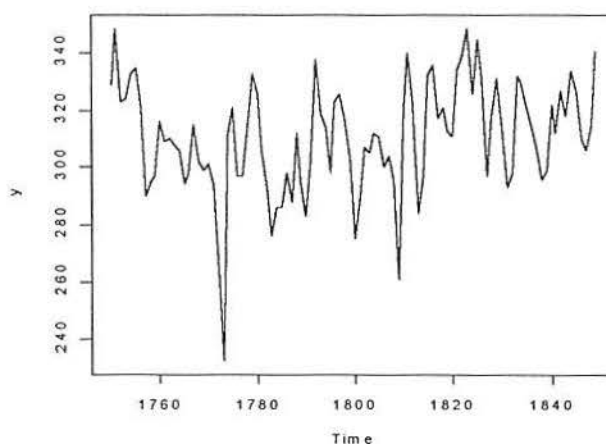


Figura 8: Taxa de fertilidade na Suécia, x1000

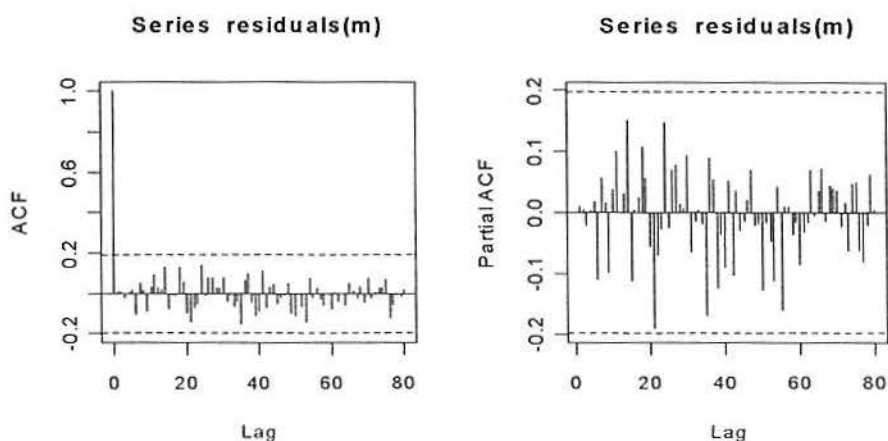


Figura 9: Funções de autocorrelação e de autocorrelação parcial da série nº2

Trabalhando com a metodologia de Box-Jenkins, escolheu-se o modelo ARIMA (4,0,0) como o mais

adequado, após a observação das funções de autocorrelação e autocorrelação parcial da série, onde foi identificada a estacionariedade da série, das funções de autocorrelação e autocorrelação parcial dos resíduos do modelo (ver anexo) e do cálculo do critério AIC para os diferentes modelos propostos.

Tabela 5: Critério AIC para modelos ARIMA sugeridos, série n°2

Modelo sugerido	AIC
ARIMA (1,0,0)	853,76
ARIMA (0,0,1)	848,86
ARIMA (1,0,1)	849,00
ARIMA (2,0,0)	850,20
ARIMA (3,0,0)	848,26
<u>ARIMA (4,0,0)</u>	<u>847,53</u>
ARIMA (5,0,0)	849,53
ARIMA (6,0,0)	851,41
ARIMA (0,0,2)	848,29
ARIMA (0,0,3)	849,70
ARIMA (1,0,2)	850,06
ARIMA (2,0,1)	850,63
ARIMA (3,0,1)	848,47
ARIMA (1,0,3)	851,40

O modelo ARIMA (4,0,0), com seus coeficientes estimados, pode ser escrito como

$$(1 - 0,6444B + 0,3008B^2 - 0,0875B^3 - 0,1673B^4)Y_t = a_t$$

Tabela 6: Previsões pelo modelo ARIMA (4,0,0), série n°2

Ano	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (k passos à frente, k=1..12)	Erro	Erro percentual
1838	296	310,92	14,92	+5,04	310,92	14,92	+5,04
1839	299	303,97	4,97	+1,66	314,04	15,04	+5,03
1840	322	307,54	-14,46	-4,49	312,79	-9,21	-2,86
1841	312	319,43	7,43	+2,38	310,22	-1,78	-0,57
1842	327	303,89	-23,11	-7,07	309,91	-17,09	-5,23
1843	318	320,21	2,21	+0,69	310,86	-7,14	-2,25
1844	334	311,91	-22,09	-6,61	311,09	-22,91	-6,86
1845	327	325,17	-1,83	-0,56	310,51	-16,49	-5,04
1846	310	317,67	7,67	+2,47	310,12	0,12	+0,04
1847	306	308,71	2,71	+0,89	310,22	4,22	+1,38
1848	315	312,92	-2,08	-0,66	310,38	-4,62	-1,47
1849	341	317,28	-23,72	-6,96	310,32	-30,68	-9,00

Modelando através de redes neurais artificiais, foram usadas quatro variáveis para representar os

períodos defasados da série (y_{t-1} , y_{t-2} , y_{t-3} e y_{t-4}). A rede também possuiu duas unidades na camada oculta. Logo, temos uma RNA (4,2).

Tabela 7: Previsões pelo modelo RNA (4,2), série n°2

Ano	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (k passos à frente, $k=1..12$)	Erro	Erro percentual
1838	296	310,53	14,53	+4,91	310,53	14,53	+4,91
1839	299	299,23	0,23	+0,08	300,19	1,19	+0,40
1840	322	303,78	-18,22	-5,66	305,29	-16,71	-5,19
1841	312	321,63	9,63	+3,09	321,16	9,16	+2,94
1842	327	301,03	-25,97	-7,94	301,83	-25,17	-7,70
1843	318	321,22	3,22	+1,01	321,31	3,31	+1,04
1844	334	313,41	-20,59	-6,16	313,19	-20,81	-6,23
1845	327	325,46	-1,54	-0,47	326,59	-0,41	-0,13
1846	310	320,73	10,73	+3,46	319,60	9,60	+3,10
1847	306	307,16	1,16	+0,38	306,93	0,93	+0,30
1848	315	312,47	-2,53	-0,80	312,85	-2,15	-0,68
1849	341	318,77	-22,23	-6,52	318,69	-22,31	-6,54

A tabela seguinte mostra a comparação entre as diferentes metodologias e formas de previsão, através da soma dos quadrados dos erros de cada modelo.

Tabela 8: Comparação entre métodos, série n°2

	ARIMA (passo a passo)	ARIMA (k passos à frente)	RNA (passo a passo)	RNA (k passos à frente)
SQE	2175,0	2657,1	2364,1	2248,8

Novamente, os resultados mostram que a previsão feita passo a passo pelo modelo ARIMA é a que nos dá melhores resultados. Assim como a análise da série da seção anterior, a comparação das duas formas de previsão dentro de cada modelo nos mostra resultados opostos: enquanto no modelo ARIMA a previsão passo a passo gera melhores resultados, no modelo de redes neurais é a previsão k passos à frente que fornece os melhores resultados. Poderíamos dizer que a ordem de “eficiência” dentre as quatro formas de prever foi a mesma na modelagem destes dois primeiros bancos de dados.

3.1.3 Série de tempo n°3: finanças

Por fim, a terceira área de estudos explorada foi a da economia. A série de tempo estudada foi o retorno dos valores do índice SP500, medidos mensalmente entre janeiro de 1926 e dezembro de 1991. O

período de estimação da amostra utilizado foi entre 1926 e 1990. O gráfico abaixo mostra o comportamento da série durante os 66 anos.

