

**REVISTA DE  
EMPREENDEDORISMO,  
NEGÓCIOS E INOVAÇÃO**

ISSN 2448-3664

**Mauri Aparecido de Oliveira**

Professor de Métodos  
Quantitativos da Escola Paulista  
de Política, Economia e  
Negócios EPPEN/UNIFESP.  
[http://lattes.cnpq.  
br/4663153865292835](http://lattes.cnpq.br/4663153865292835)

**Ricardo Luiz Pereira Bueno**

Professor de Administração  
da Escola Paulista de Política,  
Economia e Negócios EPPEN/  
UNIFESP.  
[http://lattes.cnpq.  
br/3940643168298775](http://lattes.cnpq.br/3940643168298775)

**Alessandra de Avila  
Montini**

Professora de Métodos  
Quantitativos da Faculdade de  
Economia, Administração e  
Contabilidade FEA/USP.  
[http://lattes.cnpq.  
br/8952860161247290](http://lattes.cnpq.br/8952860161247290)

**UNIVERSIDADE FEDERAL DO ABC**  
AVENIDA DOS ESTADOS, 5001  
BAIRRO BANGU, SANTO ANDRÉ - SP.  
CEP 09210-580

E-MAIL: RENEI@UFABC.EDU.BR

**COORDENAÇÃO**  
AGÊNCIA DE INOVAÇÃO INOVAUFABC



**UMA APLICAÇÃO DE REDES  
NEURAIS ARTIFICIAIS  
RECORRENTES (RTRL) E  
PROCESSOS ARIMA-GARCH  
PARA PREDIÇÃO DA SÉRIE DE  
PREÇOS DA SOJA**

***AN APPLICATION OF RECURRENT  
ARTIFICIAL NEURAL NETWORKS (RTRL)  
AND ARIMA-GARCH PROCESSES FOR  
PREDICTING THE SOYBEAN SERIES OF  
PRICES***

**RESUMO**

Neste artigo é realizado um estudo comparativo quanto à eficiência de previsão de séries temporais utilizando processos ARIMA-GARCH e redes neurais artificiais (RNA) treinadas com o algoritmo de aprendizagem recorrente em tempo real (RTRL – *real time recurrent learning*). Como experimento, a série de preços da saca de soja de 60 Kg é usada para realizar a comparação entre as duas técnicas. São realizadas previsões de 1 a 10 passos à frente, estes valores de janelas de previsão foram escolhidos arbitrariamente. Tanto o modelamento usando RNA como os processos ARIMA-GARCH exigem a transformação dos dados da série original. Os resultados das previsões são apresentados em termos dos valores da série no nível, ou seja, tem a mesma forma da série original. Pode-se verificar que as previsões das redes neurais tiveram desempenho superior quando comparadas aos resultados dos modelos econométricos tradicionais.

**Palavras-Chave:** Previsão, Redes Neurais e ARIMA-GARCH

**ABSTRACT**

*This paper describes a comparative study to measure forecasting efficiency between ARIMA-GARCH process and artificial neural networks (ANN) using the real time recurrent learning algorithm (RTRL). An experiment, applying this two techniques, is performed to compare the forecast of the soybean price series. The forecasting window choice is arbitrary, in this work are used 1 to 10 steps ahead. Both methods, ARIMA-GARCH and ANN, requires data transformation of the original series (or level series), but final forecasting results are presented in terms of level series. According to the obtained results, can be verified a superior performance generated by artificial neural networks when compared with the traditional econometrics volatility models.*

**Key Words:** Forecasting, Neural Networks and ARIMA-GARCH

Classificação JEL / JEL Classification: C01

## 1. INTRODUÇÃO

A previsão de séries temporais (sucessões cronológicas ou crônicas) tem sido uma das áreas-chave em Ciências Sociais Aplicadas. Dadas algumas observações passadas do comportamento de alguns sistemas, surge a questão de como fazer previsões sobre o comportamento futuro e quão precisas essas previsões podem ser.

