

# CONSTRUÇÃO DE INTERVALOS DE PREDIÇÃO PARA REDES NEURAS VIA REGRESSÃO MULTIVARIADA E SUA APLICAÇÃO EM SÉRIES HETEROCEDÁSTICAS

CONSTRUCTION OF PREDICTION INTERVALS FOR NEURAL NETWORKS VIA MULTIVARIATE REGRESSION AND THEIR APPLICATION TO HETEROCEDASTIC SERIES

---

Mauri Aparecido de **OLIVEIRA**

Doutorando da Faculdade de Economia, Administração e Contabilidade da USP  
Professor do Instituto Presbiteriano Mackenzie  
mauriao@usp.br

Alessandra de Ávila **MONTINI**

Professora Doutora da Faculdade de Economia, Administração e Contabilidade da USP  
amontini@usp.br

Daniel Reed **BERGMANN**

Doutorando da Faculdade de Economia, Administração e Contabilidade da USP  
Professor do Instituto Presbiteriano Mackenzie  
dreedbergmann@yahoo.com.br

## Resumo

O objetivo deste trabalho é apresentar a teoria para construção de intervalos de predição (IP) e de confiança (IC) para redes neurais artificiais (RNA) alimentadas adiante de múltiplas camadas – feedforward – baseada em regressão multivariada. Essa técnica foi aplicada em um conjunto de séries temporais univariadas ARIMA-GARCH construídas por simulação e em uma série de preço da saca da soja. O desempenho das topologias de RNA é comparado por meio do coeficiente de desigualdade de Theil, do RMSE e pela quantidade de valores que ficam fora do IP e do IC. As redes neurais têm sido utilizadas com frequência na análise de séries temporais e a utilização de intervalos de predição e confiança pode ser considerada uma ferramenta adicional para avaliar o desempenho das RNA.

**Palavras-chave:** Redes Neurais Artificiais – Intervalos de Confiança – Intervalos de Predição – Séries Temporais.

## Abstract

This work presents the theory to construct prediction (PI) and confidence intervals (CI) for feedforward artificial neural networks (ANN) based on multivariate regression. This technique was applied to a set of ARIMA-GARCH time series built by simulation and to a soybean price series. The ANN topologies performance is compared using the Theil inequality coefficient (TIC), root mean square error (RMSE) and the quantities of values outside the PI and CI. Neural networks have been intensively applied to time series analysis and the use of prediction and confidence intervals can be considered an additional tool to evaluate ANN performance.

**Key-words:** Artificial Neural Networks – Confidence Intervals – Prediction Intervals – Time Series.

## Introdução

Quando as redes neurais artificiais (RNA) são usadas em problemas de predição, é desejável que sejam construídos intervalos de predição e de confiança para os valores preditos. Os métodos de estimação desses intervalos dependem da topologia de RNA que é utilizada e esses métodos são válidos sob certas suposições.

Para modelar qualquer série econométrica é importante avaliar a qualidade do modelo construído. Isso é bastante estudado em problemas que utilizam a estrutura da teoria da regressão linear, levando a resultados bem estabelecidos e métodos comumente utilizados para construir intervalos de confiança (IC) e predição (IP) para regressão, mais recentemente, esses resultados têm sido estendidos para modelos não lineares.

Este artigo apresenta um método de construção de IC e IP baseado em regressão multivariada para RNA alimentadas adiante. Essa abordagem é construída a partir do método dos mínimos quadrados que aplica a expansão de Taylor em modelos não lineares.

Os intervalos são construídos para as previsões de séries temporais ARIMA-GARCH geradas por simulação e para a série real de preço da saca de 60 Kg da soja.

Hwang e Ding (1997) construíram intervalos de confiança e de predição para redes neurais de múltiplas camadas alimentadas adiante e aplicaram a metodologia em séries temporais simuladas. Rivals e Personnaz (2000) apresentam resultados teóricos sobre a construção de intervalos de confiança para regressão não-linear baseados na estimação de mínimos quadrados utilizando expansão de Taylor. Outros métodos para a construção de intervalos de confiança baseados em regressão não-linear foram propostos por De Vaux et al. (1998).

Métodos propostos para construção de intervalos de predição e confiança para redes neurais, podem ser encontrados por exemplo em

Chryssolouris et al. (1996), Dybowski (2001) e Castellano et al. (2003). Muitas destas técnicas são dependentes da arquitetura específica de rede neural que é utilizada. Por esse motivo, tais técnicas não se aplicam aos diversos tipos de RNA com diferentes topologias e conexões.

## 1 Referencial teórico

Uma rede alimentada adiante tem todos os sinais indo em uma direção, dos neurônios de entrada para os neurônios de saída. As redes neurais alimentadas adiante de múltiplas camadas também são chamadas de *multilayer feedforward neural networks (MLFN)*.

Uma conexão entre dois neurônios tem um valor numérico (peso) representando a influência do neurônio de entrada sobre o neurônio de saída. Um caso especial desse tipo de rede, com apenas uma camada oculta e uma saída, é mostrado na Figura 1.

Usando a Figura 1 como exemplo, os sinais de entrada  $x_1, x_2, \dots, x_y$  são linearmente combinados com os vários pesos resultando em vários sinais de entrada para a segunda camada. Esses sinais de entrada são, então, passados através de uma função de ativação para produzir os sinais de saída dos neurônios da segunda camada. A função de ativação, utilizada neste trabalho, é a função logística:

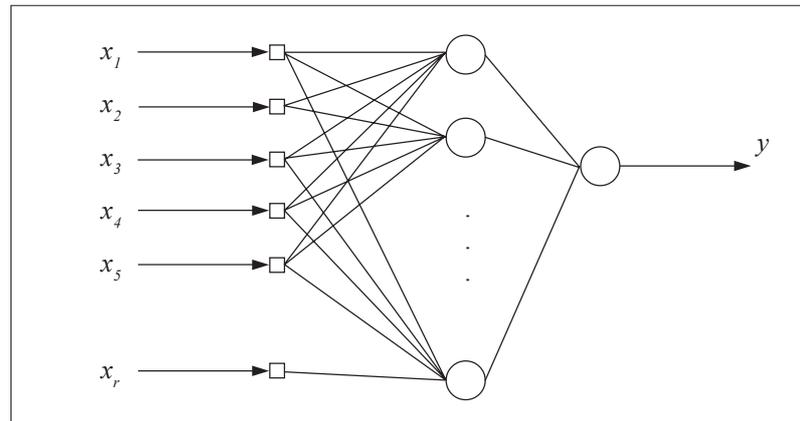
$$f(x) = \frac{1}{1+e^{-x}}$$

Tendo uma função de ativação, a saída total da rede, dada a entrada  $\mathbf{x}=(x_1, x_2, \dots, x_y)^t$ , é dada por:

$$y_w(\mathbf{x}) = w_{20} + \sum_{i=1}^k w_{2i} f(\mathbf{w}'_{li} \mathbf{x} + w_{li0}) \quad (1)$$

em que  $\mathbf{w}'_{li} = (w_{li1}, w_{li2}, \dots, w_{liy})$  e os pesos, ou parâmetros não conhecidos, são representados pela matriz  $\mathbf{W} = (w_{20}, \dots, w_{2k}, w_{110}, \dots, w_{lk0}, \mathbf{w}'_{11}, \dots, \mathbf{w}'_{lk})$ , sendo que o negrito indica notação matricial e o sobrescrito ' denota a transposta (HWANG e DING, 1997).

