

Comparação de algumas técnicas de previsão em análise de séries temporais

RAYDONAL OSPINA MARTÍNEZ*
BARTOLOMEU ZAMPROGNO**

Resumo

Neste trabalho, nós comparamos o desempenho das metodologias de previsão de séries temporais Box & Jenkins, alisamento exponencial e redes neurais para diferentes horizontes de previsão previamente definidos. São analisadas duas séries de acordo com estas metodologias e suas previsões são comparadas sob diferentes medidas de ajuste.

Palabras clave: Modelos Box & Jenkins, alisamento exponencial, redes neurais, previsão.

Abstract

In this work, we compared the forecasting performance of three approaches: Box & Jenkins, exponential smoothing and neural networks, under different forecasting horizons. Two series are considered.

Key words: Box & Jenkins models, exponential smoothing, neural networks, forecast.

*Postgraduação em Estatística, Departamento de Estatística, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brasil.

**Postgraduação em Estatística, Departamento de Estatística, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brasil.

1. Introdução

Uma série temporal é um processo estocástico gerador de observações no tempo de uma determinada variável y a qual representa medições sucessivas de algum fenômeno de interesse. O principal objetivo do análise de séries temporais é investigar o mecanismo gerador dos dados, descrever seu comportamento através da construção de gráficos para verificação da existência de tendência, ciclos e variações sazonais, por exemplo. Por fim, objetiva-se fazer previsões de valores futuros que serviriam para tomar decisões. O processo de obter previsões é muito mais que a posição simplista de aventurar valores, pois, precisa da construção de um modelo adequado do sistema a ser tratado. Os métodos de previsão, em geral, baseiam-se na idéia de que observações passadas trazem informações (memória) sobre o padrão do comportamento da série temporal.

Assim, o problema principal é tratar de distinguir estes padrões de possíveis ruídos. Um procedimento de previsão que tenta tratar os dois aspectos é o alisamento exponencial, que é um método sem justificação probabilística, mas que tem grande popularidade devido à sua simplicidade, à eficiência computacional e a sua razoável precisão. O método Box & Jenkins é uma modelagem paramétrica que consiste em ajustar modelos auto-regressivos e/ou modelos de médias móveis, $ARIMA(p, d, q)$, a um conjunto de dados. Por sua vez, as redes neurais artificiais (RNAs) são sistemas processadores de informação que aceitam entradas, por exemplo, uma série temporal, que é processada gerando saídas (previsões). As redes neurais são compostas por unidades simples, chamadas neurônios, inter-conectadas entre si por conexões ponderadas.

Neste trabalho, apresentamos, na seção 2, os métodos de previsão utilizados no análise de duas séries temporais, na seção 3 se discutem os resultados sob as diferentes metodologias de previsão aplicadas a ditas séries mostrando que o método de alisamento exponencial apresenta um ótimo desempenho para previsões de curto prazo e que os modelos ARIMA e RNAs se comportam melhor nas previsões de longo prazo segundo os critérios de avaliação adotados.

2. Os métodos de previsão

O método de *Alisamento Exponencial Simples* é apropriado para séries que não apresentam tendência nem sazonalidade¹. O nível atual da série $\{N_t\}_{t \in T}$ é estimado através de uma média ponderada das observações anteriores, sendo

¹Os movimentos sistemáticos ao longo do tempo de aumento ou decréscimo de uma série temporal são chamados de tendência. Por sua vez, a sazonalidade indica flutuações periódicas que podem aparecer quando as observações são intra-aneais, isto é, registradas mensalmente,

que os pesos decrescem exponencialmente à medida que regredimos no tempo. A expressão do nível atual é dada por

$$N_t = (1 - \alpha)N_{t-1} + \alpha y_t, \quad t \in T, \quad (2.1)$$

onde,

$$N_{t-1} = \alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + \dots, \quad 0 < \alpha < 1.$$

O problema em questão é a determinação do valor de α . Uma forma razoável de escolhê-lo é através de inspeção visual, ou seja, se a série evolui de forma suave faz sentido usar um valor alto para α , ao passo que se a série evolui de forma errática faz sentido atribuir peso pequeno à última observação. Um método mais objetivo é escolher o valor de α que minimiza a soma dos quadrados dos erros de previsão um passo à frente, $S_\alpha = \sum_{t=3}^n e_t^2$, onde

$$e_t = y_t - N_{t-1} \text{ e } N_{t-1} = \hat{y}_{t-1}(1), \quad t = 3, 4, \dots, n \quad (2.2)$$

$\hat{y}_{t-1}(1)$ denotando a previsão de y_t feita no instante $t - 1$. Os algoritmos de alisamento exponencial podem ser vistos como um sistema de aprendizado, já que partir de (2.1) e (2.2) segue que $N_t = N_{t-1} + \alpha e_t$, $0 < \alpha < 1$, ou seja, a estimativa do nível num instante é a soma da estimativa anterior e de um múltiplo do erro de previsão. Dentre esta metodologia se encontra o *algoritmo de alisamento exponencial de Holt* o qual permite obter estimativas do nível e da tendência da série. A forma de recorrência é dada por

$$N_t = \alpha y_t + (1 - \alpha)(N_{t-1} + T_{t-1}), \quad 0 < \alpha < 1,$$

$$T_t = \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1}, \quad 0 < \beta < 1,$$

onde N_t e T_t são estimativas do nível e da tendência, respectivamente, no instante t e α e β são constantes de suavização. A previsão de y_{t+h} feita no instante t é

$$\hat{y}_t(h) = N_t + hT_t, \quad h = 1, 2, \dots$$

Novamente, a escolha objetiva dos valores de α e β se dá pela minimização da soma dos quadrados dos erros de previsão um passo à frente. Uma variante do algoritmo de Holt é o *algoritmo de alisamento exponencial de Holt-Winters* cujo objetivo principal é a incorporação de movimentos sazonais. Consideremos uma série sazonal com período s , por exemplo para dados mensais, $s = 12$. O fator sazonal (F_t) é multiplicativo, enquanto que a tendência é aditiva (série sazonal multiplicativa). A forma de recorrência do algoritmo é:

trimestralmente ou semanalmente, por exemplo.

$$\begin{aligned} N_t &= \alpha y_t / F_{t-s} + (1 - \alpha)(N_{t-1} + T_{t-1}), & 0 < \alpha < 1, \\ T_t &= \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1}, & 0 < \beta < 1, \\ F_t &= \gamma y_t / N_t + (1 - \gamma)F_{t-s}, & 0 < \gamma < 1. \end{aligned}$$

As previsões dos valores futuros da série são obtidas das seguintes expressões:

$$\begin{aligned} \hat{y}_t(h) &= (N_t + hT_t)F_{t+h-s}, & h = 1, 2, \dots, s, \\ \hat{y}_t(h) &= (N_t + hT_t)F_{t+h-2s}, & h = s + 1, s + 2, \dots, 2s, \\ & \vdots & \vdots \end{aligned}$$

O procedimento anterior pode ser modificado para tratar com situações onde o fator sazonal é aditivo (série sazonal aditiva). As equações de atualização dos dados são

$$\begin{aligned} N_t &= \alpha(y_t - F_{t-s}) + (1 - \alpha)(N_{t-1} + T_{t-1}), & 0 < \alpha < 1, \\ T_t &= \beta(N_t - N_{t-1}) + (1 - \beta)T_{t-1}, & 0 < \beta < 1, \\ F_t &= \gamma(y_t - N_t) + (1 - \gamma)F_{t-s}, & 0 < \gamma < 1. \end{aligned}$$

Os valores futuros são previstos a partir das equações a seguir:

$$\begin{aligned} \hat{y}_t(h) &= N_t + hT_t + F_{t+h-s}, & h = 1, 2, \dots, s, \\ \hat{y}_t(h) &= N_t + hT_t + F_{t+h-2s}, & h = s + 1, s + 2, \dots, 2s, \\ & \vdots & \vdots \end{aligned}$$

Uma metodologia de previsão que é classificada como clássica é a modelagem Box & Jenkins a qual baseia-se na construção de funções baseadas em um ciclo iterativo que utiliza os próprios dados da série para encontrar uma estrutura que permita fazer previsões. Inicialmente, é proposta uma classe de modelos que permite a identificação de um “bom” modelo com base em alguns critérios específicos, em seguida os parâmetros são estimados e, através da análise dos resíduos, o modelo ajustado é avaliado; caso não seja adequado, o ciclo é repetido. Muitas vezes é útil avaliar mais de um modelo, pois, por exemplo, se o objetivo é fazer previsões, o melhor modelo pode ser aquele que apresente o menor erro quadrático médio de previsão. Em geral, os modelos mais convenientes são os parcimoniosos, ou seja, aqueles que contêm poucos parâmetros e tendem a fornecer previsões mais precisas.