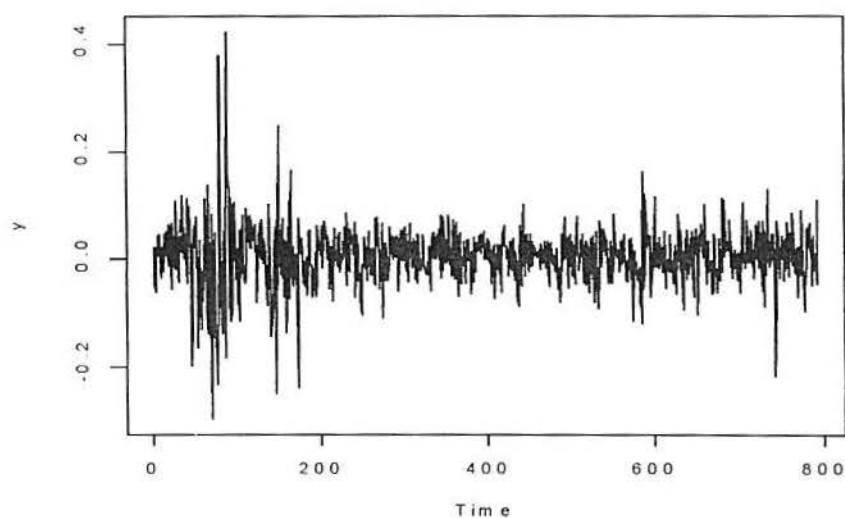


Figura 10: Retorno dos valores do índice SP500

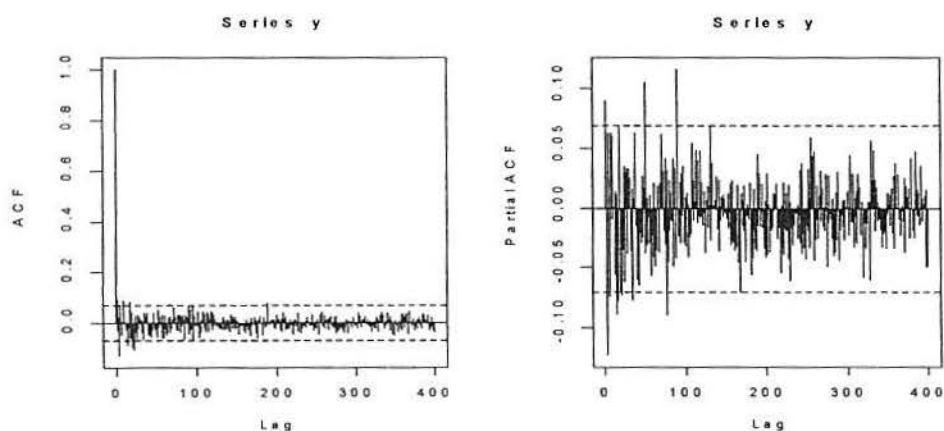


Figura 11: Funções de autocorrelação e de autocorrelação parcial da série nº3

Modelando com a metodologia de Box-Jenkins, escolheu-se o modelo ARIMA (3,0,0) como o mais adequado, após a observação das funções de autocorrelação e autocorrelação parcial da série, onde foi identificada a estacionariedade, das funções de autocorrelação e autocorrelação parcial dos resíduos do modelo (ver anexo) e, também, após o cálculo do critério AIC para os diferentes modelos propostos.

Tabela 9: Critério AIC para modelos ARIMA sugeridos, série n°3

Modelo sugerido	AIC
ARIMA (1,0,0)	-2251,5
ARIMA (0,0,1)	-2251,8
ARIMA (1,0,1)	-2249,8
ARIMA (2,0,0)	-2250,5
ARIMA (0,0,2)	-2249,8
<u>ARIMA (3,0,0)</u>	<u>-2260,5</u>
ARIMA (4,0,0)	-2259,9
ARIMA (3,0,1)	-2259,2
ARIMA (2,0,1)	-2251,4

O modelo ARIMA (3,0,0), com seus coeficientes estimados, pode ser escrito como

$$(1 - 0,089B + 0,0238B^2 - 0,1229B^3)Y_t = 0,0062 + \alpha_t$$

Tabela 10: Previsões pelo modelo ARIMA (3,0,0), série n°3

Período	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (k passos à frente, k=1..10)	Erro	Erro percentual
1991:1	0,0415	0,0078107	-0,0336893	-81%	0,0078107	-0,0336893	-81%
1991:2	0,0673	0,0020496	-0,0652504	-97%	-0,0010884	-0,0683884	-102%
1991:3	0,0222	0,0085253	-0,0136747	-62%	0,0028948	-0,0193052	-87%
1991:4	0,0003	0,0016739	0,013739	+458%	0,0056005	0,0053005	1767%
1991:5	0,0386	-0,0024032	-0,0410032	-106%	0,0068480	-0,0317520	-82%
1991:6	-0,0479	0,0073143	0,0552143	+115%	0,0063974	0,0542974	+113%
1991:7	0,0449	0,0009679	-0,0439321	-98%	0,0059878	-0,0389122	-87%
1991:8	0,0196	0,0069611	-0,0126389	-64%	0,0058073	-0,0137927	-70%
1991:9	-0,0191	0,0130083	0,0321083	+168%	0,0058574	0,0249574	+131%
1991:10	0,0119	-0,0012853	-0,0131853	-111%	0,0059175	-0,0059825	-50%
1991:11	-0,0439	0,0056070	0,0495070	+113%	0,0059441	0,0498441	+114%
1991:12	0,1116	0,0043994	-0,1072006	-96%	0,0059387	-0,1056613	-95%

Na modelagem através de redes neurais artificiais, foram usados três variáveis de entrada para representar os períodos defasados da série (y_{t-1} , y_{t-2} e y_{t-3}). A rede também possuiu duas unidades na camada oculta. Usando a notação definida anteriormente, tem-se uma RNA (3,2).

Tabela 11: Previsões pelo modelo RNA (3,2), série nº3

Período	Valor real	Valor previsto (passo a passo)	Erro	Erro percentual	Valor previsto (<i>k</i> passos à frente, <i>k</i> =1..10)	Erro	Erro percentual
1991:1	0,0415	0,0065663	-0,0349337	-84%	0,0065663	-0,0349337	-84%
1991:2	0,0673	0,0066169	-0,0606831	-90%	0,0065592	-0,0607408	-90%
1991:3	0,0222	0,0067032	-0,0154968	-70%	0,0065662	-0,0156338	-70%
1991:4	0,0003	0,0067187	0,0064187	+2140%	0,0065660	0,0062660	+2089%
1991:5	0,0386	0,0066238	-0,0319762	-83%	0,0065231	-0,0320769	-83%
1991:6	-0,0479	0,0068151	0,0547151	+114%	0,0065655	0,0544655	+114%
1991:7	0,0449	0,0067202	-0,0381798	-85%	0,0065661	-0,0383339	-85%
1991:8	0,0196	0,0066094	-0,0129906	-66%	0,0065323	-0,0130677	-67%
1991:9	-0,0191	0,0068938	0,0259938	+136%	0,0065664	0,0256664	+134%
1991:10	0,0119	0,0066914	-0,0052086	-44%	0,0065563	-0,0053437	-45%
1991:11	-0,0439	0,0068202	0,0507202	+116%	0,0065632	0,0504632	+115%
1991:12	0,1116	0,0069574	-0,1046426	-94%	0,0065662	-0,1050338	94%

A Tabela 12 mostra a comparação entre as diferentes metodologias e formas de previsão, através da soma dos quadrados dos erros de cada modelo.