**Figura 1** - Rede neural alimentada adiante de três camadas



Neste trabalho assume-se um modelo estatístico que relaciona  $y$  e  $y_w(\mathbf{x})$  da seguinte forma:

$$y = y_w(\mathbf{x}) + \varepsilon, \quad (2)$$

em que os  $\mathbf{x}_i$ 's são independentes dos erros  $\varepsilon_i$ 's e que  $(\mathbf{x}_i, \varepsilon_i)$  são iid (independentes e identicamente distribuídos).

Suponha-se que  $\varepsilon$  satisfaz as seguintes condições:

1.  $E[\varepsilon] = 0$ ;
2.  $Cov[\varepsilon] = E[\varepsilon\varepsilon'] = \sigma^2 \mathbf{I}$ .

Dada uma série de tamanho  $n$ , a rede é treinada com os vetores  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$ . Esses vetores são usados para determinar o estimador  $\hat{\mathbf{W}}$  dos pesos  $\mathbf{W}$ . Essa rede é usada para prever a saída futura de uma nova entrada  $\mathbf{x}_{n+1}$  por  $\hat{y} = y_w(\mathbf{x}_{n+1})$ .

Assume-se que para todo  $1 \leq i \leq n$ ,  $(y_i, \mathbf{x}_i)$  satisfaz (1) e (2), ou seja  $y_i = y_w(\mathbf{x}_i) + \varepsilon_i$ ,  $y_{n+1}$  é uma variável aleatória não observável que é o objetivo da predição.

O objetivo é construir intervalos de predição para  $y_{n+1}$  e intervalos de confiança para  $y(\mathbf{x}_{n+1})$ , que é a esperança condicionada de  $y_{n+1}$  dado  $\mathbf{x}_{n+1}$ .

### 1.1 Teoria de intervalo de predição

O problema de estimação dos parâmetros da equação (1) utilizando a função objetivo de mínimos

quadrados pode também ser formulado como um problema de regressão não-linear. Os métodos que fornecem intervalos de predição assintóticos para regressão não-linear podem, em teoria, ser aplicados diretamente às redes neurais. Nessa seção, fazemos uma revisão dessa teoria (SEBER e WILD, 1989, NETER et al., 1996, JOHNSON e WICHERN, 2002).

Considere o modelo de regressão não-linear definido por:

$$y = f(\mathbf{x}, \mathbf{w}) + \varepsilon \quad (3)$$

sendo  $\mathbf{w}$  um vetor de parâmetros e  $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$ .

No caso da rede neural,  $f$  é dada pelo lado direito da equação (1).

A equação (3) pode ser reescrita como:

$$y = \mathbf{x} \cdot \mathbf{w} + \varepsilon \quad (4)$$

$\begin{matrix} (n \times 1) & (n \times (r+1)) & ((r+1) \times 1) & (n \times 1) \end{matrix}$

O vetor de parâmetros ( $\mathbf{w}$ ) é estimado por meio da minimização da soma dos resíduos ao quadrado dada por:

$SQR = \frac{1}{2} (y - \hat{y})'(y - \hat{y})$  em que  $\hat{y}$  é o vetor de valores preditos.

Na vizinhança dos valores dos parâmetros, que denotaremos por  $\mathbf{w}^*$ , a função  $f$  pode ser aproximada por uma série de Taylor:

$\hat{y} = f(\mathbf{w}^*) + \mathbf{J}(\mathbf{w} - \mathbf{w}^*)$ , sendo que  $\mathbf{J}$  é a matriz cuja  $i$ -ésima entrada é  $\partial f(\mathbf{x}_i) / \partial w_j$ , avaliada no vetor de parâmetros  $\mathbf{w}^*$ , e  $\mathbf{x}_i$  é a  $i$ -ésima linha de  $\mathbf{X}$ , correspondendo ao  $i$ -ésimo conjunto de dados. Pode-se aproximar a soma dos resíduos ao quadrado por:

$$SQR \approx \frac{1}{2} (\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^* - \mathbf{J}\hat{\mathbf{w}})' (\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^* - \mathbf{J}\hat{\mathbf{w}}) \quad (5)$$

$$SQR \approx \frac{1}{2} [(\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^*) - \mathbf{J}\hat{\mathbf{w}}]' [(\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^*) - \mathbf{J}\hat{\mathbf{w}}] \\ = \frac{1}{2} [\mathbf{k}'\mathbf{k} - 2\mathbf{J}'\hat{\mathbf{w}}'(\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^*) + \hat{\mathbf{w}}'\mathbf{J}'\mathbf{J}\hat{\mathbf{w}}] \quad (6)$$

em que  $\mathbf{k} = \mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^* = \boldsymbol{\varepsilon}^* + \mathbf{J}\mathbf{w}^*$

Para se obter uma estimativa de mínimos quadrados dos parâmetros, ou seja  $\hat{\mathbf{w}}$ , pode-se igualar a zero a derivada de  $SQR$  com relação a  $\hat{\mathbf{w}}$ .

A soma quadrada dos resíduos também pode ser escrita como:

A derivada da expressão (6) com relação a  $\hat{\mathbf{w}}$  é dada por:

$$\frac{\partial SQR}{\partial \hat{\mathbf{w}}} = \mathbf{J}'\hat{\mathbf{w}} - \mathbf{J}'(\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^*) \quad (7)$$

Igualando a expressão (7) a zero, tem-se que:

$$\hat{\mathbf{w}} = (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}'(\mathbf{y} - f(\mathbf{w}^*) + \mathbf{J}\mathbf{w}^*) \quad (8)$$

A variância assintótica de  $\hat{\mathbf{w}}$  pode ser obtida pelo teorema de Gauss-Markov que resulta em:

$$Var[\hat{\mathbf{w}}] = \sigma^2 (\mathbf{J}'\mathbf{J})^{-1} \quad (9)$$

O estimador assintoticamente não-viesado de  $\sigma^2$  é dado por:

$$s^2 = \frac{SQR}{(n-r-1)} \quad (10)$$

Dados os parâmetros estimados  $\hat{\mathbf{w}}$  e um novo conjunto de observações  $\mathbf{x}_0$ , pode-se aproximar o valor predito de  $f(\mathbf{x}, \mathbf{w})$  em  $\mathbf{x}_0$ , por meio da expansão de Taylor, por:

$$\hat{y}_0 = f(\mathbf{x}_0, \hat{\mathbf{w}}) \approx f(\mathbf{x}_0, \mathbf{w}^*) + \mathbf{J}'_0(\mathbf{w} - \mathbf{w}^*) \quad (11)$$

em que  $\mathbf{J}_0$  é um vetor cuja  $i$ -ésima entrada é  $\partial f(\mathbf{x}_0, \mathbf{w}) / \partial w_i$ .