Dentre esta classe de modelos se encontra o *modelo auto-regressivo de ordem p*, denotado por $AR(p)$ o qual é definido como

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + u_t,$$

onde $u_t \sim RB(0, \sigma^2)$ é ruído branco², ϕ_1, \dots, ϕ_p são os parâmetros auto-regressivos e c é um parâmetro que permitir ao processo ter média diferente de zero. A notação mas utilizada é $\phi(B)y_t = c + u_t$, onde

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

Um outro modelo conhecido dentre esta metodologia é o denominado *processo de média móvel de ordem q*, denotado por $MA(q)$, o qual é da forma

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q},$$

com $u_t \sim RB(0, \sigma^2)$. A notação usual para descrever este processo é

$$y_t = \mu + \theta(B)u_t,$$

onde $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$. De forma mais geral definimos o *processo auto-regressivo de média móvel ARMA(p, q)* o qual tem a forma

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q},$$

onde $u_t \sim RB(0, \sigma^2)$, os ϕ 's e os θ 's são os parâmetros auto-regressivos e de médias móveis respectivamente, ou ainda $\phi(B)y_t = c + \theta(B)u_t$, onde $\phi(B)$ e $\theta(B)$ são os polinômios AR e MA usuais e as condições de estacionaridade³ em relação ao processo AR e de invertibilidade em relação ao MA permanecem. Se a série não for estacionária, o processo pode ser integrado de ordem d , ou seja, y_t é não estacionário, mas a diferencia⁴ de ordem d é estacionária. Tem-se então um processo $ARIMA(p, d, q)$ definido como

$$\phi(B)[(1 - B)^d y_t - \mu] = \theta(B)u_t.$$

Com o objetivo de levar em consideração movimentos sazonal, ampliou-se a classe de modelos $ARIMA$, pois em muitas ocasiões não é possível transformar y_t de forma a remover a sazonalidade, ou seja, a própria sazonalidade

²Uma seqüência $\{\varepsilon_t\}_{t \in T}$ de variáveis aleatórias é ruído branco quando estas variáveis são não correlacionadas com média zero e variância finita e constante.

³Neste contexto a série $\{y_t\}_{t \in T}$, é estacionária se: $E(y_t) = \mu$, $\text{Var}(y_t) = \sigma^2$, para todo t e $\text{Cov}(y_t, y_{t+h})$ depender apenas de h .

⁴Definimos a n -ésima diferença da série $\{y_t\}_{t \in T}$ como $\Delta^n y_t = \sum_{r=0}^n (-1)^r \binom{n}{r} y_{t-r}$, e o operador de defasagem $B y_t = y_{t-1}$; mais geralmente, $B^n y_t = y_{t-n}$, $n = 0, 1, 2, \dots$

pode apresentar um padrão dinâmico. Isto significa que é necessário considerar uma componente sazonal estocástica dentro do modelo com o fim de ajustar à série original. O novo modelo ARIMA é conhecido como ARIMA sazonal ou SARIMA. Seja y_t a série de interesse observada s períodos por ano e sejam

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{sP},$$

o operador auto-regressivo sazonal de ordem P , estacionário, e

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ},$$

o operador de médias móveis sazonal de ordem Q , invertível, $\Delta_s^D = (1 - B^s)^D$, onde D indica o número de “diferenças sazonais”. A classe de modelos sazonais multiplicativos $(p, d, q) \times (P, D, Q)$ é dada por

$$\begin{aligned} & (1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})[(1 - B)^d(1 - B^s)^D - \mu]y_t \\ & = (1 - \theta_1 B - \dots - \theta_p B^p)(1 - \Theta_1 B^s - \dots - \Theta_P B^{sP})u_t, \end{aligned}$$

onde $u_t \sim \text{RB}(0, \sigma^2)$, ou ainda,

$$\phi(B)\Phi(B^s)[(1 - B)^d(1 - B^s)^D - \mu] = \theta(B)\Theta(B^s)u_t.$$

Podemos supor que é possível combinar o modelo ARMA com uma estrutura de regressão da forma

$$\begin{aligned} y_t &= \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \beta_0 Z_t + \\ & \beta_1 Z_{t-1} + \dots + \beta_r Z_{t-r} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \end{aligned}$$

onde y_t é a variável dependente, Z_t é a variável independente, ε_t é um ruído branco e α , ϕ_i , θ_j são parâmetros desconhecidos. O modelo acima é conhecido como ARMAX e pode ser estendido para incluir mais de uma variável explicativa (exógena). Para mas detalhes sob a fundamentação teórica dos modelos anteriormente descritos, ver Box & Jenkins (1970).

Para nossa análise, precisamos de uma breve descrição do processo de identificação do modelo. Em modelos $\text{AR}(p)$ os gráficos das estimativas das autocorrelações parciais⁵ devem ter ordenada nula para ordem superior a p . Já em processos $\text{MA}(q)$ a ordenada nula deve corresponder a ordens superiores a q no gráfico das autocorrelações amostrais. Contudo, sendo o processo misto, a identificação por este procedimento torna-se muito difícil e artesanal. Neste caso,

⁵Entendemos a autocorrelação parcial de ordem k como a influência direta de y_{t-k} sobre y_t sem levar em consideração as observações entre elas.

existem dois critérios que permitem identificar o melhor modelo, no sentido de que maximizam a informação contida neles, a saber:

$$\begin{aligned} \text{AIC} &= 2 \log \widehat{L} + 2(p + q), \\ \text{BIC} &= -2 \log \widehat{L} + (p + q) \log T, \end{aligned}$$

onde T é o número de observações e \widehat{L} é a função de verossimilhança maximizada. O melhor modelo é aquele que apresenta o menor AIC e/ou o menor BIC. Como uma observação, o AIC (*Akaike Information Criterion*) não é consistente e o BIC (*Bayesian Information Criterion*) é consistente e mais parcimonioso. Assim, o BIC escolhe modelos cuja dimensão não ultrapassa a do modelo selecionado pelo AIC. Mills & Prasad (1992), usando simulação de Monte Carlo, concluíram que o BIC apresenta melhor desempenho. Neste trabalho será adotado o AIC como critério de escolha del melhor modelo.

Finalmente, o último passo consiste em determinar se o modelo estimado fornece uma representação adequada aos dados. Uma das formas de checagem é através de um análise das autocorrelações residuais. Segundo Box & Pierce (1970) e McLeod (1978) as distribuições assintóticas das autocorrelações amostrais dos resíduos de um modelo ARMA(p, q) é $\mathcal{N}(0, \frac{1}{T})$. Porém, é de se esperar que se o modelo estiver bem especificado as autocorrelações amostrais se encontrem dentro do intervalo $[-\frac{2}{\sqrt{T}}, \frac{2}{\sqrt{T}}]$ de confiança 95%. Um enfoque alternativo é testar conjuntamente a significância das primeiras m autocorrelações, o que pode ser feito usando um teste Portmanteau. Há duas estatísticas para este teste, a saber:

$$Q_{BP} = T \sum_{j=1}^m r_j^2,$$

proposta por Box & Pierce (1970) e

$$Q_{LB} = T(T + 2) \sum_{j=1}^m (T - j)^{-1} r_j^2,$$

proposta por Ljung & Box (1978), onde

$$r_j(\widehat{u}) = \frac{\sum_{t=j+1}^T \widehat{u}_t \widehat{u}_{t-j}}{\sum_{t=j}^T \widehat{u}_t^2}, \quad j = 1, 2, \dots,$$

e \widehat{u}_t são os resíduos do modelo escolhido. Temos que, sob a hipótese nula, $Q_i \xrightarrow{d} \chi_{m-p-q}^2$, $i = BP, LB$. Assim rejeita-se a hipótese de que o modelo está especificado corretamente se $Q_i > \chi_{\alpha, m-p-q}^2$, onde $\chi_{\alpha, m-p-q}^2$ é o valor crítico

de uma distribuição qui-quadrado para o nível de significância α e $(m - p - q)$ graus de liberdade. Se não for identificada nenhuma deficiência no modelo, então parte-se para a geração de previsões. Segundo as metodologias de previsão temos as redes neurais artificiais (RNAs) como modelos matemáticos inspirados na organização e no funcionamento dos neurônios biológicos. Existem numerosas variações das RNAs, as quais estão diretamente relacionadas com tarefas específicas. Uma RNA é basicamente um conjunto de elementos inter-conectados denominados neurônios, os quais estão distribuídos em três zonas: neurônios de entrada, neurônios intermediários e neurônios de saída. Os neurônios se inter-conectam e interagem conjuntamente com o fim de alcançar um comportamento desejado para um sistema. Cada neurônio recebe o estado de suas correspondentes entradas e, mediante uma função aplicada ao vetor de pesos, produz um sinal (saída), que é tomado pelo neurônio seguinte. As RNAs se adaptam a um ambiente de acordo com um treinamento o qual consiste em procurar a relevância das conexões que codificam o conhecimento que o sistema deve adquirir. Para determinar completamente dita estrutura as RNAs necessitam de uma arquitetura de inter-conexão, funções de ativação que medem a importância de um dado, uma função de custo que avaliam a saída da rede e um algoritmo de treinamento (aprendizagem) que modifica os parâmetros de conexão com o objetivo de otimizar a função de custo. Para mais detalhes ver Hertz, Krogh e Palmer (1991).