Tabela 12: Comparação entre métodos, série nº3

	ARIMA (passo a passo)	ARIMA (<i>k</i> passos à frente)	RNA (passo a passo)	RNA (<i>k</i> passos à frente)
SQE	0,02755	0,02619	0,02505	0,02510

Neste banco de dados, pela primeira vez as redes neurais artificiais obtiveram resultados melhores nas duas formas de previsão do que os modelos ARIMA. Em parte, isto pode ser explicado pelo modelo ARIMA utilizado, que não se adequava totalmente aos dados. Devido às dificuldades em encontrar um modelo ARIMA razoável, optou-se por utilizar o modelo “mais próximo” encontrado, nesse caso o ARIMA (3,0,0). Por este exemplo, vê-se que as redes neurais têm mais facilidade em lidar com a não-linearidade, uma característica particular desta última série.

3.2 Identificação de outliers e quebras estruturais com RNAs

Foram feitos alguns experimentos com simulações para examinar o comportamento das redes neurais artificiais na presença de *outliers* ou mudanças de nível. Foi gerada uma série de tempo x_t a partir de um modelo AR(2)

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t, \quad t = 1, 2, \dots, n, \quad (3.1)$$

com $\phi_1 = 0,7$ e $\phi_2 = -0,2$. Os erros aleatórios a_t foram gerados de uma distribuição normal padrão e o tamanho da amostra foi $n = 500$. A série x_t foi “contaminada” com um *outlier* aditivo (AO), com um *outlier* inovativo (IO) e com mudança de nível (LS), cada caso ocorrendo no tempo $t = \tau$. Isto é, a série observada y_t foi obtida como:

$$\text{(AO): } y_t = x_t + \omega I[t = \tau], \quad (3.2)$$

$$\text{(IO): } y_t = x_t + \frac{\omega I[t = \tau]}{\phi(B)}, \quad (3.3)$$

$$\text{(LS): } y_t = x_t + \frac{\omega I[t = \tau]}{(1 - B)}, \quad (3.4)$$

onde $I[A]$ é a função indicadora do evento A e $\phi(B)$ é polinômio de defasagem, neste caso, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2$. Nos casos do *outlier* aditivo e do *outlier* inovativo, um único *outlier* de magnitude $\omega = 15$ ocorre no tempo $\tau = n/4$. No outro caso, ocorre uma mudança de nível permanente de mesma magnitude e no mesmo tempo.

Foram ajustados modelos de séries temporais por RNA como dado em (2.9), com $p = 2$ e $q = 1, 2$ e 3. Os *outliers* aditivo e inovativo, bem como a mudança de nível, ocorrem a $1/4$ da amostra, em $n = 125$. As saídas da camada oculta $\sum_{j=1}^q \hat{\beta}_j G(x'_i \hat{\gamma}_j)$ para RNAs com $p = 2$, para os casos AO, IO e LS, estão nas Figuras 12 a 23.

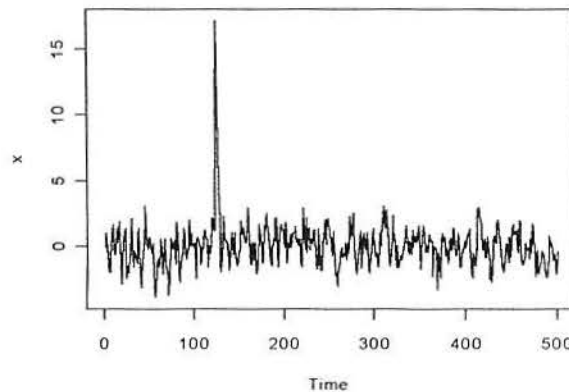


Figura 12: Série simulada “contaminada” com um *outlier* aditivo

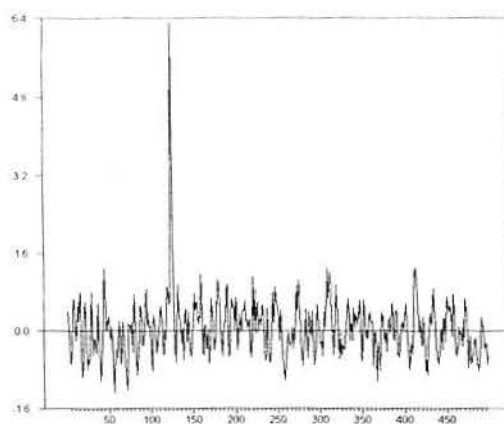


Figura 13: Saída da camada oculta para série simulada, com outlier aditivo, $p=2$ e $q=1$

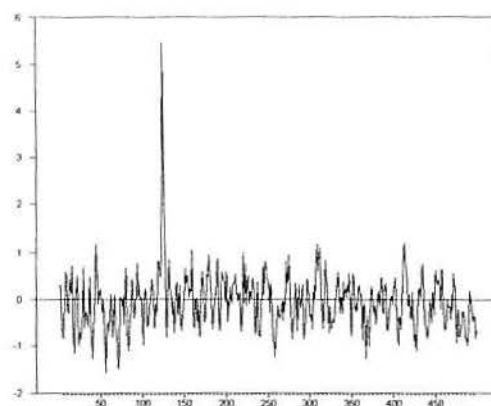


Figura 14: Saída da camada oculta para série simulada, com outlier aditivo, $p=2$ e $q=2$

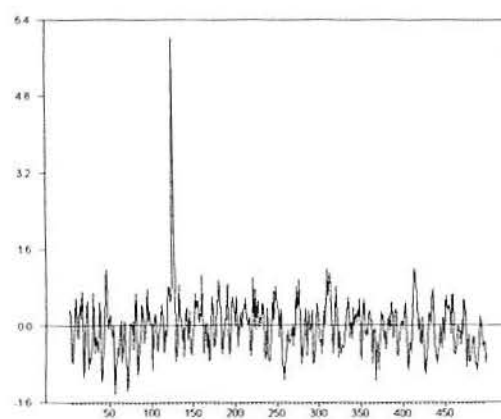


Figura 15: Saída da camada oculta para série simulada, com outlier aditivo, $p=2$ e $q=3$

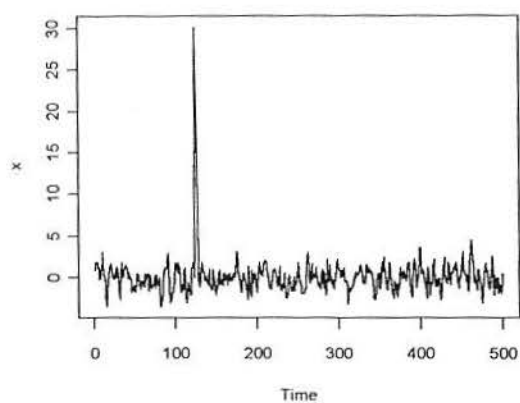


Figura 16: Série simulada “contaminada” com um *outlier* inovativo

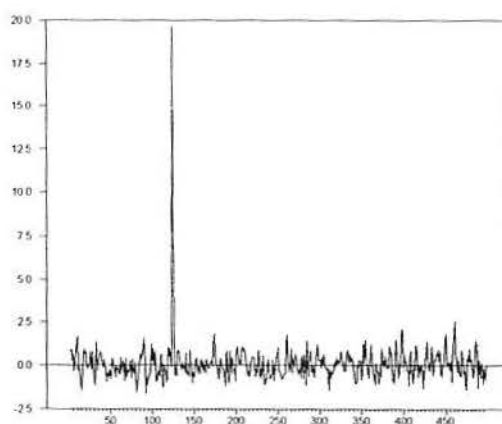


Figura 17: Saída da camada oculta para série simulada, com *outlier* inovativo, $p=2$ e $q=1$