A expressão (10) pode ser aproximada por:

$$\hat{y}_0 \approx f(\mathbf{x}_0, \mathbf{w}^*) + \mathbf{J}'_0 [(\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}'(\boldsymbol{\varepsilon}^* + \mathbf{J}\mathbf{w}^*) - \mathbf{w}^*] \\ = f(\mathbf{x}_0, \mathbf{w}^*) + \mathbf{J}'_0 [(\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}'\boldsymbol{\varepsilon}^* + \mathbf{w}^* - \mathbf{w}^*] \\ = f(\mathbf{x}_0, \mathbf{w}^*) + \mathbf{J}'_0 (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}'\boldsymbol{\varepsilon}^*. \quad (12)$$

A variância de  $\hat{y}$  condicionada a  $\mathbf{x}_0$  pode ser aproximada por

$$Var[\hat{y} | \mathbf{x}_0] = Var[f(\mathbf{x}_0, \mathbf{w}^*) + \mathbf{J}'_0 (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}'\boldsymbol{\varepsilon}^*],$$

como a variância de  $f(\mathbf{x}_0, \mathbf{w}^*)$  é igual a zero pelo fato de que os estados  $\mathbf{x}_0$  e  $\mathbf{w}^*$  são conhecidos, e como se assumiu que o vetor de erros é não correlacionado com nenhum dos outros vetores de parâmetros, conclui-se que a variância condicionada pode ser aproximada por:

$$Var[\hat{y} | \mathbf{x}_0] = \sigma^2 \mathbf{J}'_0 (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}_0 \quad (13)$$

Tem-se que,  $n$  é o número de valores que correspondem às variáveis preditas e  $r$  é o número de parâmetros. Como  $y_0$  denota o valor de uma resposta quando a variável preditora tem os valores  $\mathbf{x}'_0 = [1, x_{01}, \dots, x_{0r}]$ , de acordo com o modelo da equação (4), o valor esperado de  $y_0$  é:

$$E[y_0 | \mathbf{x}_0] = w_0 + w_1 x_{01} + \dots + w_r x_{0r} = \mathbf{x}'_0 \mathbf{w}.$$

Para o modelo de regressão da equação (4),  $\mathbf{x}'_0 \hat{\mathbf{w}}$  é o estimador não-viesado de  $E[y_0 | \mathbf{x}_0]$  com variância mínima dada por

$$Var[\hat{y} | \mathbf{x}_0] = \sigma^2 \mathbf{J}'_0 (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}_0$$

Quando os erros  $\boldsymbol{\varepsilon}$  são normalmente distribuídos, o intervalo com  $100(1-\alpha)\%$  de confiança (IC) para  $E[y_0 | \mathbf{x}_0] = \mathbf{x}'_0 \mathbf{w}$  é dado por (JOHNSON e WICHERN, 2002, p. 374):

$$\mathbf{x}'_0 \hat{\mathbf{w}} \pm t_{n-r-1} \sqrt{s^2 [\mathbf{J}'_0 (\mathbf{J}'\mathbf{J})^{-1} \mathbf{J}_0]} \quad (14)$$

em que  $t_{n-r-1}$  é o  $100(1-\alpha)$ -ésimo percentil da distribuição t de Student com  $n-r-1$  graus de liberdade.

A predição de uma nova observação, tal como  $y_0$ , em  $\mathbf{x}'_0 = [1, x_{01}, \dots, x_{0r}]$ , é mais incerta que estimar o valor esperado de  $y_0$ .

Quando os erros  $\mathbf{\varepsilon}$  têm uma distribuição normal, um intervalo de predição (IP) de  $100(1-\alpha)\%$  para  $y_0$  é dado por (JOHNSON e WICHERN, 2002, p. 375):

$$\mathbf{x}'_0 \hat{\mathbf{w}} \pm t_{n-r-1} \sqrt{s^2 [\mathbf{1} + \mathbf{J}'_0 (\mathbf{J}' \mathbf{J})^{-1} \mathbf{J}_0]} \quad (15)$$

em que  $t_{n-r-1}$  é o  $100(1-\alpha)$ -ésimo percentil da distribuição t de Student com  $n-r-1$  graus de liberdade.

## 2 Modelo AR-GARCH

Os modelos auto-regressivos com heterocedasticidade condicional generalizado – GARCH – foram proposto por Bollerslev (1986). Esses modelos são muito utilizados para a modelagem da volatilidade de séries temporais, ou seja, eles são utilizados quando a volatilidade da série não é constante (MORETTIN, 2004, HAMILTON, 1994, ENDERS, 2003).

Uma série temporal segue um modelo AR([1])-GARCH(1,1) quando puder ser escrita na forma:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + v_t,$$

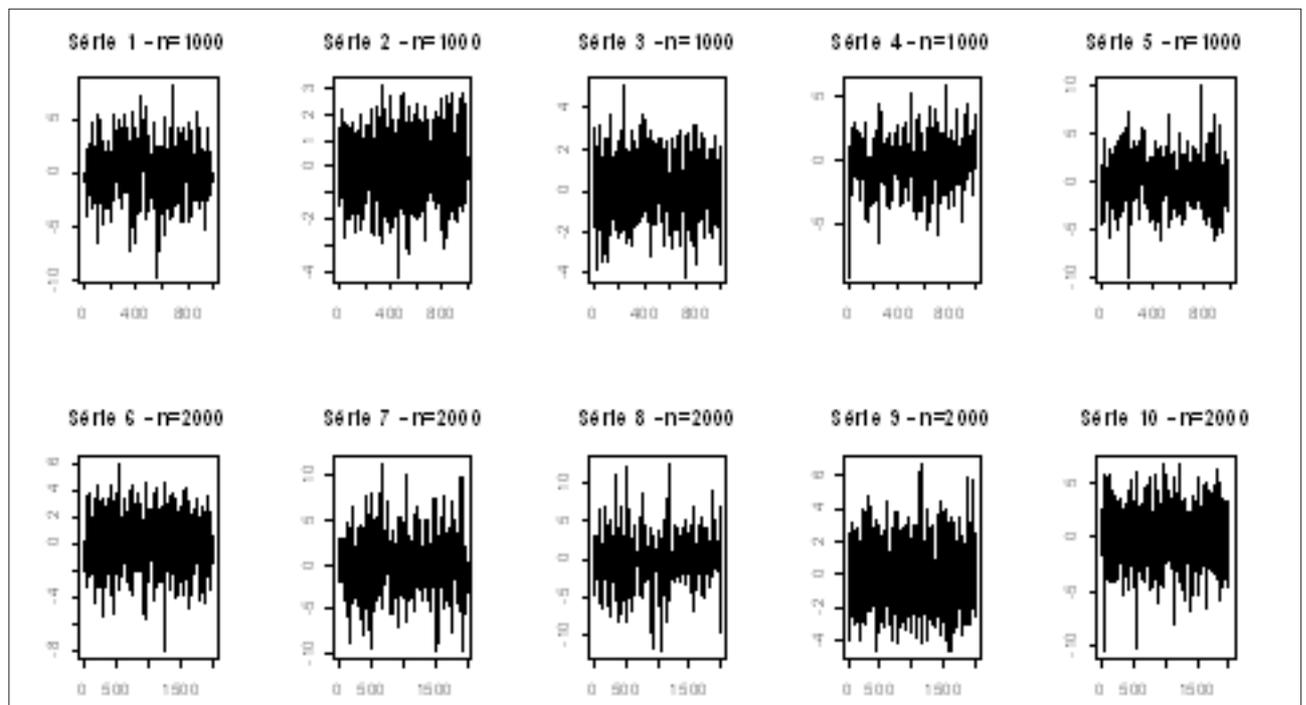
$$v_t = \sqrt{\sigma_t^2} \eta_t \quad (16)$$

sendo que  $\sigma_t^2 = \alpha_0 + \alpha_1 v_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ , e a variável aleatória  $\eta_t$  possui média 0 e variância 1.