O problema de previsão das séries temporais pode ser tratado como a busca de uma função (sistema dinâmico) $\mathbf{F} : \mathbf{R}^d \rightarrow \mathbf{R}$ tal que $y_t = \mathbf{F}(y_{t-1}, \dots, y_{t-p}) + \varepsilon_t$, onde $\{\varepsilon_t\}_{t \in T}$ corresponde a um ruído. Um perceptrón é uma RNA composta por uma coleção de neurônios de entrada e um neurônio de saída. Assim, um perceptrón simples com p neurônios de entrada e função de ativação linear na capa de saída corresponde a um modelo $\text{AR}(p)$ com coeficientes estimados por mínimos quadrados. Nota-se que uma RNA compete com os modelos lineares para os quais existe uma fundamentação teórica sem questionamento; ver Gershenfeld & Weigend (1994). A metodologia de RNAs é mais eficiente na previsão de séries temporais não lineares, tratando o problema de previsão como um problema de aproximação. Tomando os valores pasados y_{t-1}, \dots, y_{t-p} encontrar o valor y_t corresponde a aproximar uma função \mathbf{F} que, em geral, tem solução. O perceptrón multi-capas é a RNA mais utilizada na previsão de séries temporais, a arquitetura típica corresponde a três capas: uma capa de entrada com p neurônios que contém os valores y_{t-1}, \dots, y_{t-p} , uma capa intermedia com k neurônios e uma capa de saída com um só neurônio que contém o valor y_t . As funções de ativação são as sigmóides; para mais detalhes ver Cheng & Titterington (1994).

Não existe uma fórmula geral para determinar o número de neurônios de

capa intermediária; se k é um número muito grande, existe o problema de que a rede não ajusta bem os dados, usualmente se determina esta capa de maneira “heurística” dependendo o problema a ser resolvido existem teoremas sob dimensão que podem fornecer idéias sobre o número adequado de neurônios a ser usado; para mas detalhes ver Pi & Peterson (1994). O treinamento de este tipo de rede neurais pode ser feito em ciclo aberto, i.e., não utiliza os valores de saída da rede como novas entradas para predizer y_t . No caso de se desejar prever vários valores y_t, \dots, y_{t+s} , então deve ser feito em ciclo fechado, o que implica usar como entradas o valor de saída da rede no instante anterior; para mas detalhes sobre este tipo de modelos ver Torres (1995). Na tabela 1 são apresentados os critérios de avaliação utilizados neste trabalho que permitiram determinar a qualidade das metodologias de previsão anteriormente mencionadas e que levam em consideração o grau de precisão das previsões consideradas.

Tabela 1: Critérios de avaliação sobre a qualidade das previsões.

Critério	Notação	Expressão
Erro quadrático médio	EQM	$\frac{1}{h} \sum_{j=1}^h e_{t+j}^2$
Erro absoluto médio	EAM	$\frac{1}{h} \sum_{j=1}^h e_{t+j} $
Erro médio	EM	$\frac{1}{h} \sum_{j=1}^h e_{t+j}$
Erro absoluto total	EAT	$\sum_{j=1}^h e_{t+j} $
Erro absoluto total	ET	$\sum_{j=1}^h e_{t+j}$

3. Modelagem e previsões

Neste trabalho foram analisadas duas séries temporais, com a finalidade de gerar previsões, sob as diferentes metodologias discutidas anteriormente. A primeira série é composta de 264 observações mensais do PIB (Produto Interno Bruto) do Brasil, que cobre o período de janeiro de 1980 a dezembro de 2001. A fonte dos dados é o instituto de Pesquisa Econômica Aplicada (IPEA)⁶. O horizonte de previsão é de 12 meses. O segundo conjunto de dados

⁶<http://www.ipeadata.gov.br>

corresponde à série de variações do IPC (Índice de preços ao consumidor) da Colômbia, a qual é tomada mensalmente para o período compreendido entre janeiro de 1992 e fevereiro de 1998, cobrindo um total de setenta e quatro (74) meses. A fonte dos dados é o Departamento Administrativo Nacional de Estatística da Colômbia (DANE)⁷. O horizonte de previsão é de sete meses: agosto de 1997 a fevereiro de 1998. Os gráficos das figuras 1 e 2 representam o comportamento destas duas séries. Para a análise das séries na parte de modelos lineares e alisamento exponencial será utilizado o pacote R⁸, no entanto para a previsão por redes neurais será utilizado o software *Trajan Neuronal Network Simulator*⁹.

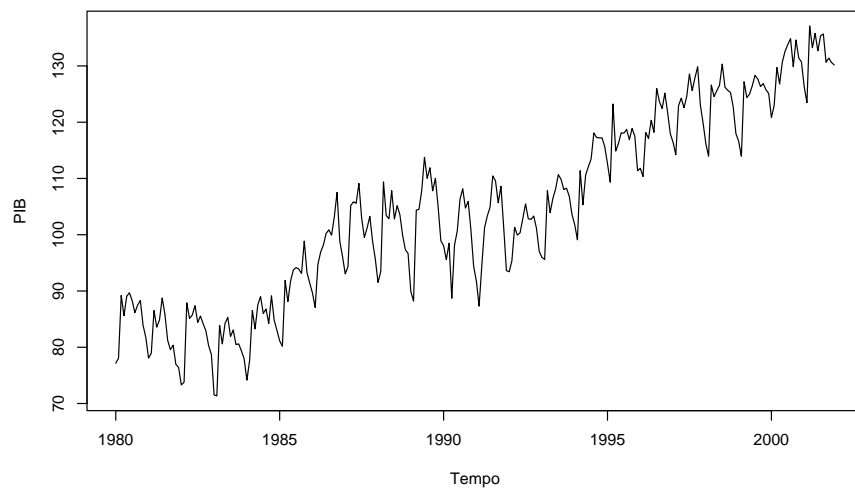


Figura 1: Série do PIB do Brasil.

Para a análise por redes neurais, o software *Trajan Neuronal Network Simulator* procura automaticamente a melhor especificação de rede neural usando análise de sensibilidade (Golberg, 1989), seleção de características (Hunter, et al., 2000) e regularização de Weigend (Weigend, Rumelhart & Huberman, 1991).

⁷<http://www.dane.gov.co>

⁸<http://www.r-project.org>

⁹Versão demonstração, <http://www.trajan-software.demon.co.uk>

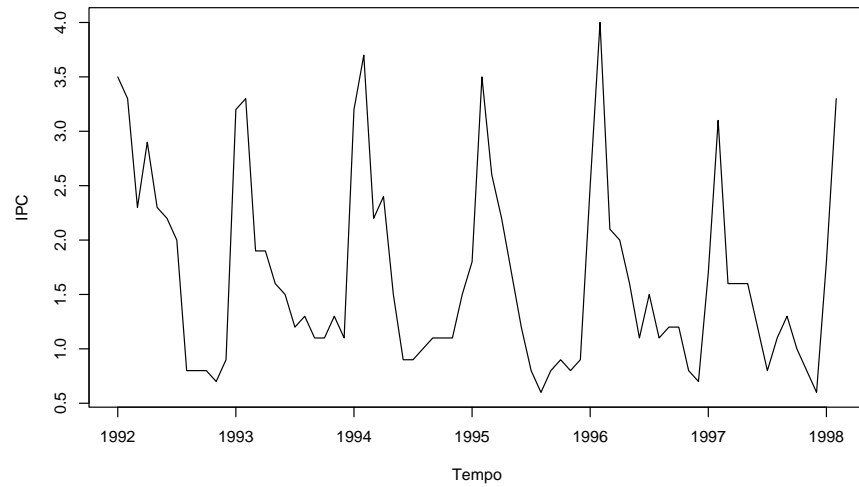


Figura 2: Série do IPC da Colômbia (variações).

Entre os tipos de RNAs usadas temos: Multilayer Perceptron (MLP) (Rumelhart, Hinton & Williams, 1986), Linear Network (Linear) (Golub & Kahan, 1965) e Radial Basis Function Network (RBFN) (Broomhead & Lowe, 1988). No software são apresentados os valores do número de neurônios de entrada (NE), número de neurônios intermediários (NI), a menor razão de regressão (RZ) e o menor erro de aprendizagem (ERA). A RZ e o ERA são duas medidas do ajuste da rede aos dados de entrada e servem para determinar a qualidade da rede a ser usada na previsão. Nas tabelas 2 e 3 se apresentam as melhores arquiteturas de previsão encontradas para as séries do PIB e IPC, respectivamente, ressaltando-se que a rede R_6 , nas duas séries, é aparentemente a que apresentou a melhor arquitetura de previsão, pois mostra os menores valores de RZ e ERA. Escolhidas estas arquiteturas foram feitas as previsões \hat{y}_{t+h} , sendo h o horizonte de previsão.

Tabela 2: Redes neurais ajustadas com os dados do PIB.

Rede	Tipo	NE	NI	RZ	ERA
R ₁	linear	3	0	0.2392	0.06552
R ₂	MLP	6	8	0.2351	0.06425
R ₃	MLP	6	8	0.2257	0.06095
R ₄	RBF	6	9	0.2609	0.01848
R ₅	RBF	6	18	0.2498	0.01645
R ₆	MLP	12	2	0.2214	0.05812

Tabela 3: Redes neurais ajustadas com os dados do IPC.