Neste trabalho vamos analisar duas técnicas utilizadas para previsão de séries temporais: os modelos ARIMA-GARCH e as redes neurais treinadas com o algoritmo RTRL. Com relação as redes neurais, existem vários algoritmos disponíveis para treinamento dos pesos sinápticos baseados nos fluxos de entrada e saída dos dados. Um dos algoritmos mais amplamente utilizado é o de aprendizagem recorrente em tempo real (*real time recurrent learning* - RTRL). Este algoritmo é baseado na computação do gradiente de uma medida do erro de saída com relação aos pesos da rede (Williams e Zipser, 1995). Uma das habilidades desse modelo é conseguir realizar complicadas tarefas de reconhecimento de seqüências temporais (Aussem et al., 1995). A utilização de redes neurais tem aumentado significativamente em aplicações econométricas, principalmente por causa de sua natureza quantitativa de lidar com séries temporais não-lineares. As tarefas de previsão, avaliação de risco e decisões de portfólios de médio prazo constituem o centro da análise financeira e não são apenas exercícios acadêmicos. Deseja-se prever de forma mais precisa para que melhores decisões sejam tomadas, tal como comprar ou vender determinados ativos (Mcnelis, 2002).

Do ponto de vista teórico, o processamento de sinais não-lineares (Lapedes e Faber, 1987), incorporação do tempo na rede neural (Elman, 1988) e o modelamento não-linear para predição de séries temporais caóticas (Casdagli, 1989) têm levado a aplicação de redes neurais como ferramenta na tomada de decisão em finanças (Hawley; Johnson; Raina, 1990; Refenes et al., 1997; Taylor,

2000), análise de mercado (Fishman et al., 1991), modelamento não-linear e previsão (Casdagli e Eubank., 1992; Azzof, 1993; Clements e Hendry, 1999). Mais recentemente tem havido a preocupação de comparar e relacionar a tecnologia de redes neurais com a abordagem estatística tradicional (Cheng; Titterington, 1994; Ripley, 1993, 1994, 1996; Hwang e Ding, 1997; Medeiros et al., 2006), sob a perspectiva econométrica (Kuan e White, 1994), de engenharia financeira (Abu-Mostafa et al., 2001) e macroeconômica (Teräsvirta et al., 2005).

A despeito das várias técnicas que existem para análise de séries temporais (Oliveira e Fávero, 2002), neste trabalho concentramos nosso interesse apenas em um dos inúmeros algoritmos de redes neurais que estão sendo propostos e aplicados por várias empresas e pesquisadores.

## 2. METODOLOGIA

Como a nossa questão de pesquisa pretende avaliar a aderência estatística da curva de previsão do modelo neural com os dados históricos da série de preço da soja, pode-se dizer que se trata de um método de procedimento estatístico. Segundo Andrade (1999, p. 26), o método de procedimento estatístico fundamenta-se na utilização da teoria estatística das probabilidades. Suas conclusões apresentam grande probabilidade de serem verdadeiros, embora admitam certa margem de erro. A manipulação estatística permite comprovar as relações dos fenômenos entre si, e obter generalizações sobre sua natureza, ocorrência ou significado.

Segundo Martins (2002, p. 34-39), este trabalho poderá ser classificado conforme as seguintes modalidades de pesquisa:

- Empírico-analíticas: são abordagens que apresentam em comum a utilização de técnicas de coleta, tratamento e análise de dados marcadamente quantitativos. Privilegiam estudos práticos. Suas propostas têm caráter técnico, restaurador e incrementalista.

Têm forte preocupação com a relação causal entre variáveis. A validação da prova científica é buscada através de testes dos instrumentos, graus de significância e sistematização das definições operacionais.

- *Ex post facto*: tipo de investigação empírica na qual o pesquisador não tem controle direto sobre a(s) variável(eis) independente(s), porque suas manifestações já ocorreram, ou porque ela(s) é (são), por sua natureza, não manipulável(eis).

### 3. MODELOS ARCH

O modelo ARCH - Autoregressive Conditional Heteroskedasticity, ou seja de Heterocedasticidade Autoregressiva Condicional foi proposto inicialmente por Engle (1982). Desde então, têm aparecido inúmeros estudos com refinamentos e modificações do modelo ARCH básico e também com aplicações empíricas em taxa de câmbio, mercado de ações e muitos outros tipos de ativos (Bollerslev, Engle e Nelson, 1994).