## 3 Resultados

A teoria para a construção de intervalos de confiança e predição, apresentada neste artigo, foi aplicada a séries temporais AR([1])-GARCH(1,1) geradas por simulação. Foram geradas, utilizando o programa Mathematica versão 5.0, cinco séries com 1000 observações e cinco com 2000 observações, ver Figura 2.

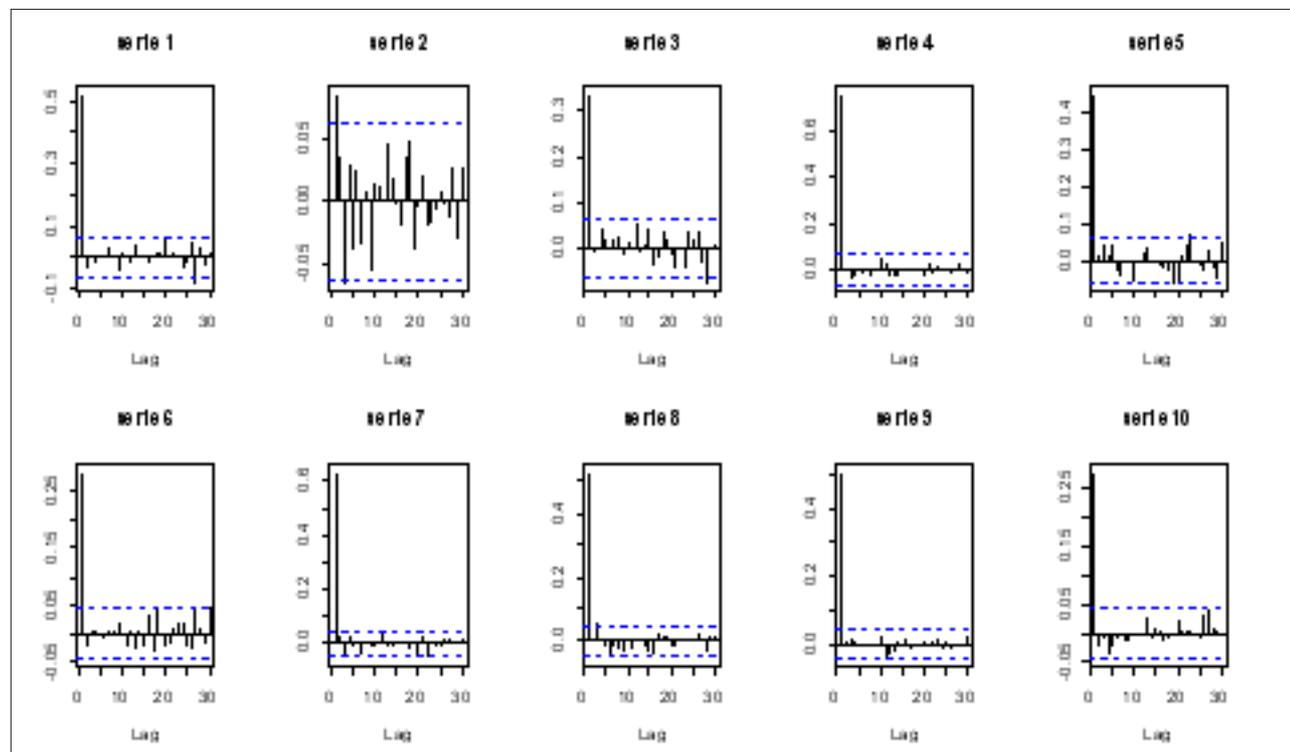
**Figura 2** - Séries AR([1])-GARCH(1,1) geradas por simulação



Pode-se verificar, a partir das funções de auto-correlação parcial (FACP) das séries simuladas, ver Figura 3, que a auto-correlação parcial de cada

uma dessas séries é truncada no “lag” um, caracterizando modelos auto-regressivos de primeira ordem (AR(1)).

**Figura 3** – Funções de auto-correlação parcial das séries geradas por simulação



Os parâmetros estimados das séries temporais AR([1])-GARCH(1,1) geradas por simulação,

dados pela equação 16, são apresentados na Tabela 1.

**Tabela 1** – Parâmetros estimados das séries AR([1])-GARCH(1,1) geradas

Série	Tamanho da série	$\phi_1$	$\alpha_0$	$\alpha_1$	$\beta_1$
Série1	1000	0,517	0,707	0,183	0,609
Série2	1000	0,083	0,057	0,038	0,915
Série3	1000	0,353	0,478	0,182	0,478
Série4	1000	0,761	0,234	0,281	0,564
Série5	1000	0,444	0,333	0,143	0,772
Série6	2000	0,289	0,978	0,186	0,351
Série7	2000	0,668	0,768	0,333	0,452
Série8	2000	0,581	0,753	0,750	0,082
Série9	2000	0,504	0,811	0,051	0,479
Série10	2000	0,284	1,255	0,304	0,338

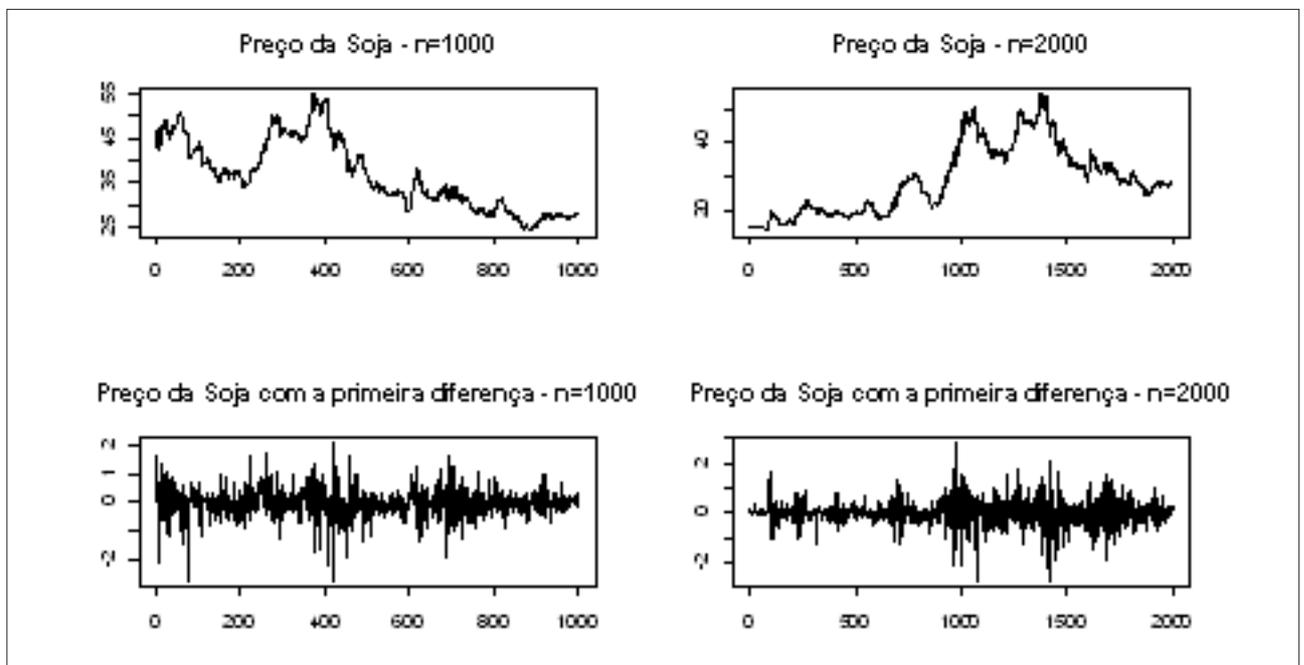
A metodologia para a construção dos IC e IP apresentada foi aplicada a série do preço da saca de 60 quilos de soja, de 8/9/98 a 22/9/06. Essas séries com 1000 e 2000 observações, respectivamente, serão denominadas por soja-1000 e soja-2000. Estas séries foram obtidas do CEPEA - Centro de Estudos Avançados em Economia Aplicada da ESALQ.