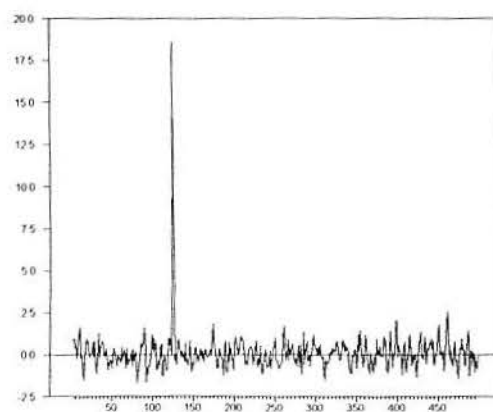


Figura 18: Saída da camada oculta para série simulada, com *outlier* inovativo, $p=2$ e $q=2$

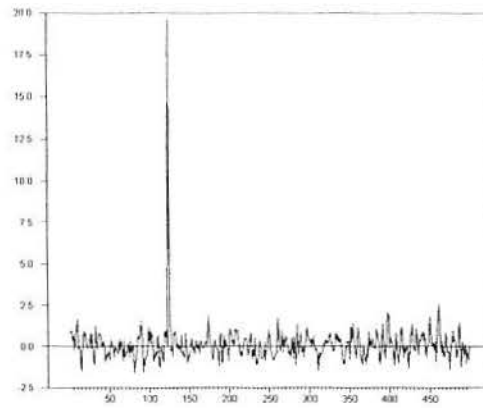


Figura 19: Saída da camada oculta para série simulada, com *outlier* inovativo, $p=2$ e $q=3$

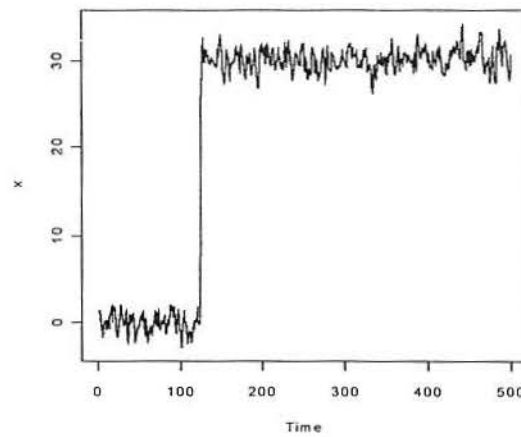


Figura 20: Série simulada “contaminada” com mudança de nível

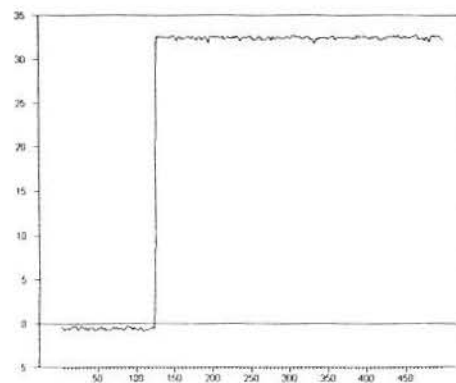


Figura 21: Saída da camada oculta para série simulada, com mudança de nível, $p=2$ e $q=1$

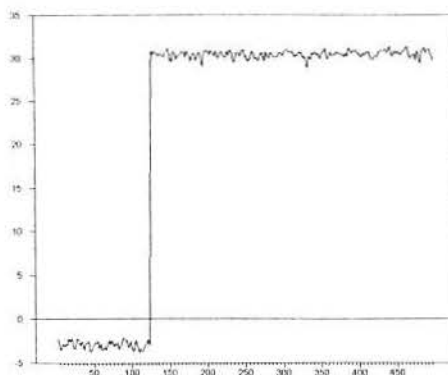


Figura 22: Saída da camada oculta para série simulada, com mudança de nível, $p=2$ e $q=2$

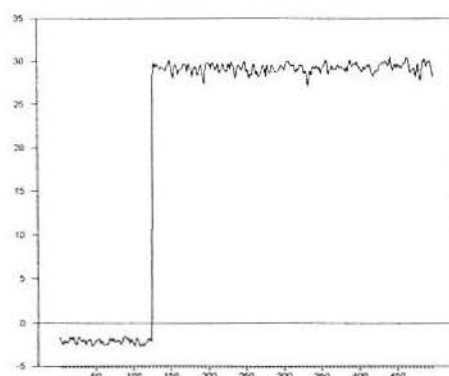


Figura 23: Saída da camada oculta para série simulada, com mudança de nível, $p=2$ e $q=3$

Pode-se ver claramente que, para as séries AO e IO, a saída da camada oculta é notadamente diferente quando o *outlier* ocorre. No experimento da mudança de nível, vemos na saída exatamente uma mudança de nível, que absorve parte da variação observada na série. Desta forma, concluímos que uma RNA é capaz de reconhecer e modelar eventos atípicos.

CONCLUSÃO

Este trabalho teve como objetivo comparar os modelos ARIMA e as redes neurais artificiais, procurando verificar qual desses modelos gera melhores resultados na previsão de séries temporais. As duas classes de modelos foram comparadas quanto à sua capacidade de prever as doze últimas observações das três séries temporais analisadas, todas contendo dados reais. Foram estimados modelos das duas classes para cada série, bem como diferentes formas de previsão para cada modelo: na primeira, previu-se “passo a passo”, com a atualização dos dados depois da previsão um passo à frente; na segunda, foi feita a previsão das doze observações em aberto de uma única vez.

Os resultados obtidos corroboram os estudos anteriores, mostrando que os modelos ARIMA, em grande parte dos casos, reproduzem razoavelmente as características das séries. Contudo, a classe das redes neurais artificiais reproduz de forma mais fidedigna estas características, principalmente na terceira série estudada, que é caracterizada por uma não-linearidade. Pode-se dizer que a capacidade em lidar com modelos não-lineares é o ponto forte das redes neurais artificiais.

Os principais problemas com as redes neurais artificiais parecem ser a falta de capacidade de explicação e de uma metodologia propriamente construída para a definição da arquitetura da rede. Além disso, uma vez que não há considerações estatísticas envolvidas na modelagem por RNAs, esta suporta somente previsões pontuais.

Em um tópico extra, foi dada atenção à habilidade das RNAs em modelar séries que contêm *outliers* e mudanças de nível. Através de dados simulados, obtiveram-se resultados similares a estudos anteriores, sugerindo que as RNAs são adequadas para lidar com eventos anômalos como *outliers* e mudanças de nível.

Em suma, apesar do bom desempenho dos modelos ARIMA, principalmente nas previsões feitas “passo a passo”, este trabalho mostra que a previsão de séries via redes neurais artificiais pode ser vantajosa, na medida em que importantes características das séries são descritas com mais exatidão por este tipo de modelo.