Rede	Tipo	NE	NI	RZ	ERA
R ₁	MLP	12	2	1.0445	0.1890
R ₂	RBF	6	7	0.7458	1.0414
R ₃	Linear	7	1	0.7523	0.2141
R ₄	Linear	6	1	0.7452	0.2136
R ₅	MLP	6	8	0.6682	0.1899
R ₆	MLP	8	8	0.6298	0.1795

As tabelas 4 e 5 mostram os resultados das previsões. Usando o EQM como uma medida da qualidade da previsão, na tabela 4 se observa, para os dados do PIB, que as redes R₄ e R₆ têm o melhor desempenho em previsão dois passos à frente. No entanto, a rede R₁ apresentou a melhor capacidade preditiva a longo prazo. De forma análoga, na tabela 5 se observa que para os dados do IPC a rede R₆ tem o melhor desempenho em previsão dois passos à frente e que a rede R₁ apresentou maior poder preditivo a longo prazo. Assim, neste trabalho será adotada a rede R₁, tanto para os dados do PIB, quanto para os do IPC, para serem comparadas suas previsões com as previsões obtidas a partir das outras metodologias. Na análise por alisamento exponencial e modelos Box & Jenkins foram inicialmente considerados os dados do PIB brasileiro.

Na figura 1, apresentamos o gráfico das observações e notamos um crescimento da série ao longo do tempo, caracterizando assim uma componente de tendência; adicionalmente, a série aparenta ter uma periodicidade marcante de 12 meses, indicando assim a presença de uma componente sazonal. A figura 3 apresenta o gráfico da função de autocorrelação amostral da série versus defasagens, as linhas tracejadas são limites assintóticos de significância ao nível de 95%, considerando-se estatisticamente nulas a autocorrelações entre elas. O mesmo pode ser dito sobre as linhas do gráfico da função de autocorrelação parcial, figura 4. Verificamos através da figura 3 um decaimento muito lento em direção a zero, indicando um comportamento não-estacionário da série PIB.

Tabela 4. Comparação das previsões da série por RNA versus a série original do PIB.

h	$\hat{y}_{t+h R_1}$	$\hat{y}_{t+h R_2}$	$\hat{y}_{t+h R_3}$	$\hat{y}_{t+h R_4}$	$\hat{y}_{t+h R_5}$	$\hat{y}_{t+h R_6}$	EQM R_1	EQM R_2	EQM R_3	EQM R_4	EQM R_5	EQM R_6
1	133.00	131.0	131.0	125.0	132.8	128.3	44.8900	22.09	22.09	1.69	42.25	4.00
2	132.00	130.8	130.0	125.0	132.6	125.7	118.8500	76.85	65.65	4.25	126.89	9.29
3	132.20	130.7	129.8	124.5	132.5	124.4	142.8600	117.81	118.94	163.01	148.05	170.58
4	132.00	130.7	129.6	124.0	132.4	127.8	144.3000	124.06	131.90	247.65	148.69	199.74
5	131.90	130.6	129.6	122.8	132.3	125.2	158.7400	150.07	169.11	414.06	160.25	309.99
6	131.08	130.2	128.5	122.5	131.3	128.3	161.0504	155.83	185.92	516.07	161.94	328.48
7	131.07	129.5	128.6	122.4	131.6	129.2	178.9433	189.47	230.81	682.48	175.63	365.69
8	131.07	129.2	128.4	122.3	131.4	130.2	199.4642	230.43	282.65	859.37	193.27	394.85
9	131.05	129.7	127.8	122.2	131.3	128.9	199.6667	231.24	290.49	929.93	193.76	397.74
10	131.04	128.1	127.7	122.1	130.7	130.1	199.7343	241.48	303.45	1014.57	194.12	399.18
11	131.02	128.0	127.6	121.0	130.3	128.7	199.9107	248.24	312.45	1106.73	194.21	402.79
12	131.01	128.3	127.5	121.1	130.1	128.7	200.7388	251.48	319.21	1187.73	194.21	404.75

Tabela 5. Comparação das previsões da série por RNA versus a série original do IPC.

h	$\hat{y}_{t+h R_1}$	$\hat{y}_{t+h R_2}$	$\hat{y}_{t+h R_3}$	$\hat{y}_{t+h R_4}$	$\hat{y}_{t+h R_5}$	$\hat{y}_{t+h R_6}$	EQM R_1	EQM R_2	EQM R_3	EQM R_4	EQM R_5	EQM R_6
1	1.60	0.65	1.38	1.62	1.38	0.90	0.250000	0.2025	0.0784	0.270400	0.0784	0.0400
2	1.01	1.24	1.55	1.22	1.42	1.08	0.334100	0.2061	0.1409	0.276800	0.0928	0.0884
3	1.10	1.70	1.90	1.53	1.83	1.57	0.344100	0.6961	0.9509	0.526800	0.7328	0.4133
4	1.02	1.98	1.92	1.78	2.28	1.84	0.384904	2.0885	2.2053	1.487200	2.9232	1.4949
5	1.51	2.60	1.95	1.97	2.48	2.22	1.194904	6.0885	4.0278	3.177200	6.1632	4.1193
6	2.80	2.67	1.85	1.92	2.11	2.21	2.194904	6.8454	4.0303	3.192576	6.2532	4.2793
7	2.13	1.80	1.72	1.75	1.55	1.87	3.634904	9.0954	6.5267	5.595076	9.3157	6.3242

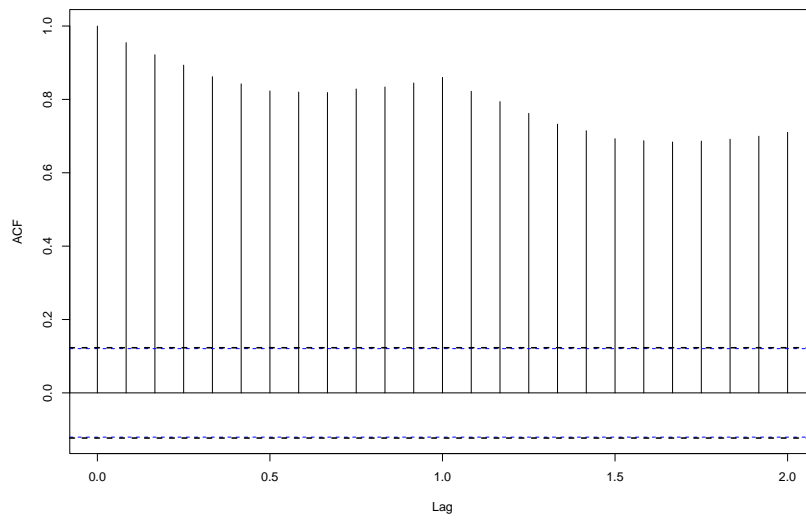


Figura 3: Correlograma da série PIB.

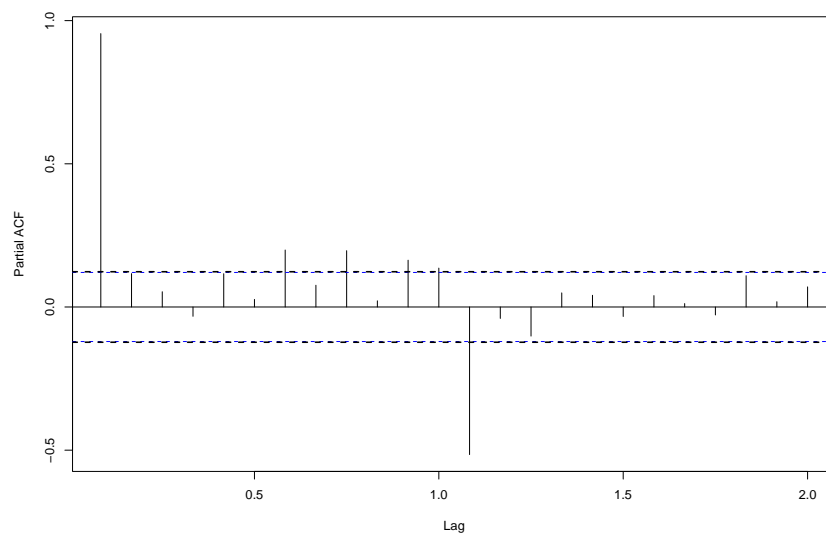


Figura 4: Correlograma parcial da série PIB.

Um procedimento utilizado para ser atingida a estacionariedade é realizar diferenças nas séries. É de se observar que nas figuras 5 e 6 estão apresentadas as funções de autocorrelação e autocorrelação parcial amostrais da série PIB, com diferenciação de ordem 1. Além disso, a análise do figura 4 evidencia a possibilidade da existência de uma componente sazonal. Estas percepções são consideradas para a escolha dos modelos mais adequados. A decomposição STL (Seasonal Decomposition of Time Series) decompõe uma série temporal em suas componentes sazonal, de tendência e irregular, utilizando um ajuste de regressão polinomial local. Uma referência sobre STL é Cleveland, et al. (1990). As figuras 7 e 8 apresentam a decomposição STL na série PIB e em sua diferenciação de ordem 1, respectivamente.