A Figura 4 apresenta as séries e as respectivas séries

originais com uma diferença de ordem 1.

As séries originais da soja não são estacionárias, pois a média é afetada pelo tempo. Aplicam-se o teste de Dickey-Fuller às duas séries e verifica-se que há evidência de que as séries são não estacionárias ( $p\text{-value} = 0,8011$  para a série soja-2000 e  $p\text{-value}=0,3040$  para a série soja-1000). Nota-se que as séries com a primeira diferença são heterocedásticas, ver Figura 4.

**Figura 4** - Séries soja-1000 e soja-2000 e as respectivas séries com a primeira diferença



As funções de auto-correlação parcial dessas duas séries estão apresentadas na Figura 5. Para a série do preço da soja, com a primeira diferença, com 999 observações, ajustam-se um AR(18) incompleto GARCH(1,1) dado por:

$$y_t = 0,257363y_{t-1} - 0,065839y_{t-18} + v_t$$

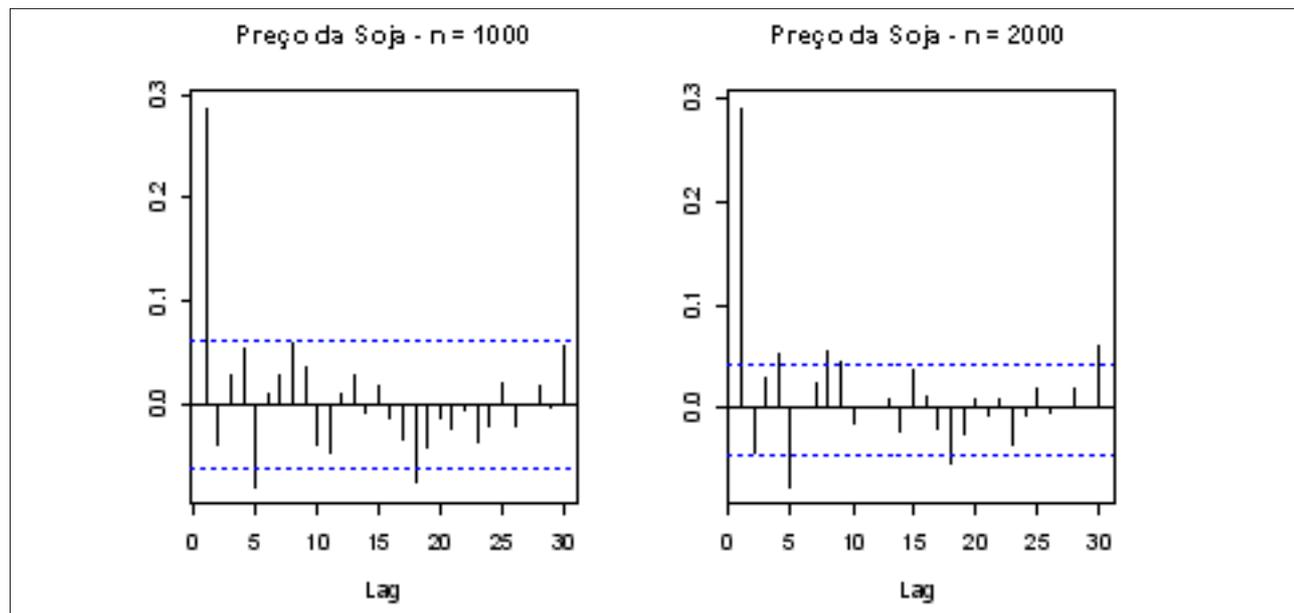
$$\sigma_t^2 = 0,000958 + 0,081936v_{t-1}^2 + 0,917299\sigma_{t-1}^2$$

Para a série do preço da soja, com a primeira diferença, com 1999 observações, ajusta-se um AR(8) incompleto GARCH(1,1) dado por:

$$y_t = 0,270855y_{t-1} + 0,052553y_{t-4} + 0,045338y_{t-8} + v_t$$

$$\sigma_t^2 = 0,00152 + 0,126364v_{t-1}^2 + 0,874073\sigma_{t-1}^2$$

**Figura 5** – Funções de auto-correlação das séries do preço da soja com uma diferença para a série soja-1000 e soja-2000 observações



### 3.1 Construção de intervalos de confiança e de predição para uma RNA utilizada para prever uma série temporal econométrica

Neste trabalho, está-se restrito ao uso de uma rede alimentada adiante com uma única camada oculta, totalmente conectada, algoritmo de minimização de Levenberg-Marquardt e a rede apresenta uma saída.

As redes neurais são modelos não-lineares que podem ser treinadas para mapear valores passados e futuros de séries temporais. No caso da predição de séries heterocedásticas, presentes em grande parte dos modelos econométricos, os modelos utilizando RNAs tornam-se uma ferramenta importante e cada vez mais aplicada (KHIRIPET, 2001).

As séries temporais, utilizadas neste trabalho, apresentam tamanhos 1000 e 2000, no caso da série com 1000 observações utilizam-se 900 valores para treinamento e para série com 2000 observações foram usadas 1900 observações para treinamento. Os 100 valores restantes são utilizados para predição e construção do IP e IC.

Procedendo dessa forma, pode-se testar algumas configurações de RNA. Para realizar a predição, utilizam-se, como configuração de partida, as seguintes topologias de rede RNA: 5-1-1, 5-2-1, 5-3-1, 5-4-1, 5-5-1 e 5-8-1.

Para cada topologia de RNA, mostram-se as respectivas quantidades de valores que ficaram fora do intervalo de confiança (FIC), fora do intervalo de predição (FIP), coeficiente de desigualdade de Theil (TIC) e o RMSE de cada uma delas. Os resultados foram obtidos após a fase de treinamento com um número máximo de 5000 épocas.

Após a obtenção dos valores preditos, são construídos os intervalos de confiança e predição, com 95% de confiança, utilizando-se a metodologia apresentada na seção 1.1.

Na Tabela 2, têm-se os resultados obtidos pelas várias topologias de RNA para todas as séries temporais simuladas. Verifica-se que, para todas as séries, os intervalos de predição estão de acordo com o esperado para o nível de significância adotado.

Note-se que para as séries 4, 5 e 7 que possuem os maiores parâmetros estimados associados à estacionariedade na média ( $\phi_1$ ) ou à volatilidade ( $\beta_1$ ), ver Tabela 1, as topologias de rede que apresentaram os menores RMSE foram 5-1-1 e 5-2-1. A série 2 que apresenta o maior parâmetro estimado associado à volatilidade e o menor parâmetro estimado associado a estacionariedade na média teve como rede de menor RMSE a topologia 5-5-1. As demais séries tiveram os menores RMSE para as topologias 5-5-1, 5-8-1, 5-3-1 e 5-4-1.