REFERÊNCIAS

- Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.
- Bosque, M. (2002) *Understanding 99% of Artificial Neural Networks: Introduction & Tricks*. Paperback.
- Box, G.E.P. e Jenkins, G.M. (1970) *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- Carpenter, G. e Grossberg, S. (1987) A massive parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, **37**, 54-115.
- Carrol, S.M. e Dickinson, B.W. (1989) Construction of neural nets using the random transform. *Proceedings of the IEEE Conference on Neural Networks (Washington D.C.)*, New York: IEEE Press, 607-11.
- Chatfield, C. (1984) *The Analysis of Time Series*, 3ª edição. Chapman & Hall: New York.
- Chatfield, C. (1989) *The Analysis of Time Series*, 4ª edição. Chapman & Hall: New York.
- Cybenko, G. (1989) Approximation by superpositions of a sigmoid function. *Mathematics of Control Signals and Systems*, **2**, 303-14.
- Dayhoff, J.E. (1998) *Neural Networks and Pattern Recognition*. Academic Press.
- Diebold, F.X. e Mariano, R.S. (1995) Comparing predictive accuracy. *Journal of Business & Economic Statistics*, **13**, 253-63.
- Donaldson, R.G. e Kamstra, M. (1996) Forecast combining with neural networks. *Journal of Forecasting*, **15**, 49-61.
- Fan, J. e Gijbels, I. (1996) *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.
- Fausett, L. (1994) *Fundamentals of Neural Networks: Architectures, Algorithms and Application*. New

Jersey: Prentice Hall, Englewood Cliffs.

Franses, P.H. e van Dijk, D. (2000) *Non-linear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.

Franses, P.H. e van Griensven, K. (1998) Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics and Econometrics*, **2**, 109-16.

Franses, P.H. e van Homelen, P. (1998) On forecasting exchange rates using neural networks. *Applied Financial Economics*, **8**, 589-96.

Funabashi, K. (1989) On the approximate realization of continuous mappings by neural networks. *Neural Networks*, **2**, 183-92.

Gençay, R. (1996) Non-linear prediction of security returns with moving average rules. *Journal of Forecasting*, **15**, 165-74.

Gençay, R. (1998) Optimization of technical trading rules and the profitability in security markets. *Economics Letters*, **59**, 249-54.

Gençay, R. (1999) Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics*, **47**, 91-107.

Gençay, R. e Stengos, T. (1998) Moving average rules, volume and the predictability of security returns with feedforward networks. *Journal of Forecasting*, **17**, 401-14.

Haefke, C. e Helmenstein, C. (1996a) Forecasting austrian IPOs: an application of linear and neural network error-correction models. *Journal of Forecasting*, **15**, 237-51.

Haefke, C. e Helmenstein, C. (1996b) Neural networks in the capital markets: an application to index forecasting. *Computational Economics*, **9**, 37-50.

Härdle, W., Lütkepohl, H. e Chen, W. (1997) A review of nonparametric time series analysis. *International Statistical Review*, **65**, 49-72.

Hebb, D.O. (1949) *The Organization of Behavior*. New York: John Wiley & Sons.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554-58.

- Hornik, K., Stinchcombe, M. e White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**, 359-66.
- Hornik, K., Stinchcombe, M. e White, H. (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, **3**, 551-60.
- Hutchinson, J.M., Lo, A.W. e Poggio, T. (1994) A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, **49**, 851-89.
- Jenkins, G.M. e Watts, D.G. (1968) *Spectral Analysis and its Applications*. San Francisco: Holden-Day.
- Kohonen, T. (1987) *Self-Organization and Associative Memory*. New York: Springer-Verlag.
- Kuan, C.-M. e Liu, T. (1995) Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*, **10**, 347-64.
- Kuan, C.-M. e White, H. (1994) Artificial neural networks: an econometric perspective. *Econometrics Reviews*, **13**, 1-143.
- Lavine, R.A. (1983) *Neurophysiology: The Fundamentals*. The Collamore Press.
- LeBaron, B. e Weigend, A.S. (1998) A bootstrap evaluation of the effect of data splitting on financial time series. *IEEE Transactions on Neural Networks*, **9**, 213-20.
- McCleary, R. e Hay, R. (1980) *Applied Time Series Analysis for the Social Sciences*. Beverly Hills: Sage.
- McCulloch, W.S. e Pitts, W. (1943) A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, **5**, 115-33.
- Minsky, M. e Papert, S. (1969) *Perceptrons: An Introduction to Computational Geometry*. Cambridge: MIT Press.
- Morettin, P.A. e Toloí, C.M.C. (1981) *Modelos para Previsão de Séries Temporais*. Rio de Janeiro: Impa.
- Nicholls, J.G. (2000) *From Neuron To Brain*. Sinauer Associates, Inc.
- Pena, D., Tiao, G. e Tsay, R. (2001) *A Course in Time Analysis*. New York: John Wiley.
- Pesaran, M.H. e Timmermann, A. (1992) A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, **10**, 461-5.

- Portugal, M. (1995) Neural networks versus time series methods: a forecasting exercise. *Revista Brasileira de Economia*, **49**, 611-29.
- Pötscher, B.M. e Prucha, I.V. (1997) *Dynamic Nonlinear Econometric Models – Asymptotic Theory*. Berlin: Springer-Verlag.
- Qi, M. e Maddala, G.S. (1995) Option pricing using ANN: the case of S&P 500 index call options. *Neural Networks in Financial Engineering: Proceedings of the 3rd International Conference on Neural Networks in the Capital Markets*, London, 78-91.
- Qi, M. e Maddala, G.S. (1999) Economic factors and the stock market: a new perspective. *Journal of Forecasting*, **18**, 151-66.
- Rosenblatt, R. (1959) *Principles of Neurodynamics*. New York: Spartan Books.
- Rumelhart, D., Hinton, G. e Williams, R. (1986) Learning internal representation by error propagation. *Rumelhart & McClelland: Parallel Distributed Processing – Explorations in the Microstructure of Cognition – Vol. 1: Foundations*. Cambridge: MIT Press.
- Swanson, N.R. e White, H. (1995) A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, **13**, 265-75.
- Swanson, N.R. e White, H. (1997a) Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, **13**, 439-61.
- Swanson, N.R. e White, H. (1997b) A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, **79**, 540-50.
- Warner, B. & Misra, M. (1996) Understanding neural networks as statistical tools. *American Statistician*, **50**, 284-93.
- White, H. (1989) Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association*, **84**, 1003-13.
- Ziegelmann, F.A. (2003) *NonParametric and Semi-Parametric Methods in Time Series: The Kernel Smoothing Approach*. ESTE 10.

ANEXOS

Tabela 13: média mensal da temperatura na cidade de Recife entre 1953 e 1962 (em °C)

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1953	26,8	27,2	27,1	26,3	25,4	23,9	23,8	23,6	25,3	25,8	26,4	26,9
1954	27,1	27,5	27,4	26,4	24,8	24,3	23,4	23,4	24,6	25,4	25,8	26,7
1955	26,9	26,3	25,7	25,7	24,8	24,0	23,4	23,5	24,8	25,6	26,2	26,5
1956	26,8	26,9	26,7	26,1	26,2	24,7	23,9	23,7	24,7	25,8	26,1	26,5
1957	26,3	27,1	26,2	25,7	25,5	24,9	24,2	24,6	25,5	25,9	26,4	26,9
1958	27,1	27,1	27,4	26,8	25,4	24,8	23,6	23,9	25,0	25,9	26,3	26,6
1959	26,8	27,1	27,4	26,4	25,5	24,7	24,3	24,4	24,8	26,2	26,3	27,0
1960	27,1	27,5	26,2	28,2	27,1	25,4	25,6	24,5	24,7	26,0	26,5	26,8
1961	26,3	26,7	26,6	25,8	25,2	25,1	23,3	23,8	25,2	25,5	26,4	26,7
1962	27,0	27,4	27,0	26,3	25,9	24,6	24,1	24,3	25,2	26,3	26,4	26,7