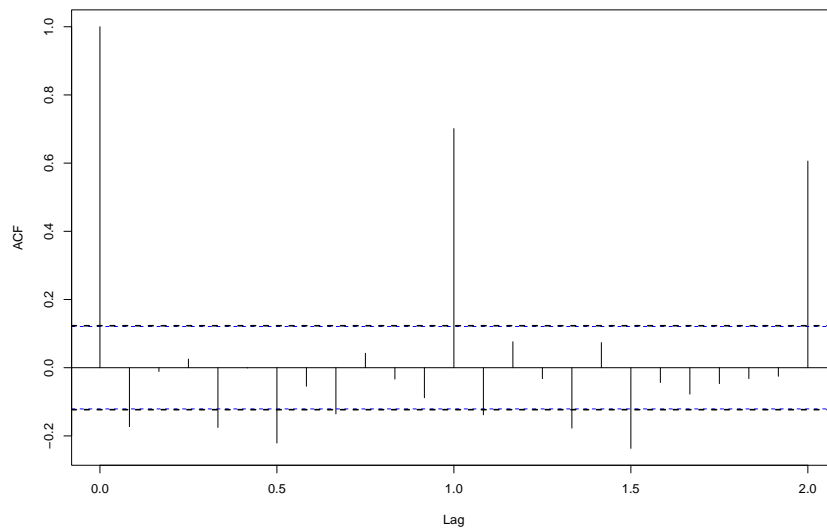


Figura 5: Correlograma da série PIB com diferenciação de ordem 1.

Em cada um dos gráficos o painel superior apresenta o comportamento temporal da série; o segundo painel fornece a estimativa da componente sazonal PIB, sem deixar de lado a as características cabíveis a cada figura. O terceiro painel, por sua vez, contém o gráfico da tendência. O último painel apresenta a estimativa da componente irregular. Observase que o comentado sobre a figura 2 se confirma a partir da análise da figura 7, ou seja, existe uma componente de tendência, indicando um comportamento do crescimento do PIB, e uma componente sazonal, que se torna mais evidente nesta decomposição.

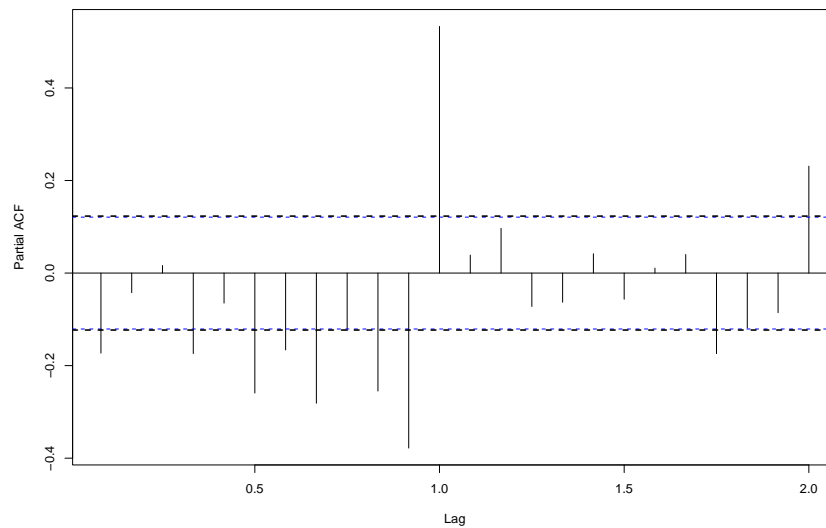


Figura 6: Correlograma parcial da série PIB com diferenciação de ordem 1.

A análise da figura 8 indica que, mesmo após diferenciada com ordem 1, a componente sazonal ainda “perturba” a série. De acordo com as características que a série apresenta aplicamos o algoritmo de Holt-Winters, que leva em consideração as componentes de tendência e sazonalidade. Consideramos sazonalidade aditiva e multiplicativa e quanto às constantes de suavização, estas foram escolhidas de forma automática minimizando-se a soma dos erros quadrados de previsão um passo à frente.

A tabela 6 apresenta os resultados dos critérios de seleção. O critério de seleção utilizado para os modelos de Box-Jenkins foi o AIC e em contrapartida apresentamos o BIC dos modelos selecionados por AIC apenas como informação adicional. Observando os gráficos da função de autocorrelação amostral e da função de autocorrelação parcial amostral, figuras 3 e 4, percebe-se que o decaimento da função de autocorrelação é lento, o que indica não-estacionariedade da série, e que a sazonalidade está presente. Diante disto, checamos os modelos variando as partes auto-regressivas e médias móveis variando a ordem dos parâmetros até grau 3 em suas componentes. Adicionalmente, utilizamos diferenciações $d = 0, 1, 2$ e $D = 0, 1, 2$, no modelo SARIMA.

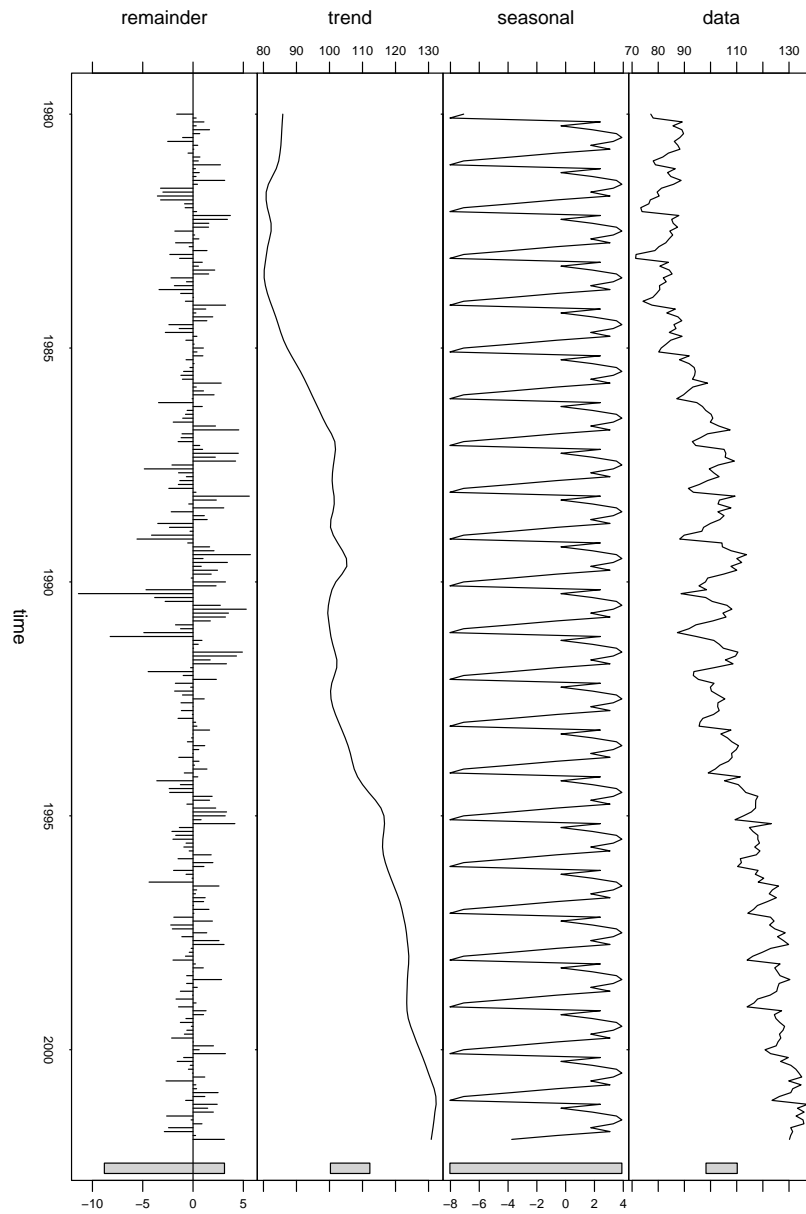


Figura 7: Decomposição STL da série PIB entre janeiro de 1980 a dezembro de 2001.

Os melhores resultados são apresentados na tabela 6. Para testar se o modelo estiver bem especificado foi investigado a correlação serial dos resíduos e feitos testes de Portmanteau dos modelos da tabela 6 e se concluiu que todos os modelos considerados eram adequados. Além dos modelos SARIMA também consideramos modelos SARIMAX, i.e., o modelo ARMAX com sazonalidade. Assim, investigamos a inclusão de algumas variáveis exógenas tais como: resíduo, resíduo elevado ao quadrado, tendência, sazonalidade e combinações entre tendência e sazonalidade, como também tendência e resíduo. Entre as variáveis exógenas avaliadas, a que apresentou melhores resultados foi a multiplicação entre a tendência e a sazonalidade da série.

Tabela 4: Critérios AIC e BIC aplicados nos modelos SARIMA e SARIMAX para ajuste na série PIB.

$(p,d,q)(P,D,Q)$	AIC	BIC
SARIMA(3,1,3) \times (1,0,1)	1181.23	1212.99
SARIMA(3,1,3) \times (2,0,1)	1181.99	1217.29
SARIMA(3,1,3) \times (0,1,1)	1115.47	1143.70
SARIMA(3,1,3) \times (0,1,2)	1116.29	1148.05
SARIMA(3,1,3) \times (1,1,1)	1116.16	1147.93
SARIMA(3,2,3) \times (1,1,1)	1131.99	1163.76
SARIMAX _{ts} (3,1,3) \times (2,0,1)	1161.65	1204.00

Para ajustarmos o modelo SARIMAX utilizamos a ordem dos polinômios do modelo SARIMA(3,1,3) \times (2,0,1) devido a este modelo apresentar, sobre nossas considerações, melhores resultados de previsão. Esta última combinação está representada na última linha da tabela 6 e também na última da tabela 7 b e denotamos por SARIMAX_{ts}. Apresentamos na tabela 7 medidas de erro de previsão para todos os modelos comentados até o momento. Observamos que de acordo com todas as medidas consideradas o modelo de Holt-Winters com sazonalidade aditiva se comporta melhor para $h = 1$ e $h = 3$, ou seja, um e três passos à frente, entretanto perde poder de previsão nos demais horizontes considerados. Dentre a classe de modelos SARIMA, o modelo (3,1,3) \times (2,0,1) apresentou em geral os melhores resultados.