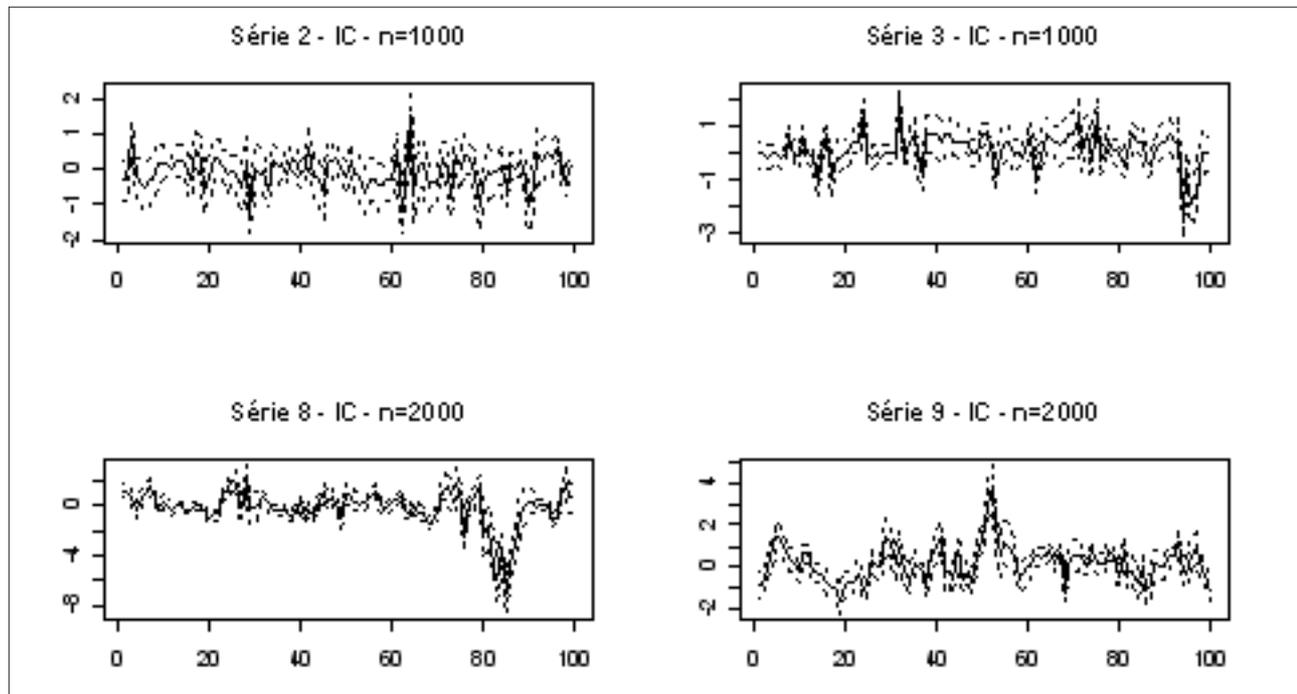
**Tabela-2** Estatísticas para diferentes topologias de RNA para as séries simuladas

Serie_1 Tamanho 1000					Serie_2 Tamanho 1000				
Topologia	FIC	FIP	TIC	RMSE	Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	39	<b>0</b>	0,301879	0,803311	RNA 5-1-1	61	3	0,785141	1,060820
RNA 5-2-1	38	1	0,399675	1,427120	RNA 5-2-1	61	<b>0</b>	0,723386	0,959579
RNA 5-3-1	34	1	0,334592	0,875776	RNA 5-3-1	51	2	0,684038	0,943482
RNA 5-4-1	29	1	0,386736	1,448970	RNA 5-4-1	49	<b>0</b>	0,543451	0,809758
RNA 5-5-1	27	<b>0</b>	0,256966	<b>0,711886</b>	RNA 5-5-1	47	<b>0</b>	<b>0,485354</b>	<b>0,756618</b>
RNA 5-8-1	<b>22</b>	<b>0</b>	<b>0,255647</b>	0,718414	RNA 5-8-1	<b>45</b>	<b>0</b>	0,531842	0,802062
Serie_3 Tamanho 1000					Serie_4 Tamanho 1000				
Topologia	FIC	FIP	TIC	RMSE	Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	59	1	0,502663	0,82656	RNA 5-1-1	40	<b>0</b>	<b>0,170818</b>	<b>0,503313</b>
RNA 5-2-1	57	1	0,479186	0,80053	RNA 5-2-1	41	<b>0</b>	0,175167	0,511368
RNA 5-3-1	55	1	0,482115	0,80778	RNA 5-3-1	42	<b>0</b>	0,206802	0,595457
RNA 5-4-1	<b>52</b>	1	0,419940	0,89248	RNA 5-4-1	44	<b>0</b>	0,199870	0,576690
RNA 5-5-1	58	<b>0</b>	0,433090	0,75401	RNA 5-5-1	43	<b>0</b>	0,190966	0,549347
RNA 5-8-1	<b>52</b>	<b>0</b>	<b>0,373570</b>	<b>0,68996</b>	RNA 5-8-1	<b>36</b>	<b>0</b>	0,193525	0,559327
Serie_5 Tamanho 1000					Serie_6 Tamanho 2000				
Topologia	FIC	FIP	TIC	RMSE	Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	49	<b>0</b>	0,350871	1,163410	RNA 5-1-1	<b>57</b>	<b>0</b>	0,574315	1,082940
RNA 5-2-1	49	<b>0</b>	<b>0,325705</b>	<b>1,103810</b>	RNA 5-2-1	<b>57</b>	<b>0</b>	0,578806	1,088040
RNA 5-3-1	47	1	0,373065	1,224880	RNA 5-3-1	<b>57</b>	<b>0</b>	0,518792	<b>1,018410</b>
RNA 5-4-1	<b>46</b>	1	0,560629	3,545520	RNA 5-4-1	<b>57</b>	<b>0</b>	0,549549	1,056170
RNA 5-5-1	<b>46</b>	1	0,334280	1,132640	RNA 5-5-1	<b>57</b>	1	0,575418	1,094320
RNA 5-8-1	47	1	0,342072	1,485300	RNA 5-8-1	58	<b>0</b>	<b>0,515938</b>	1,023570
Serie_7 Tamanho 2000					Serie_8 Tamanho 2000				
Topologia	FIC	FIP	TIC	RMSE	Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	21	<b>0</b>	0,197814	<b>1,064670</b>	RNA 5-1-1	58	<b>0</b>	0,566012	1,250430
RNA 5-2-1	<b>10</b>	1	<b>0,193400</b>	1,091480	RNA 5-2-1	<b>57</b>	<b>0</b>	0,525961	1,191850
RNA 5-3-1	11	1	0,225797	1,205190	RNA 5-3-1	58	<b>0</b>	0,515991	<b>1,177860</b>
RNA 5-4-1	19	1	0,245945	1,282960	RNA 5-4-1	58	<b>0</b>	0,563687	1,249090
RNA 5-5-1	16	1	0,224315	1,194070	RNA 5-5-1	58	<b>0</b>	0,536157	1,209090
RNA 5-8-1	11	1	0,218734	1,344130	RNA 5-8-1	58	<b>0</b>	<b>0,514995</b>	1,178320
Serie_9 Tamanho 2000					Serie_10 Tamanho 2000				
Topologia	FIC	FIP	TIC	RMSE	Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	23	2	0,300660	1,119920	RNA 5-1-1	49	<b>0</b>	0,321003	0,766997
RNA 5-2-1	28	3	0,352499	1,265730	RNA 5-2-1	50	<b>0</b>	0,323810	0,772407
RNA 5-3-1	<b>22</b>	<b>1</b>	<b>0,267687</b>	<b>1,041700</b>	RNA 5-3-1	51	<b>0</b>	0,319560	0,764513
RNA 5-4-1	<b>22</b>	2	0,346810	1,251340	RNA 5-4-1	52	<b>0</b>	<b>0,302390</b>	<b>0,736179</b>
RNA 5-5-1	25	2	0,314559	1,170280	RNA 5-5-1	<b>46</b>	1	0,312437	0,752868
RNA 5-8-1	30	2	0,293214	1,216080	RNA 5-8-1	52	1	0,322271	0,773927