Fonte: Chatfield (1989)

Tabela 14: taxa de fertilidade na Suécia, medida anualmente (1750-1849), x1000

Ano	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9
175.	329	349	323	324	333	335	321	290	295	297
176.	316	309	310	308	306	294	298	315	302	299
177.	301	294	264	233	311	321	297	297	314	333
178.	326	306	293	276	286	286	298	288	312	295
179.	283	303	338	319	314	298	323	326	317	304
180.	275	290	307	305	312	311	300	304	296	261
181.	321	340	322	284	297	332	336	317	321	313
182.	311	334	339	349	326	345	331	297	318	331
183.	314	293	298	332	330	322	315	307	296	299
184.	322	312	327	318	334	327	310	306	315	341

Fonte: McCleary e Hay (1980) *apud* www-personal.buseco.monash.edu.au/~hyndman/TSDL/

Tabela 15: Retornos mensais dos valores do índice SP500

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1926	0,0225	-0,044	-0,0591	0,0227	0,0077	0,0432	0,0455	0,0171	0,0229	-0,0313	0,0223	0,0166
1927	-0,0208	0,0477	0,0065	0,0172	0,0522	-0,0094	0,065	0,0445	0,0432	-0,0531	0,0678	0,019
1928	-0,0051	-0,0176	0,1083	0,0324	0,0127	-0,0405	0,0125	0,0741	0,024	0,0145	0,1199	0,0029
1929	0,0571	-0,0058	-0,0023	0,0161	-0,0428	0,1124	0,0456	0,098	-0,0489	-0,1993	-0,1337	0,0253

1930	0,0625	0,0215	0,0799	-0,0095	-0,0165	-0,1646	0,0367	0,0075	-0,1301	-0,0888	-0,0218	-0,0742
1931	0,0489	0,1144	-0,0692	-0,0959	-0,1372	0,139	-0,0742	0,0095	-0,2994	0,0844	-0,0978	-0,1453
1932	-0,0283	0,0507	-0,1182	-0,2025	-0,2333	-0,0089	0,377	0,3754	-0,0369	-0,1386	-0,0589	0,0519
1933	0,0073	-0,1844	0,0336	0,4222	0,1587	0,1317	-0,088	0,1146	-0,1136	-0,0885	0,1027	0,0223
1934	0,1059	-0,0367	-0,0009	-0,027	-0,0813	0,0208	-0,1152	0,0541	-0,0055	-0,0319	0,0829	-0,0042
1935	-0,0421	-0,0396	-0,0309	0,0956	0,0323	0,0678	0,0831	0,0217	0,0239	0,0751	0,0393	0,0371
1936	0,0655	0,0168	0,0254	-0,0771	0,0458	0,0306	0,0681	0,0088	0,0013	0,075	0,0041	-0,0058
1937	0,0378	0,0146	-0,0094	-0,0831	-0,0103	-0,0529	0,1026	-0,0554	-0,1421	-0,1017	-0,1011	-0,0504
1938	0,0133	0,0608	-0,2504	0,1412	-0,0443	0,247	0,0727	-0,0274	0,0149	0,076	-0,0334	0,0377
1939	-0,0689	0,0325	-0,1354	-0,0055	0,0623	-0,0638	0,1087	-0,0714	0,1646	-0,0146	-0,0491	0,0238
1940	-0,0352	0,0066	0,0099	-0,0049	-0,2395	0,0766	0,0311	0,0262	0,0095	0,0394	-0,0424	-0,0028
1941	-0,0482	-0,0149	0,004	-0,0653	0,0043	0,0535	0,0548	-0,0087	-0,0097	-0,0686	-0,0421	-0,0451
1942	0,0138	-0,025	-0,0675	-0,0437	0,064	0,0184	0,0313	0,007	0,0267	0,0644	-0,0138	0,0517
1943	0,0716	0,0506	0,0527	0,0009	0,0449	0,0198	-0,0543	0,0103	0,0237	-0,0132	-0,0755	0,059
1944	0,0154	-0,0025	0,0169	-0,0125	0,0404	0,051	-0,0208	0,0087	-0,0031	0	0,0039	0,0351
1945	0,0143	0,0616	-0,0462	0,088	0,0115	-0,0033	-0,0201	0,058	0,0419	0,0303	0,0324	0,0099
1946	0,0697	-0,0695	0,0463	0,0376	0,0224	-0,0391	-0,0255	-0,0729	-0,1015	-0,008	-0,0115	0,0429
1947	0,0235	-0,0147	-0,0169	-0,0389	-0,0089	0,0526	0,0362	-0,0279	-0,0137	0,0212	-0,0285	0,0207
1948	-0,0399	-0,047	0,0771	0,0265	0,0782	0,003	-0,0532	0,0076	-0,0301	0,0678	-0,1082	0,0305
1949	0,0013	-0,0394	0,0301	-0,0212	-0,0373	-0,0021	0,0621	0,012	0,0237	0,0295	0,0012	0,0436
1950	0,0173	0,01	0,0041	0,0451	0,0393	-0,058	0,0085	0,0325	0,0559	0,0041	-0,001	0,0461
1951	0,0612	0,0065	-0,0183	0,0481	-0,0406	-0,026	0,0687	0,0393	-0,0009	-0,0138	-0,0026	0,0389
1952	0,0156	-0,0365	0,0477	-0,0431	0,0232	0,0461	0,0176	-0,0146	-0,0196	-0,0008	0,0465	0,0355
1953	-0,0072	-0,0182	-0,0236	-0,0265	-0,0032	-0,0163	0,0253	-0,0578	0,0013	0,051	0,009	0,002
1954	0,0512	0,0027	0,0302	0,049	0,0329	0,0007	0,0572	-0,034	0,0831	-0,0195	0,0808	0,0508
1955	0,0181	0,0035	-0,0049	0,0377	-0,0013	0,0823	0,0607	-0,0078	0,0113	-0,0305	0,0749	-0,0007
1956	-0,0365	0,0347	0,0693	-0,0021	-0,0657	0,0392	0,0515	-0,0381	-0,0455	0,0051	-0,011	0,0353
1957	-0,0418	-0,0326	0,0196	0,037	0,0369	-0,0013	0,0114	-0,0561	-0,0619	-0,0321	0,0161	-0,0415
1958	0,0428	-0,0206	0,0309	0,0318	0,015	0,0261	0,0431	0,0119	0,0484	0,0254	0,0224	0,052
1959	0,0038	-0,0002	0,0005	0,0388	0,0189	-0,0036	0,0349	-0,015	-0,0456	0,0113	0,0132	0,0276
1960	-0,0715	0,0092	-0,0139	-0,0175	0,0269	0,0195	-0,0248	0,0261	-0,0604	-0,0024	0,0403	0,0463
1961	0,0632	0,0269	0,0255	0,0038	0,0191	-0,0288	0,0328	0,0196	-0,0197	0,0283	0,0393	0,0032
1962	-0,0379	0,0163	-0,0059	-0,062	-0,086	-0,0818	0,0636	0,0153	-0,0482	0,0044	0,1016	0,0135
1963	0,0491	-0,0289	0,0355	0,0485	0,0143	-0,0202	-0,0035	0,0487	-0,011	0,0322	-0,0105	0,0244
1964	0,0269	0,0099	0,0152	0,0061	0,0115	0,0164	0,0182	-0,0162	0,0287	0,0081	-0,0052	0,0039
1965	0,0332	-0,0015	-0,0145	0,0342	-0,0077	-0,0486	0,0134	0,0225	0,032	0,0273	-0,0088	0,009
1966	0,0049	-0,0179	-0,0218	0,0205	-0,0541	-0,0161	-0,0135	-0,0778	-0,007	0,0475	0,0031	-0,0015
1967	0,0782	0,002	0,0394	0,0422	-0,0524	0,0175	0,0453	-0,0117	0,0328	-0,0291	0,0011	0,0263
1968	-0,0438	-0,0312	0,0094	0,0819	0,0112	0,0091	-0,0185	0,0115	0,0385	0,0072	0,048	-0,0416
1969	-0,0082	-0,0474	0,0344	0,0215	-0,0022	-0,0556	-0,0602	0,0401	-0,025	0,0442	-0,0353	-0,0187
1970	-0,0765	0,0527	0,0015	-0,0905	-0,061	-0,05	0,0733	0,0445	0,033	-0,0114	0,0474	0,0568
1971	0,0405	0,0091	0,0368	0,0363	-0,0416	0,0007	-0,0413	0,0361	-0,007	-0,0418	-0,0025	0,0862
1972	0,0181	0,0253	0,0059	0,0044	0,0173	-0,0218	0,0023	0,0345	-0,0049	0,0093	0,0456	0,0118
1973	-0,0171	-0,0375	-0,0014	-0,0408	-0,0189	-0,0066	0,038	-0,0367	0,0401	-0,0013	-0,1139	0,0166
1974	-0,01	-0,0036	-0,0233	-0,0391	-0,0336	-0,0147	-0,0778	-0,0903	-0,1193	0,163	-0,0532	-0,0202
1975	0,1228	0,0599	0,0217	0,0473	0,0441	0,0443	-0,0677	-0,0211	-0,0346	0,0616	0,0247	-0,0115
1976	0,1183	-0,0114	0,0307	-0,011	-0,0144	0,0409	-0,0081	-0,0051	0,0226	-0,0222	-0,0078	0,0525
1977	-0,0505	-0,0217	-0,014	0,0002	-0,0236	0,0454	-0,0162	-0,021	-0,0025	-0,0434	0,027	0,0028
1978	-0,0615	-0,0248	0,0249	0,0854	0,0042	-0,0176	0,0539	0,0259	-0,0073	-0,0916	0,0166	0,0149