Considerando uma série temporal  $\varepsilon_t$ , a qual pode ser descrita como:

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{1}$$

sendo que:

$$h_t = Var[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = E[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = E[\eta_t^2 (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)]$$

$\eta_t \sim NID(0;1)$  (2)

como  $E[\eta_t^2] = 1$  resulta que:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \alpha_0 > 0, \quad 0 < \alpha_1 < 1 \tag{3}$$

sendo que a equação (1) descreve o comportamento de  $\varepsilon_t$  condicionado ao conjunto informacional  $I_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  e NID significa normal e identicamente distribuído.

Devido a  $h_t$  depender do período atrasado

$\varepsilon_{t-1}^2$ , a série  $\varepsilon_t$  é então chamada de modelo ARCH de ordem 1. A expressão (3) indica que a variância condicional de  $\varepsilon_t$  é variante no tempo e é importante notar que não há um termo de erro adicional. A equação (2) assume que todas as observações  $\eta_t$  têm as mesmas propriedades distribucionais.

Para analisar as propriedades dos dados do tipo ARCH é conveniente escrever (1) e (3) como:

$$\varepsilon_t = \eta_t (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{1/2}, \quad \eta_t \sim NID(0;1) \tag{4}$$

Uma vez que  $\eta_t$  é um ruído branco e é independente de  $\varepsilon_{t-1}$ , é fácil mostrar que os elementos da seqüência  $\{\varepsilon_t\}$  têm média zero e são não-correlacionados. Tomando a esperança não condicionada de  $\varepsilon_t$  e sabendo que  $E[\eta_t] = 0$ , segue que:

$$E[\varepsilon_t] = E[\eta_t (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{1/2}] = E[\eta_t] E[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{1/2}] = 0 \tag{5}$$

Uma vez que  $E[\eta_t \eta_{t-h}] = 0$ , também tem-se que  $E[\varepsilon_t \varepsilon_{t-h}] = 0$ , para  $h \neq 0$ .

A obtenção da variância não-condicional de  $\varepsilon_t$  é obtida da seguinte forma:

$$E[\varepsilon_t^2] = E[\eta_t^2 (\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)] = E[\eta_t^2] E[(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)] \tag{6}$$

Sendo  $\sigma_\eta^2 = 1$  e a variância não-condicional de  $\varepsilon_t$  idêntica a de  $\varepsilon_{t-1}$  (isto é,  $E[\varepsilon_t^2] = E[\varepsilon_{t-1}^2]$ ) implica que a variância não-condicional de  $\varepsilon_t^2$  será dada por:  $E[\varepsilon_t^2] = \alpha_0 / (1 - \alpha_1)$  (7)

Portanto, a média e a variância não-condicionadas não são afetadas pelos processos dados por (4). De forma similar, pode-se mostrar que a média condicional de  $\varepsilon_t$  é igual a zero. Dado que  $\eta_t$  e  $\varepsilon_t$  são independentes e que  $E[\eta_t] = 0$ , a média condicional de  $\varepsilon_t$  é:

$$E_{t-1}[\varepsilon_t] = E_{t-1}[\eta_t] E_{t-1} \left[ \left( \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \right)^{1/2} \right] = 0$$

(8)

Sabendo que  $E[\eta_t^2] = 1$ , a variância de  $\varepsilon_t$  condicionada ao passado  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$  será:

$$E[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

(9)

A expressão em (3) mostra que valores absolutos grandes (ou pequenos) de  $\varepsilon_t$  são esperados ser seguidos por valores absolutos grandes (ou pequenos), enquanto houver a igualdade  $E[\varepsilon_t \varepsilon_{t-h}] = 0$ , ou seja, a série  $\varepsilon_t$  é não-correlacionada. Conseqüentemente, um modelo ARCH pode descrever uma série temporal com seqüências de dados pontuais que parecem com *outliers*, onde o fato de que estes *outliers* aparecem em *clusters* (aglomerados) é causado pela equação de variância e não pelas autocorrelações no nível da série temporal.

#### 4. MODELOS GARCH

O modelo ARCH generalizado, conhecido como GARCH- Generalized ARCH foi primeiramente proposto por Bollerslev (1986).