Como exemplo, apresentamos na Figura 6 os intervalos com 95% de confiança para as séries 2, 3, 8 e 9 geradas por simulação. As linhas tracejadas

indicam os limites superiores e inferiores associados ao intervalo de confiança e a linha contínua representa os valores preditos pela RNA.

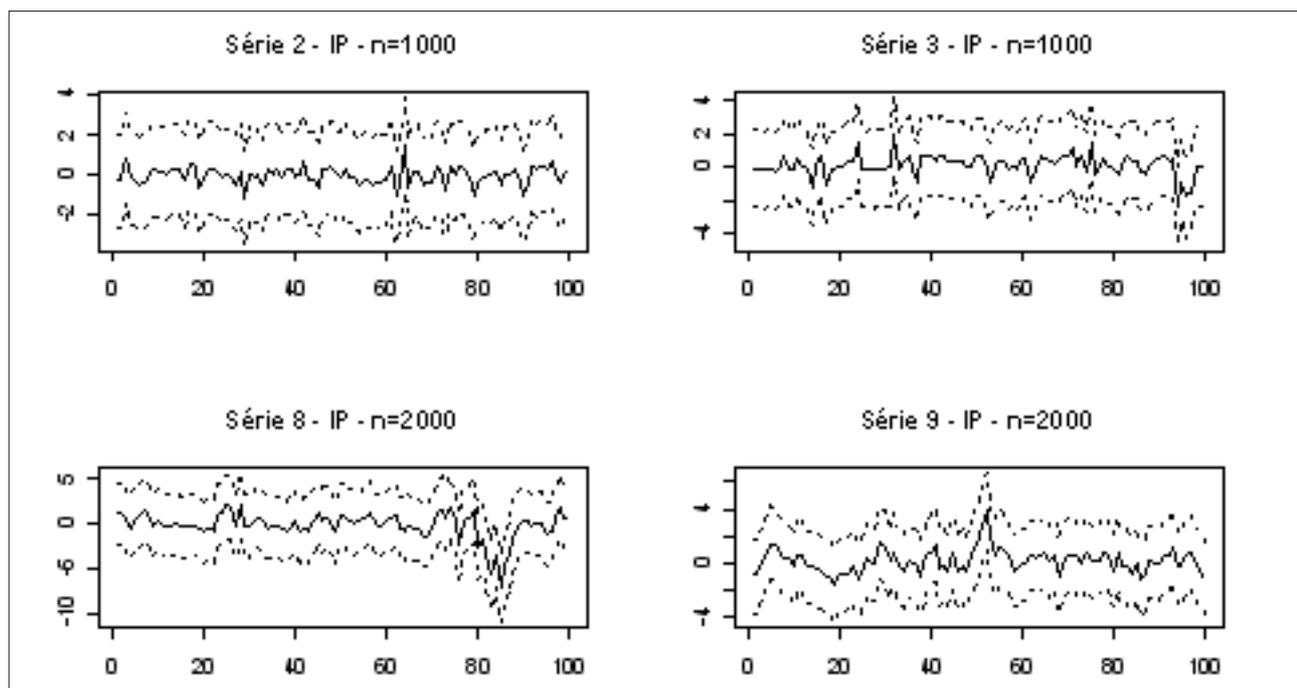
**Figura 6** – Intervalos de confiança para as séries 2, 3, 8 e 9 geradas por simulação



Na Figura 7, apresentam-se os intervalos de predição relacionados às séries 2, 3, 8 e 9

apresentadas na Tabela 1.

**Figura 7** – Intervalos de predição para as séries 2, 3, 8 e 9 geradas por simulação



A Tabela 3 apresenta as estatísticas FIC, FIP, TIC e RMSE obtidas por meio da previsão dos

últimos 100 valores da série da soja com 1000 observações.

**Tabela 3** - Estatísticas para diferentes topologias de RNA para a série do preço da soja (n = 1000)

Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	44	2	0,536179	0,191219
RNA 5-2-1	46	2	0,565791	0,199080
RNA 5-3-1	42	2	0,499710	0,182794
RNA 5-4-1	35	2	0,506254	0,184784
RNA 5-5-1	37	2	<b>0,496727</b>	<b>0,182317</b>
RNA 5-8-1	43	2	0,533281	0,190789

Com os resultados obtidos para a série do preço da soja de tamanho 1000, verifica-se que à medida que se aumenta a quantidade de neurônios na camada intermediária até a topologia 5-5-1, ocorre uma melhora da previsão da rede, visto que o RMSE e o TIC decaem gradativamente. Note-se que a RNA 5-4-1 foi a topologia que apresentou a

menor quantidade de valores fora do intervalo de confiança (FIC) e que todas as topologias tiveram 2 valores fora do intervalo de predição (FIP).

Considerando as 2000 observações da série do preço da soja, obtêm-se os resultados apresentados na Tabela 4.