No entanto, na maioria das medidas consideradas, para os horizontes 1 e 3, este modelo apresenta resultados de previsão de qualidade inferior aos demais modelos investigados, contudo este modelo fornece as melhores previsões de longo prazo. Já o melhor modelo de redes neurais, i.e., a rede R_1 , de acordo com a maioria das medidas, apresenta resultados de previsões com erros bem elevados nos primeiros lags e no decorrer do tempo de previsão passa a ter desempenho superior a alguns modelos SARIMA, constituindo assim um bom

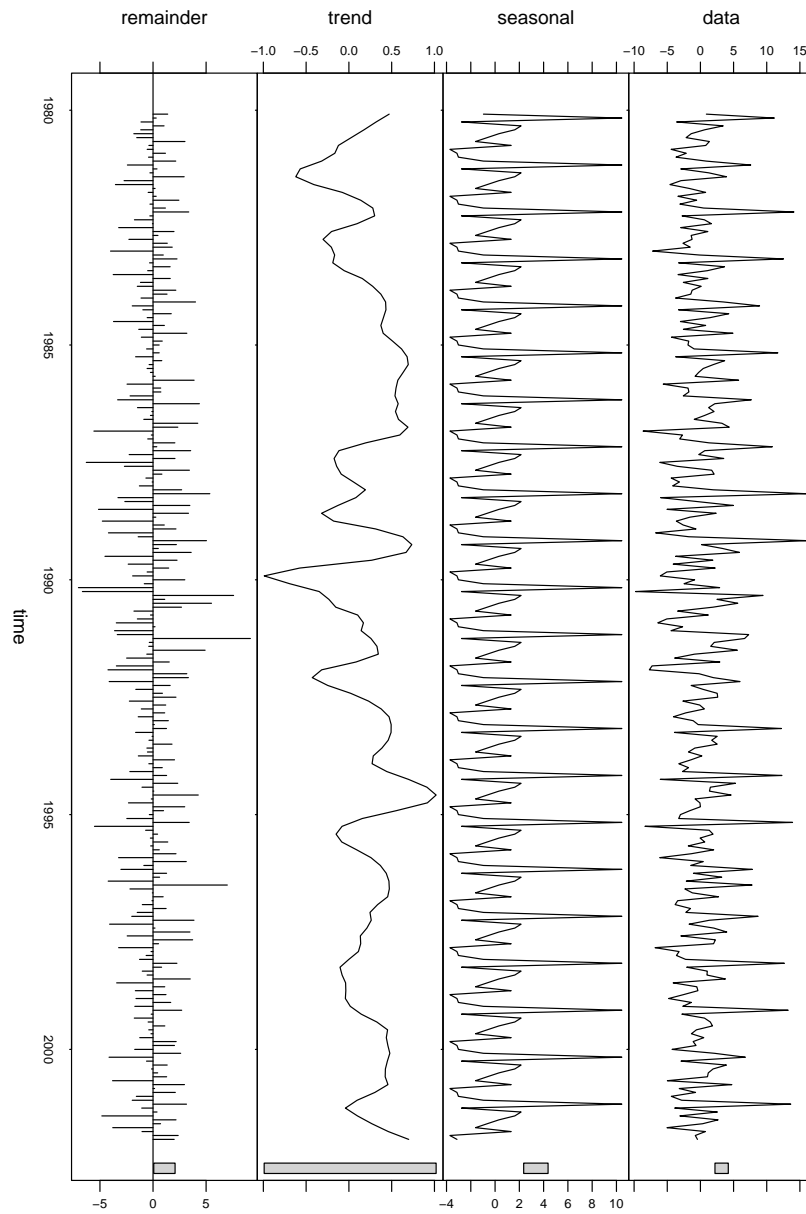


Figura 8: Decomposição STL da série PIB (diferenciada) entre janeiro de 1980 a dezembro de 2001.

Tabela 5: Medidas de previsão dos modelos SARIMA, do algoritmo dealisamento exponencial de Holt-Winters, redes neurais e SARIMAX, para $s = 12$ (Série PIB).

Método	h	EQM	EAM	EM	EAT	ET
Saz. Aditiva A.E. HoltWinters	1	0.85	0.92	-0.92	-0.92	-0.92
	3	3.85	1.59	-1.59	-4.79	-4.79
	6	16.66	3.20	-3.20	-19.23	-19.23
	9	30.02	4.62	-4.62	41.62	-41.62
	12	29.21	4.49	-4.49	-53.98	-53.98
saz. Multiplicativa A.E. HoltWinters	1	1.50	1.22	1.22	1.22	1.22
	3	8.28	2.04	2.04	6.14	6.14
	6	11.52	2.75	2.75	16.54	16.54
	9	8.50	2.36	2.01	21.25	18.17
	12	8.43	2.34	0.94	28.10	11.32
$(3,1,3) \times (1,0,1)$ SARIMA	1	2.81	1.67	-1.67	1.67	-1.67
	-1.11	8.06	2.60	2.04	7.80	-3.96
	6	5.36	1.84	-1.20	11.08	-7.23
	9	8.27	2.45	-2.03	22.13	-18.28
	12	12.53	2.93	-2.61	35.23	-31.38
$(3,1,3) \times (2,0,1)$ SARIMA	1	2.30	1.51	-1.51	1.51	-1.51
	3	8.23	2.60	-1.25	7.82	-3.75
	6	5.68	1.86	-1.18	11.16	-7.09
	9	8.19	2.43	-1.97	21.88	-17.81
	12	12.39	2.90	-2.56	34.89	-30.83
$(3,1,3) \times (0,1,1)$ SARIMA	1	2.45	1.56	-1.56	1.56	-1.56
	3	7.11	2.40	-1.32	7.21	-3.98
	6	5.38	1.90	-1.36	11.44	-8.20
	9	9.44	2.64	2.28	23.84	-20.60
	12	14.29	3.15	-2.88	37.87	-34.64

modelo para previsões de longo prazo devido a sua capacidade de aprendizado ao longo do tempo. Na figura 9 apresentamos o gráfico que corresponde à série do PIB entre o período de janeiro de 2000 e dezembro de 2001 juntamente com as previsões obtidas com os melhores modelos das metodologias de Holt-Winters, Box-Jenkins e redes neurais.

Tabela 6: Medidas de previsão dos modelos SARIMA, do algoritmo dealisamento exponencial de Holt-Winters, Redes neurais e SARIMAX, para $s = 12$ (Série PIB).

Método	h	EQM	EAM	EM	EAT	ET
(3,1,3)×(0,1,2) SARIMA	1	2.03	1.42	-1.42	1.42	-1.42
	3	7.34	2.42	-1.26	7.28	-3.80
	6	5.70	1.92	-1.34	11.53	-8.05
	9	9.32	2.62	-2.23	23.58	-20.10
	12	14.09	3.12	-2.83	37.52	-34.04
(3,1,3)×(1,1,1) SARIMA	1	1.99	1.41	-1.41	1.41	-1.41
	3	7.37	2.43	-1.26	7.29	-3.79
	6	5.76	1.92	-1.34	11.56	-8.06
	9	9.25	2.61	-2.22	23.50	-20.00
	12	13.99	3.11	-2.82	37.40	-33.90
(3,2,3)×(1,2,1) SARIMA	1	2.67	1.63	-1.63	1.63	-1.63
	3	9.28	2.33	-2.09	6.99	-6.29
	6	11.40	2.83	-2.71	16.99	-16.09
	9	21.89	4.05	-3.97	36.51	-35.81
	12	31.97	4.91	-4.85	58.92	-58.22
Redes neurais	1	42.23	6.49	-6.49	6.49	-6.49
	3	48.87	6.74	-3.66	20.22	-11.00
	6	26.90	4.32	-0.88	25.93	-5.29
	9	21.59	3.84	0.23	34.63	2.09
	12	16.24	2.97	0.26	35.73	3.18
(3,1,3)×(2,0,1) SARIMAX _{ts}	1	128.87	11.35	-11.35	11.35	-11.35
	3	121.85	9.21	-8.59	27.63	-25.78
	6	65.54	6.05	-5.74	36.32	-34.47
	9	46.24	4.94	-4.73	44.46	-42.61
	12	45.24	5.29	-5.14	63.53	-61.68

Em nossa análise, encontramos que o modelo de Holt-Winters com sazonalidade multiplicativa foi o que apresentou as melhores previsões em comparação com as demais metodologias. De maneira semelhante à série do PIB brasileiro será analisada a série do IPC da Colômbia. Na figura 23 foi apresentado o

gráfico da série do IPC ao longo do tempo, onde é razoável considerar a existência de uma componente sazonal. Na figura 10 é apresentado o gráfico da função de autocorrelação amostral da série, onde pode-se notar um decaimento senoidal em direção a zero o que sugere que a série possui uma parte auto-regressiva e uma componente sazonal bem definidas. Na figura 11 é apresentado o gráfico da função de autocorrelação parcial amostral, onde percebe-se a componente sazonal da série. Para confirmar a situação descrita acima, na figura 12 apresentamos a decomposição STL na série IPC, confirmando-se a existência da componente sazonal e verificamos a presença de uma componente de tendência que decresce levemente.