1979	0,0397	-0,0365	0,0552	0,0017	-0,0263	0,0387	0,0087	0,0531	0	-0,0686	0,0426	0,0168
1980	0,0576	-0,0044	-0,1018	0,0411	0,0466	0,027	0,065	0,0058	0,0252	0,016	0,1024	-0,0339
1981	-0,0457	0,0133	0,036	-0,0235	-0,0017	-0,0104	-0,0022	-0,0621	-0,0538	0,0491	0,0366	-0,0301
1982	-0,0175	-0,0605	-0,0102	0,04	-0,0392	-0,0203	-0,023	0,116	0,0076	0,1104	0,0361	0,0152
1983	0,0331	0,019	0,0331	0,0749	-0,0123	0,0352	-0,033	0,0113	0,0102	-0,0152	0,0174	-0,0088
1984	-0,0092	-0,0389	0,0135	0,0055	-0,0594	0,0175	-0,0165	0,1063	-0,0035	-0,0001	-0,0151	0,0224
1985	0,0741	0,0087	-0,0029	-0,0046	0,0541	0,0121	-0,0048	-0,012	-0,0347	0,0425	0,0651	0,0451
1986	0,0024	0,0715	0,0528	-0,0141	0,0502	0,0141	-0,0587	0,0712	-0,0854	0,0547	0,0215	-0,0283
1987	0,1318	0,0369	0,0264	-0,0115	0,006	0,0479	0,0482	0,035	-0,0242	-0,2176	-0,0853	0,0729
1988	0,0404	0,0418	-0,0333	0,0094	0,0032	0,0433	-0,0054	-0,0386	0,0397	0,026	-0,0189	0,0147
1989	0,0711	-0,0289	0,0208	0,0501	0,0351	-0,0079	0,0884	0,0155	-0,0065	-0,0252	0,0165	0,0214
1990	-0,0688	0,0085	0,0243	-0,0269	0,092	-0,0089	-0,0052	-0,0943	-0,0512	-0,0067	0,0599	0,0248
1991	0,0415	0,0673	0,0222	0,0003	0,0386	-0,0479	0,0449	0,0196	-0,0191	0,0119	-0,0439	0,1116

Fonte: Pena, Tiao e Tsay (2001) *apud* www-personal.buseco.monash.edu.au/~hyndman/TSDL/