Dado que um modelo AR(p)-ARCH(q) pode ser representado por:

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{10}$$

$$\eta_t \sim NID(0;1) \tag{11}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

(12)

com  $\alpha_0 > 0, \alpha_i \geq 0$  ( $i = 1, 2, \dots, q$ ) e  $\phi_p(1) > 0$  e  $\sum_{i=1}^q \alpha_i < 1$ . Para muitas séries temporais financeiras, o valor de q em (12) pode assumir inconvenientemente valores grandes de tal forma que seja necessário estimar muitos parâmetros. Além disso, pode ser inconveniente impor as restrições  $\alpha_i \geq 0$  para todos os i em (12). Portanto, pode ser útil aproximar o polinômio de q-ésima

ordem em (12) por um polinômio de ordem (p,q) similar a um modelo padrão ARMA para uma série  $y_t$ .

O modelo GARCH é então expresso por:

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 - \beta_1 h_{t-1} - \dots - \beta_p h_{t-p}$$

(13)

ou seja, é a equação GARCH(p,q), sendo

que  $\sum_{j=1}^q \alpha_j + \sum_{i=1}^p \beta_i < 1$ . Na prática, o valor

de q em (13) é muito menor que em (12) e é típico encontrar que o modelo GARCH(1,1) produz uma descrição adequada de muitas séries temporais financeiras (Bollerslev, Chou e Kroner, 1992). Expressões explícitas para a curtose e as autocorrelações de uma série  $\varepsilon_t$  gerada pelas equações de (10) a (12) podem ser geradas mostrando que um modelo GARCH(p,q) pode descrever séries temporais com seqüências de grandes observações tanto positivas como negativas.

A função de autocorrelação (FAC) de  $\varepsilon_t^2$  pode ser encontrada aplicando as mesmas técnicas dos modelos ARCH (Bollerslev, 1986). Por exemplo, a equação GARCH(1,1):

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

(14)

pode ser expressa como:

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1) \varepsilon_{t-1}^2 + v_t - \beta_1 v_{t-1}$$

(15)

$$\text{onde } v_t = \varepsilon_t^2 - h_t.$$

Uma vez que a série  $v_t$  é não correlacionada com o seu próprio passado, a equação (15) indica que o modelo GARCH(1,1) implica em que a FAC de  $\varepsilon_t^2$  parece com a FAC de um modelo ARMA(1,1). Sendo que quando  $\alpha_1 = 0$ , o parâmetro  $\beta_1$  em (15) é não determinado. Conseqüentemente, para qualquer processo GARCH(p,q) o valor de q deverá ser no mínimo igual a 1. Uma das motivações para considerar modelos do tipo GARCH é que tais modelos permitem a previsão de volatilidade condicional. Por exemplo, com um modelo GARCH(1,1), pode ser gerada a previsão de j passos à frente para  $h_t$  como (Enders, 2003 p. 137):

$$h_{t+j} = \alpha_0 + \alpha_1 \varepsilon_{t+j-1}^2 + \beta_1 h_{t+j-1}$$

(16)

Tomando a esperança condicionada:

$$E_t[h_{t+j}] = \alpha_0 + \alpha_1 E_t[\varepsilon_{t+j-1}^2] + \beta_1 E_t[h_{t+j-1}]$$

(17)

sabendo que  $E_t[\varepsilon_{t+j}^2] = E_t[h_{t+j}]$  a expressão

(17) torna-se:

$$E_t[h_{t+j}] = \alpha_0 + (\alpha_1 + \beta_1) E_t[h_{t+j-1}]$$

(18)

e dado  $h_{t+1}$  pode-se utilizar (18) para prever todos os valores subseqüentes da variância condicional como:

$$E_t[h_{t+j}] = \alpha_0 [1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \dots + (\alpha_1 + \beta_1)^{j-1}] + (\alpha_1 + \beta_1)^j h_t$$

(19)

Se  $\alpha_1 + \beta_1 < 1$ , a previsão condicionada  $h_{t+j}$  irá convergir para o valor:

$$E_t[h_t] = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

(20)

## 5. ALGORITMO DE APRENDIZAGEM RECORRENTE EM TEMPO REAL (RTRL)

Uma introdução ao assunto de redes neurais recorrentes pode ser obtido em Haykin (2001), para uma leitura inicial sobre a aplicação de redes neurais em previsão de séries financeiras podem ser consultados Azoff (1994) e Gately (1996).