Dadas as características desta série é aplicado o algoritmo de Holt-Winters, onde a escolha das constantes de suavização foi feita de forma automática. Quanto à modelagem de Box-Jenkins, analisamos a série original e também a série obtida depois de aplicar a função logaritmo aos dados com o objetivo de tentar estabilizar a variância. Checamos os modelos variando as partes auto-regressivas e médias móveis variando a ordem dos parâmetros até grau 3 em suas componentes e utilizamos diferenciações $d = 0, 1$ e $D = 0, 1$ para o modelo SARIMA. De acordo com os critérios AIC e BIC foram escolhidos alguns modelos, os quais estão apresentados na tabela 8.

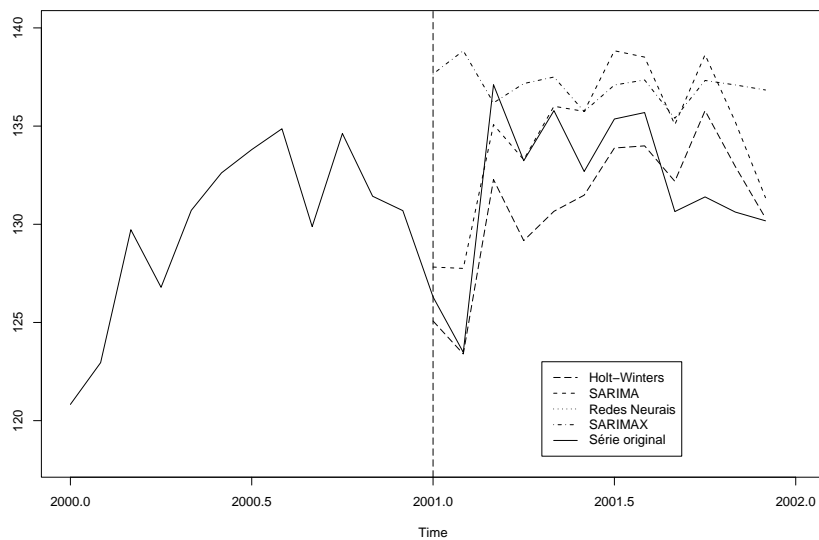


Figura 9: Previsões série PIB.

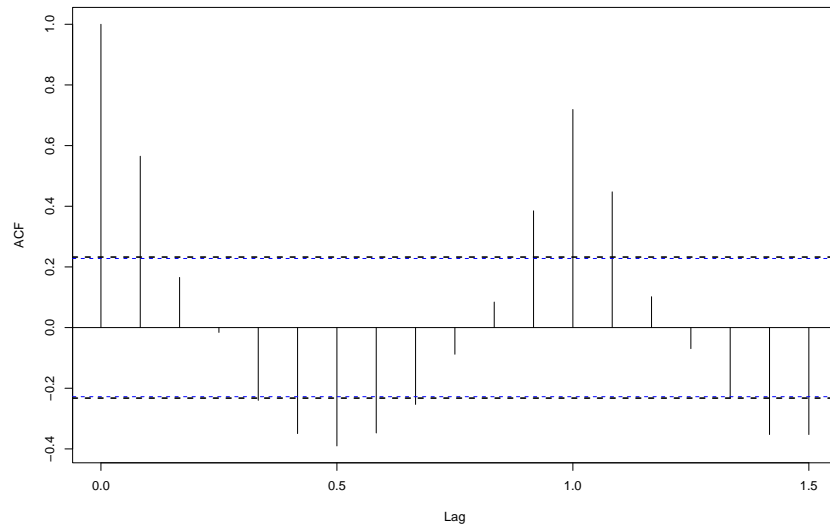


Figura 10: Correlograma da série IPC.

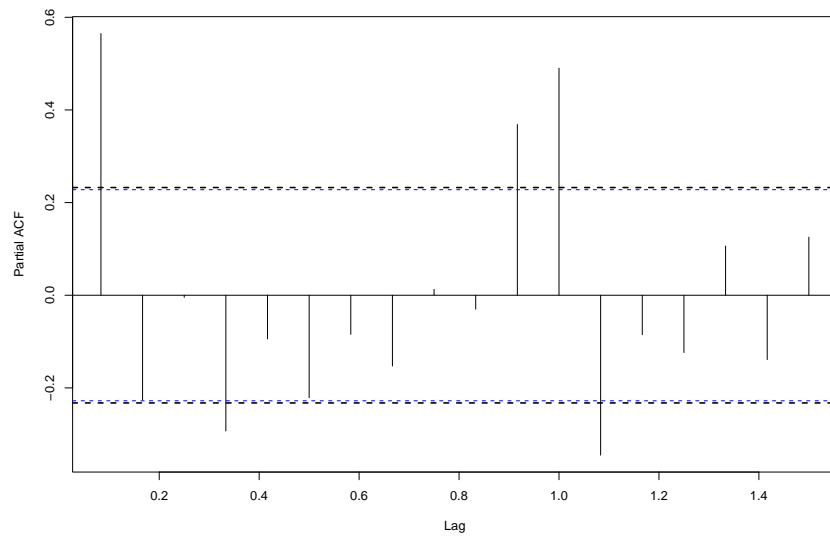


Figura 11: Correlograma parcial da série IPC

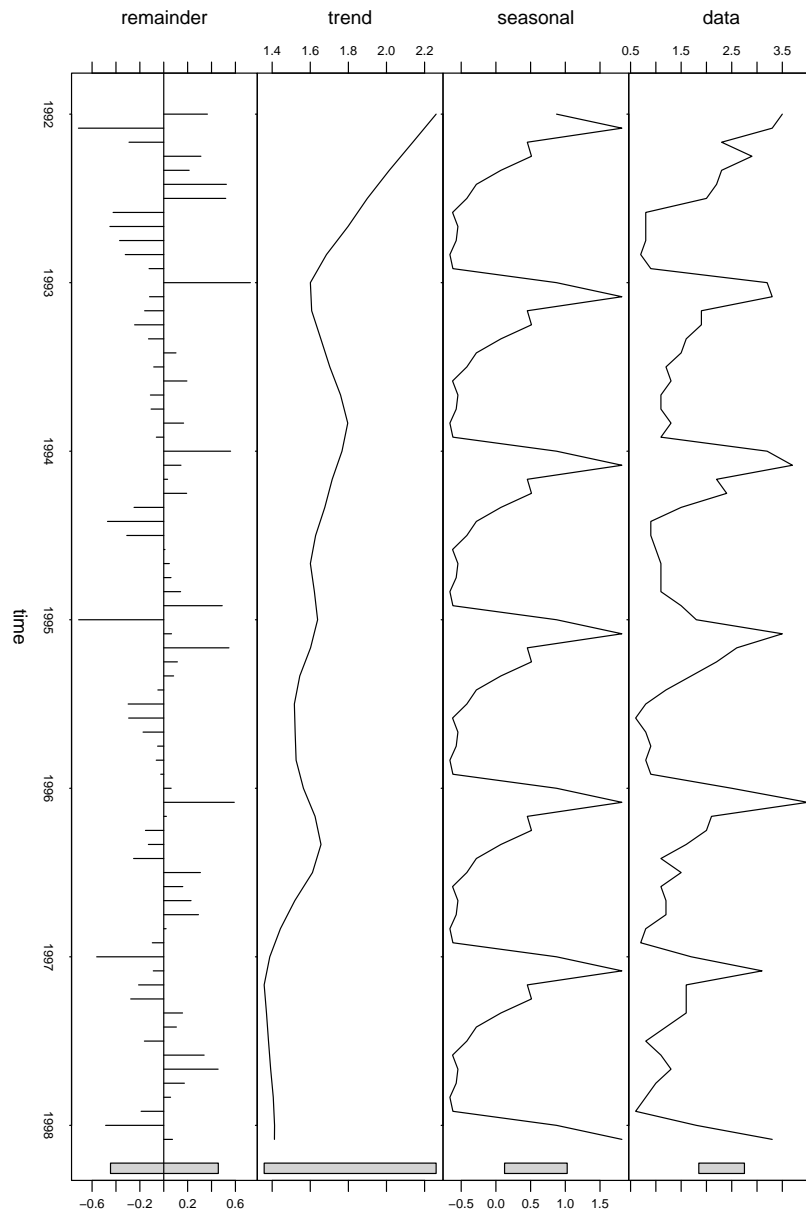


Figura 12: Decomposição STL da série IPC.

A análise da autocorrelação serial foi feita em todos estes modelos e não foi rejeitada a hipótese de adequação, o que indica uma forte inclinação a serem aplicados estes modelos para realizar previsões. Usando uma argumentação análoga à utilizada na análise da série do PIB, foram obtidas previsões sob as metodologias descritas anteriormente. Na tabela 9 são apresentadas as medidas de erro de previsão para todos os modelos para a série IPC da Colômbia. Dentre a classe dos modelos SARIMA considerados, o modelo $(2,0,3) \times (1,1,1)$ foi o que gerou as melhores resultados de previsões. Quanto ao modelo de redes neurais, observou-se um desempenho ruim nas previsões, que indica que este modelo não se ajustou adequadamente à série. Uma análise desta tabela indica que os algoritmos de Holt-Winters foram superiores a todos os outros modelos em praticamente todas as medidas consideradas. Observamos que o algoritmo de Holt-Winters com sazonalidade multiplicativa é o que apresenta o melhor desempenho em termos da qualidade das previsões, isto se confirma praticamente para todos os horizontes e para todas as medidas consideradas.