**Tabela 4** - Estatísticas para diferentes topologias de RNA para a série do preço da soja (n = 2000)

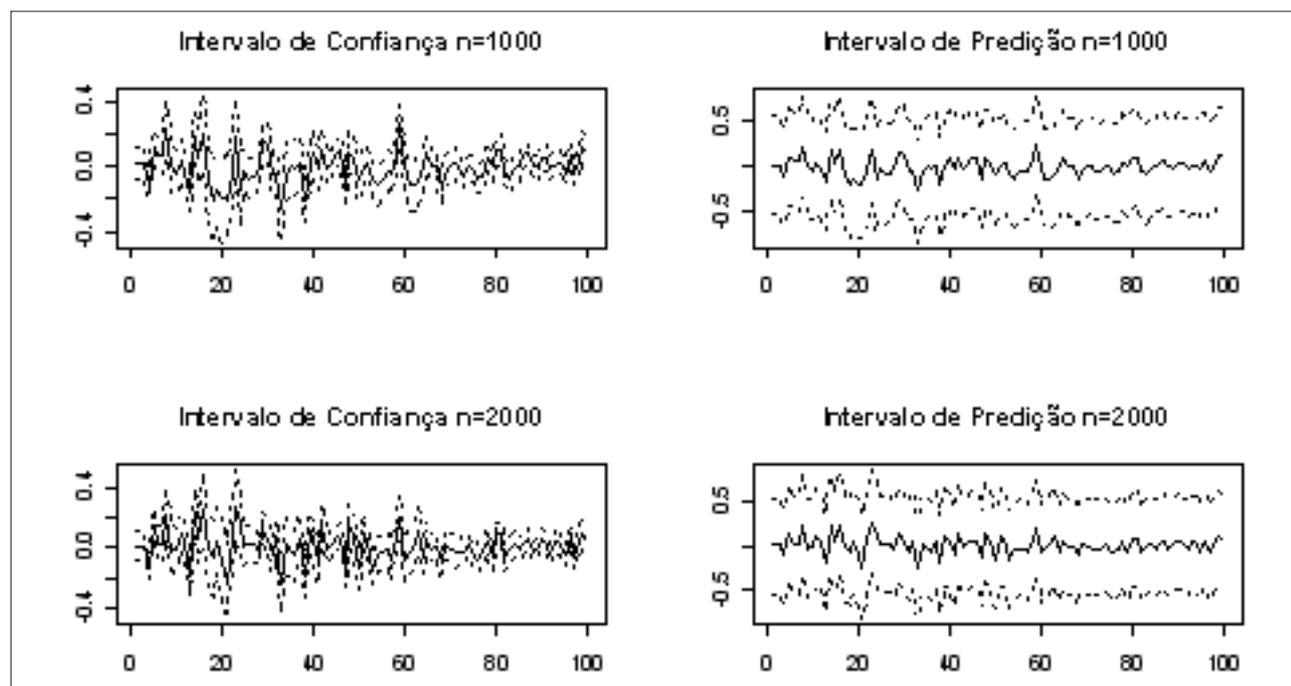
Topologia	FIC	FIP	TIC	RMSE
RNA 5-1-1	43	2	0,516048	0,185952
RNA 5-2-1	43	2	0,530198	0,189386
RNA 5-3-1	45	2	0,504771	0,183038
RNA 5-4-1	44	2	0,531320	0,189688
RNA 5-5-1	31	2	<b>0,490997</b>	<b>0,180060</b>
RNA 5-8-1	54	2	0,535707	0,199795

Dos resultados acima, podem-se verificar que, da mesma forma que a série com 1000 observações, a rede alimentada adiante com três camadas 5-5-1, também foi a topologia que apresentou os melhores resultados. Nesse caso,

o RMSE, o TIC e o FIC decaem gradativamente.

A Figura 8 apresenta os intervalos de confiança e de predição com 95 % de confiança para as séries do preço da soja com 1000 e 2000 observações.

**Figura 8** – Intervalos de confiança e de predição para a série do preço da soja



Neste trabalho, focou-se a previsão de um passo à frente. O principal motivo para isso é que as previsões de um passo à frente são mais utilizadas na prática, uma vez que se pode recalibrar o modelo e gerar novas previsões à medida que novos dados se tornam disponíveis.

Além disso, para muitos modelos, os erros das previsões de um passo à frente são independentes uns dos outros. Conseqüentemente, previsões de um passo são melhores para discriminar entre modelos concorrentes. Uma comparação simples de *performance* de métodos e modelos pode ser obtida simplesmente comparando o MSE (erro quadrático médio) ou o RMSE dessas previsões (KAJITANI et al., 2003).

A construção dos intervalos de predição e de confiança é uma adicional para avaliar a adequação do modelo construído. Estender o seu uso para redes neurais é uma maneira de incorporar novas técnicas de avaliar diferentes topologias.

Para obtenção dos resultados, foi utilizado um computador Pentium IV, 80 GB de HD e 640 MB de RAM. A rede 5-8-1, utilizada na série de tamanho 2000, apresentou o maior tempo de processamento de 2.281 segundos.

#### 4 Conclusão

Neste artigo, foi apresentada a metodologia para construção de intervalos de predição e de confiança para redes neurais de múltiplas camadas alimentadas adiante totalmente conectadas por meio da técnica de regressão multivariada. A metodologia foi aplicada às séries AR-GARCH simuladas e à série real do preço da saca da soja de 60 Kg.

Os métodos de estimação dos intervalos de predição e confiança são válidos sob certas suposições que podem não ser satisfeitas na prática.

Para a série do preço da soja, a rede neural alimentada adiante com três camadas 5-5-1 foi a topologia que apresentou os melhores resultados. Com relação às séries simuladas, obtiveram-se as seguintes topologias : 5-1-1, 5-2-1, 5-3-1, 5-4-1, 5-5-1 e 5-8-1.

O método, apresentado neste trabalho, pode ser estendido para aplicações de RNAs com mais de uma camada intermediária, mais de uma saída e diferentes formas de entrada.

## Referências

- BOLLERSLEV, T. Generalised Autoregressive Conditional Heteroscedasticity, **Journal of Econometrics**, n. 31, p. 307-327, 1986.
- CASTELLANO, G., FANELLI, A. M. e MENCAR, C. Deriving prediction intervals for neurofuzzy networks, CILAB – Computational Intelligence Laboratory. **Mathematical and Computer Modelling: An International Journal**, Kastoria, Grécia, v.42, p. 719-726, 2003.
- CHRYSSOLOURIS, G., LEE, M., RAMSEY, A. **Confidence interval prediction for neural network models**, IEEE Trans. Neural Networks n.7 v.1, pag. 229-232, 1996.
- DE VAUX, R. D., SCHUMI, J., SCHWEINSBERG, J.; UNGAR, L. H. Prediction Intervals for Neural Networks via Nonlinear Regression, **Technometrics**, v. 40, n. 4, nov. 1998.
- DYBOWSKI, S. R. Confidence intervals and prediction intervals for feed-forward neural networks, DYBOWSKI, S. R. (Org.). In: **Clinical Applications of Artificial Neural Networks**, 1. ed., Cambridge University Press, p. 298-326, 2001.
- ENDERS, W. **Applied econometric time series**, Hoboken, Wiley, p. 138-184, 2003.
- HAMILTON, J. D. **Time series analysis**, Princeton: Princeton University Press, 1994.
- HWANG, J. T. G. E DING, A. A. Prediction Intervals for Artificial Neural Networks, **Journal of the American Statistical Association**, v. 92, p. 748-757, 1997.
- JOHNSON, R. A. E WICHERN, D. W. **Applied Multivariate Statistical Analysis**, 4. ed., New Jersey, Prentice Hall, 2002. p. 374 - 375
- KAJITANI, Y., KEITH, W. H., MCLEOD, I. **Forecasting Nonlinear Time Series with Feed-Forward Neural Networks: A Case Study of Canadian Lynx Data**, Journal of Forecasting, n. 24, p. 105-117, 2003.
- KHIRIPET, N., **An Architecture for Intelligent Time Series Prediction with Causal Information**, Tese - Georgia: Georgia Institute of Technology, Atlanta, 2001.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**, São Paulo, Edgard Blucher, p.361-396, 2004.
- NETER J., KUTNER, M. H., NACHTSHEIM, C. J. E WASSERMAN, W. **Applied Linear Regression Models**, Chicago, Irwin, 1996. ed. 3.
- RIVALS, I. E PERSONNAZ, L. Construction of confidence intervals for neural networks base don least square estimation, **Neural Networks**, v. 13, p. 463-484, 2000.
- SEBER, G. A.; WILD, C. J. **Nonlinear Regression**, New York: Wiley, 1989.