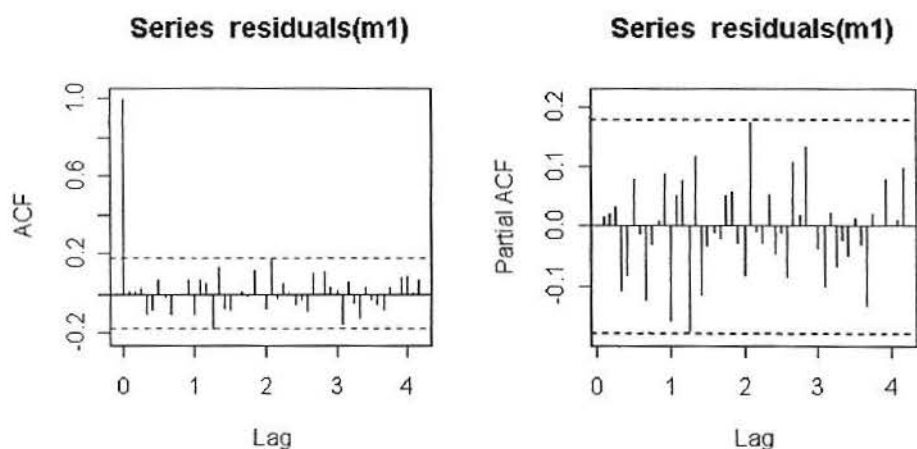


Figura 24: Resíduos do modelo SARIMA (3,0,0)x(0,1,1) aplicado na série n°1

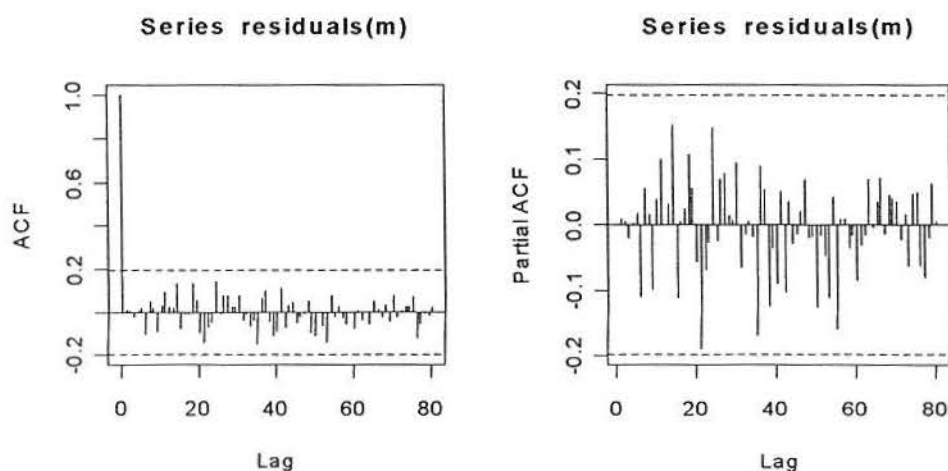


Figura 25: Resíduos do modelo ARIMA (4,0,0) aplicado na série n°2

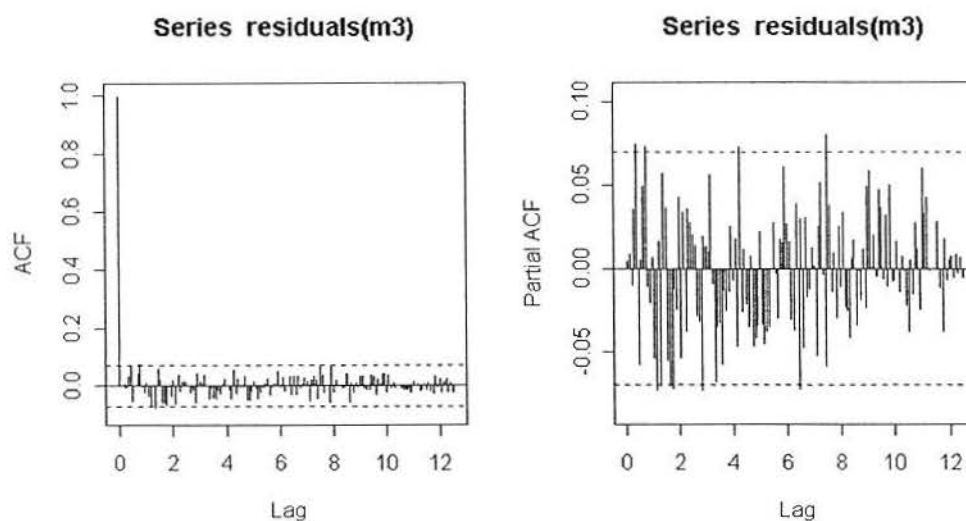


Figura 26: Resíduos do modelo ARIMA (3,0,0) aplicado na série nº3

Programa utilizado no *software* R para análise de modelos ARIMA, série de tempo nº 2.

```

y12<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem12).txt")
y12<-ts(y12,c(1750))
m12<-arima(y12,c(4,0,0))
M12<-predict(m12,n.ahead=12)
M12$pred
y11<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem11).txt")
y11<-ts(y11,c(1750))
m11<-arima(y11,c(4,0,0))
M11<-predict(m11,n.ahead=1)
M11$pred
y10<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem10).txt")
y10<-ts(y10,c(1750))
m10<-arima(y10,c(4,0,0))
M10<-predict(m10,n.ahead=1)
M10$pred
y9<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem9).txt")
y9<-ts(y9,c(1750))
m9<-arima(y9,c(4,0,0))
M9<-predict(m9,n.ahead=1)
M9$pred
y8<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem8).txt")
y8<-ts(y8,c(1750))
m8<-arima(y8,c(4,0,0))

```

```

M8<-predict(m8,n.ahead=1)
M8$pred
y7<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem7).txt")
y7<-ts(y7,c(1750))
m7<-arima(y7,c(4,0,0))
M7<-predict(m7,n.ahead=1)
M7$pred
y6<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem6).txt")
y6<-ts(y6,c(1750))
m6<-arima(y6,c(4,0,0))
M6<-predict(m6,n.ahead=1)
M6$pred
y5<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem5).txt")
y5<-ts(y5,c(1750))
m5<-arima(y5,c(4,0,0))
M5<-predict(m5,n.ahead=1)
M5$pred
y4<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem4).txt")
y4<-ts(y4,c(1750))
m4<-arima(y4,c(4,0,0))
M4<-predict(m4,n.ahead=1)
M4$pred
y3<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem3).txt")
y3<-ts(y3,c(1750))
m3<-arima(y3,c(4,0,0))
M3<-predict(m3,n.ahead=1)
M3$pred
y2<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem2).txt")
y2<-ts(y2,c(1750))
m2<-arima(y2,c(4,0,0))
M2<-predict(m2,n.ahead=1)
M2$pred
y1<-scan("C:/WINDOWS/Desktop/TCC/mccleary/mccleary15(sem1).txt")
y1<-ts(y1,c(1750))
m1<-arima(y1,c(4,0,0))
M1<-predict(m1,n.ahead=1)
M1$pred

```

Programa utilizado no *software* WinRats para análise de modelos de RNA, série de tempo nº 2.

```

all 96
data(format=rats) / yt4 yt3 yt2 yt1 yt
nnlearn(save=v,hidden=2,cvcrit=0.01) 1 84
#yt4 yt3 yt2 yt1

```

```

#yt
nntest 85 96 v
#yt4 yt3 yt2 yt1
#z
print / yt z
nnlearn(save=v1,hidden=2,cvcrit=0.01) 1 85
#yt4 yt3 yt2 yt1
#yt
nntest 86 86 v1
#yt4 yt3 yt2 yt1
#z1
print 86 86 z1
nnlearn(save=v2,hidden=2,cvcrit=0.01) 1 86
#yt4 yt3 yt2 yt1
#yt
nntest 87 87 v2
#yt4 yt3 yt2 yt1
#z2
print 87 87 z2
nnlearn(save=v3,hidden=2,cvcrit=0.01) 1 87
#yt4 yt3 yt2 yt1
#yt
nntest 88 88 v3
#yt4 yt3 yt2 yt1
#z3
print 88 88 z3
nnlearn(save=v4,hidden=2,cvcrit=0.01) 1 88
#yt4 yt3 yt2 yt1
#yt
nntest 89 89 v4
#yt4 yt3 yt2 yt1
#z4
print 89 89 z4
nnlearn(save=v5,hidden=2,cvcrit=0.01) 1 89
#yt4 yt3 yt2 yt1
#yt
nntest 90 90 v5
#yt4 yt3 yt2 yt1
#z5
print 90 90 z5
nnlearn(save=v6,hidden=2,cvcrit=0.01) 1 90
#yt4 yt3 yt2 yt1
#yt
nntest 91 91 v6

```

```
#yt4 yt3 yt2 yt1
#z6
print 91 91 z6
nnlearn(save=v7,hidden=2,cvcrit=0.01) 1 91
#yt4 yt3 yt2 yt1
#yt
nntest 92 92 v7
#yt4 yt3 yt2 yt1
#z7
print 92 92 z7
nnlearn(save=v8,hidden=2,cvcrit=0.01) 1 92
#yt4 yt3 yt2 yt1
#yt
nntest 93 93 v8
#yt4 yt3 yt2 yt1
#z8
print 93 93 z8
nnlearn(save=v9,hidden=2,cvcrit=0.01) 1 93
#yt4 yt3 yt2 yt1
#yt
nntest 94 94 v9
#yt4 yt3 yt2 yt1
#z9
print 94 94 z9
nnlearn(save=v10,hidden=2,cvcrit=0.01) 1 94
#yt4 yt3 yt2 yt1
#yt
nntest 95 95 v10
#yt4 yt3 yt2 yt1
#z10
print 95 95 z10
nnlearn(save=v11,hidden=2,cvcrit=0.01) 1 95
#yt4 yt3 yt2 yt1
#yt
nntest 96 96 v11
#yt4 yt3 yt2 yt1
#z11
print 96 96 z11
```