Tabela 7: Critérios AIC e BIC para ajuste na série IPC.

$(p,d q)(P,D,Q)$	AIC	BIC
$(2,0,3) \times (1,1,1)$	65.93	83.56
$(2,0,3) \times (1,1,1) \log(\text{IPC})$	14.02	31.66
$(0,1,2) \times (1,1,1) \log(\text{IPC})$	17.49	28.51
$(1,0,1) \times (1,1,1)$	15.67	26.69

Na figura 13 apresentamos a parte correspondente da série do IPC entre o período de julho de 1996 a fevereiro de 1998 juntamente com as previsões obtidas com os melhores modelos das metodologias de Holt-Winters, Box-Jenkins e redes neurais.

4. Conclusões

Na seção anterior foram apresentados resultados que avaliam a qualidade das previsões sob diferentes metodologias de estimação de duas séries temporais: a série do PIB do Brasil e a série do IPC da Colômbia. Como critérios de avaliação foram utilizados o EQM, EAM, EM, EAT, ET de previsão sob diferentes horizontes de previsão. Os resultados obtidos revelam que as previsões de melhor qualidade um passo à frente para a série do PIB são obtidas através do modelo de Holt-Winters com sazonalidade aditiva; dentro da classe de modelos SARIMA, as previsões obtidas com o modelo $\text{SARIMA}(3, 1, 3) \times (2, 0, 1)$ apresentam um melhor comportamento para longos períodos de previsão.

Tabela 8: Erros de previsão dos modelos SARIMA, do algoritmo dealisamento exponencial de Holt-Winters e Redes neurais, para $s = 12$ (Série IPC).

Método	h	EQM	EAM	EM	EAT	ET
Saz. Aditiva A.E. HoltWinters	1	0.05	0.24	0.24	0.24	0.24
	3	0.07	0.23	0.23	0.69	0.69
	5	0.04	0.17	0.11	0.89	0.55
	7	0.04	0.16	0.04	1.16	0.32
Saz. Multiplicativa A.E. HoltWinters	1	0.01	0.14	0.14	0.14	0.14
	3	0.03	0.15	0.11	0.46	0.33
	5	0.03	0.16	-0.003	0.81	-0.01
	7	0.03	0.15	-0.04	1.11	-0.31
(2,0,3)×(1,1,1) SARIMA	1	0.05	0.24	0.24	0.24	0.24
	3	0.04	0.19	0.19	0.57	0.57
	5	0.04	0.19	0.03	0.95	0.18
	7	0.03	0.15	0.009	1.07	0.06
(2,0,3)×(1,1,1) SARIMA log (ipc)	1	0.07	0.26	0.26	0.26	0.26
	3	0.05	0.22	0.22	0.67	0.67
	5	0.09	0.24	0.22	1.24	0.11
	7	0.08	0.23	-0.04	1.66	-0.30
(0,1,2)×(1,1,1) SARIMA log (ipc)	1	0.21	0.46	0.46	0.46	0.46
	3	0.13	0.34	0.34	1.02	1.02
	5	0.10	0.27	0.14	1.39	0.72
	7	0.07	0.23	0.10	1.64	0.73
(1,0,1)×(1,1,1) SARIMA	1	0.15	0.39	0.39	0.39	0.39
	3	0.08	0.24	0.22	0.73	0.68
	5	0.09	0.25	0.03	1.26	0.15
	7	0.07	0.21	-0.012	1.51	-0.09
Redes neurais	1	0.25	0.5	-0.5	0.5	-0.5
	3	0.11	0.29	-0.10	0.89	-0.31
	5	0.23	0.39	-0.28	1.99	-1.41
	7	0.51	0.59	-0.17	4.19	-1.21

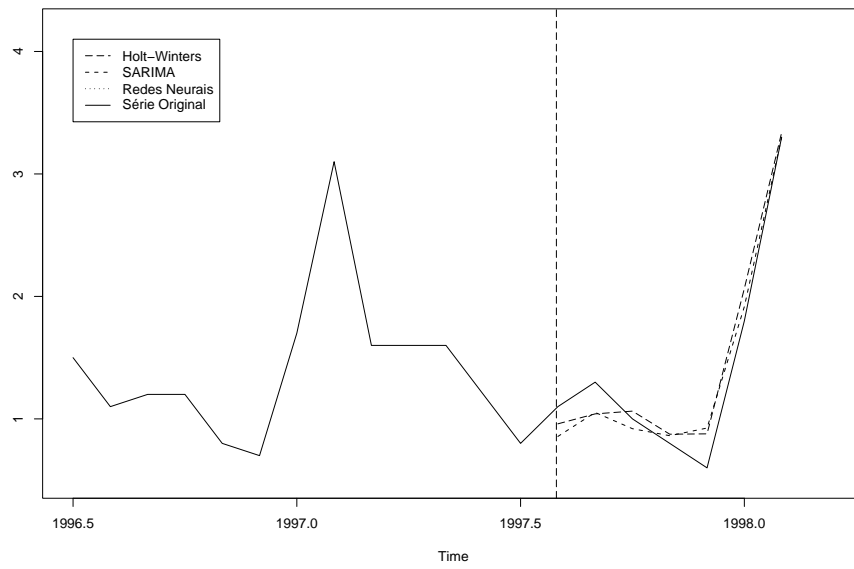


Figura 13: Previsões série IPC.

De todos os modelos avaliados, o algoritmo de Holt-Winters com sazonalidade multiplicativa é o que apresenta os melhores resultados de previsão. Além disso, o custo computacional deste algoritmo é bem menor do que os modelos de Box-Jenkins. O modelo selecionado de redes neurais apresenta um desempenho melhor para longos períodos de tempo, pois parece que se ajusta mais rapidamente às mudanças estruturais através do tempo. De forma análoga, na análise da série do IPC, observamos que a metodologia de Holt-Winters se mostrou superior à modelagem de Box-Jenkins, onde o modelo de Holt-Winters com sazonalidade multiplicativa foi superior a todos os demais modelos em análise. Uma alternativa interessante a ser analisada para um trabalho futuro de pesquisa é o de tratar de combinar as previsões de uma série temporal sob as metodologias tratadas neste trabalho com o fim de aumentar o poder preditivo e proporcionar previsões mais confiáveis.

5. Agradecimentos

Os autores agradecem a Francisco Cribari-Neto pela motivação e pelas pertinentes recomendações ao longo deste trabalho. Os autores agradecem também à CAPES e ao CNPq pelo apoio financeiro.

References

- Box, G. E. P. & Pierce, D. A. (1970), 'Distribution of residual autocorrelations in autoregressive integrated moving average time series models', *Journal of the American Statistical Association* **65**, 1509–1526.
- Box, G. P. & Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco, Holden Day.
- Broomhead, D. S. & Lowe, D. (1988), 'Multivariable functional interpolation and adaptive networks', *Complex Systems* **2**, 321–355.
- Cheng, B. & Titterton, D. M. (1994), 'A review from statistical perspective', *Statistical Science* **9**, 2–54.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. (1990), 'Stl: a seasonal-trend decomposition procedure based on loess', *Journal of Official Statistics* **6**, 3–73.

- Gershenfeld, N. A. & Weigend, A. (1994), *The Future of Time Series: Learning and Understanding. Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison Wesley, New York.
- Goldberg, D. E. (1989), *Genetic Algorithms*, Reading: Addison Wesley, New York.
- Golub, G. & Kahan, W. (1965), 'Calculating the singular values and pseudo-inverse of a matrix', *SIAM Numerical Analysis B* **2**, 205–224.
- Hertz, J., Krogh, A. & Palmer, R. (1991), *Introduction to the Theory of Neuronal Computation*, Addison Wesley, New York.
- Hunter, A., Kennedy, L., Henry, J. & Ferguson, R. I. (2000), 'Application of neural networks and sensitivity analysis to improved prediction of trauma survival', *Computer Methods and Algorithms in Biomedicine* **62**, 11–19.
- Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time serie moldels', *Biometrika* **65**, 297–303.
- McLeod, A. I. (1978), 'On the distribution and application of residual autocorrelations in box-jenkins moldels', *Journal of the Royal Statistical Society B* **40**, 296–302.
- Mills, J. A. & Prasad, K. (1992), 'A comparison of moldel selection criteria', *Econometrika* **11**, 201–233.
- Pi, H. & Peterson, C. (1994), 'Finding the embeddindg dimension and variable dependecies in time series', *Neural Computation* **6**(3), 509–520.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986), *Learning Internal Representations by Error Propagation*, Vol. 1, Cambridge, MA: MIT Press, Parallel Distributed Processing.
- Torres, L. G. (1995), 'Redes neuronales y aproximación de funciones', *Boletín de matemáticas* **2**, 35–57.
- Weigend, A. S., Rumelhart, D. E. & Huberman, B. A. (1991), *Generalization by weight-elimination with application to forecasting*, Vol. 3 of *Advances in Neural Information Processing Systems*, pp. 875–882.