Seja uma rede neural com  $n$  unidades e  $m$  linhas de entrada externas. Fazendo com que  $y(t)$  denote um conjunto de  $n$ -elementos das unidades da rede no tempo  $t$ , e fazendo com que  $x(t)$  denote um conjunto de  $m$ -elementos de sinais externos de entrada para a rede no tempo  $t$ . Os conjuntos  $y(t)$  e  $x(t)$  podem ser concatenados para formar o conjunto de  $(m+n)$ -elementos  $z(t)$ , com  $U$  denotando o

conjunto de índices  $k$  de tal forma que  $z_k$  é a saída de uma unidade na rede e  $I$  o conjunto de índices  $k$  para os quais  $z_k$  é uma entrada externa. Os índices de  $Y$  e  $X$  são escolhidos para corresponder àqueles de  $Z$ , de tal forma que:

$$z_k(t) = \begin{cases} x_k(t) & \text{se } k \in I \\ y_k(t) & \text{se } k \in U \end{cases}$$

(21)

Fazendo  $W$  denotar a matriz de pesos para a rede, com um único peso entre todos os pares de unidades e também de cada linha de entrada à cada unidade. Pela adoção da convenção descrita, pode-se incorporar todos os pesos dentro de uma matriz  $n$ -por- $(m+n)$ . Para permitir que cada unidade tenha um peso para o bias, simplesmente pode-se incluir entre as  $m$  linhas de entrada uma outra entrada para a qual seu valor seja sempre um.

Para prosseguir, adota-se o uso de uma formulação de tempo discreta e assume-se que a rede consista inteiramente de unidades semi-lineares; isto permite estender a abordagem para um tempo contínuo e outras formas de unidade computacional diferenciável (Williams e Zipser, 1989). Fazendo com que:

$$v_k(t+1) = \sum_{l \in U \cup I} w_{kl} z_l(t)$$

(22)

denote a rede de entrada para a  $k$ -ésima unidade no tempo  $t$ , para  $k \in U$ , com a sua saída no próximo passo de tempo sendo:

$$y_k(t+1) = \varphi_k(v_k(t+1))$$

(23)

sendo que  $\varphi_k$  é a unidade da função de ativação. Portanto, o sistema de equações (22) e (23), constituem o sistema dinâmico total da rede, onde os valores  $z_k$  são definidos pela equação (21). É importante notar que a entrada externa no tempo  $t$  não influencia a saída de qualquer unidade até o tempo  $t + 1$ , portanto a rede é um sistema dinâmico discreto. Usualmente, tanto as unidades (ou neurônios) escondidos e de saída terão

funções de ativação não-lineares.

Algumas das unidades em  $U$  são unidades de saída, para as quais um valor desejado é definido. No entanto, pode não haver um valor desejado definido para cada simples entrada da rede. Pode-se derivar um algoritmo de treinamento para esta rede, chamado de *aprendizagem temporal supervisionada*, significando que certos valores das unidades de saída devem igualar-se a valores desejados definidos em tempos específicos. Fazendo  $T(t)$  denotar o conjunto de índices  $k \in U$  para os quais existe um valor desejado especificado  $d_k(t)$  para o qual a saída da  $k$ -ésima unidade deverá igualar-se no tempo  $t$ . Então definindo um conjunto  $e$  de  $n$ -elementos variando no tempo, como:

$$e_k(t) = \begin{cases} d_k(t) - y_k(t) & \text{se } k \in T(t) \\ 0 & \text{caso contrário} \end{cases} \quad (24)$$

Esta formulação permite a possibilidade de que os valores desejados sejam especificados para diferentes unidades em tempos diferentes. O conjunto de unidades consideradas como “visíveis” podem, portanto, ser variantes no tempo.

Agora, fazendo com que:

$$E(t) = \frac{1}{2} \sum_{k \in U} [e_k(t)]^2 \quad (25)$$

denote o erro global da rede no tempo  $t$ . Por enquanto, deve-se assumir está iniciando no tempo  $t_0$  até algum tempo final  $t_1$ . O objetivo será a minimização do erro total, dado por:

$$E_{total}(t) = \sum_{t=t_0+1}^{t_1} E(t) \quad (26)$$

Isto pode ser feito utilizando-se o método do gradiente descendente, ajustando o peso ao longo de  $\nabla W E_{total}(t_0, t+1)$ , conforme mostrado na Figura-1. Uma vez que o erro total é apenas a soma dos erros nos passos individuais de tempo, uma forma para computar este gradiente é através da acumulação dos valores de  $\nabla W E(t)$  para

cada passo de tempo ao longo da trajetória. A mudança de peso global para qualquer peso particular  $w_{ij}$  na rede pode, portanto, ser escrito como:

$$\Delta w_{ij} = \sum_{t=t_0+1}^{t_1} \Delta w_{ij}(t) \quad (27)$$

sendo que:

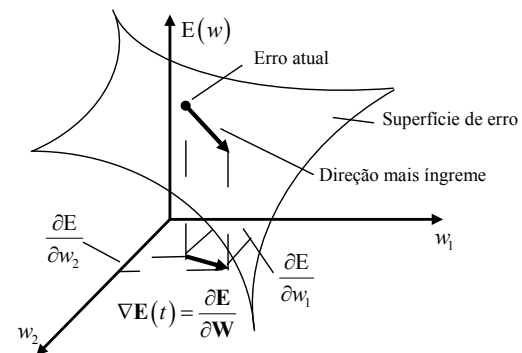
$$\Delta w_{ij}(t) = -\alpha \frac{\partial E(t)}{\partial w_{ij}} \quad (28)$$

e  $\alpha$  é alguma taxa de aprendizagem positiva.

Agora tem-se que:

$$-\alpha \frac{\partial E(t)}{\partial w_{ij}} = \sum_{k \in U} e_k(t) \frac{\partial y_k(t)}{\partial w_{ij}} \quad (29)$$

Figura-1 Representação gráfica do princípio da descida do gradiente (Szilas, 1997)



O ponto principal para entender o algoritmo RTRL é entender o que o fator  $\partial y_k(t) / \partial w_{ij}$  expressa. Ele é essencialmente a medida de sensibilidade do valor de saída da unidade de entrada  $k$  no tempo  $t$  para uma pequena mudança no valor de  $w_{ij}$ , levando em conta que o efeito de uma tal mudança no peso sobre a trajetória de toda a rede de  $t_0$  até  $t$ . É importante notar que  $w_{ij}$  não precisa estar conectado a unidade  $k$ . Portanto, este algoritmo é não local e  $\partial y_k(t) / \partial w_{ij}$  é computado pela diferenciação da rede dinâmica formada pelas equações (22) e (23).

## 6. TESTE PARA NÃO-LINEARIDADE

Na análise de séries heterocedásticas, antes de começar uma procura por uma especificação geral para encontrar um modelo que se ajuste ao seu conjunto particular de dados, é importante testar a presença de não-linearidade. Os testes para verificar não-linearidade que têm sido mais utilizados são o Brock, Dechert, e Scheinkman (1987), o teste de McLeod e Li (1983), um teste desenvolvido por Hsieh (1989) e um teste sugerido por Teräsvirta, Lin e Granger (1993). Neste trabalho vamos considerar o teste de McLeod e Li (1983).

Na estimação de um modelo ARMA, a função de autocorrelação (FAC) pode auxiliar a selecionar os valores de p e q, e o FAC dos resíduos é uma importante ferramenta de diagnóstico. Infelizmente, o FAC tal como é usado em modelos lineares pode levar a falsas conclusões nos modelos não-lineares. A razão é que os coeficientes de autocorrelação medem o grau de associação linear entre  $y_t$  e  $y_{t-i}$ . Dessa forma, a FAC pode falhar na detecção de importantes relações não-lineares presentes nos dados. Tendo interesse nas relações não-lineares dos dados, uma ferramenta de diagnóstico útil é examinar a FAC dos quadrados e dos cubos dos valores de uma série.

O teste de McLeod-Li (1983) visa determinar se existem autocorrelações significantes nos resíduos quadrados de uma equação linear. Para realizar o teste, deve-se estimar a série utilizando o melhor modelo de ajuste linear e chamar os resíduos de  $\hat{\epsilon}_t$ . Tal como em um teste formal para erros ARCH, constrói-se as autocorrelações dos resíduos quadrados. Fazendo  $\rho_i$  denotar o coeficiente de correlação da amostra entre os resíduos  $\hat{\epsilon}_t^2$  e  $\hat{\epsilon}_{t-i}^2$  utiliza-se a estatística de Ljung-Box para determinar se os resíduos quadrados exibem correlação serial. Conseqüentemente, tem-se que:

$$Q = T(T+2) \sum_{i=1}^n \frac{\rho_i}{(T-i)} \quad (30)$$

O valor de Q tem uma distribuição  $\chi^2$  assintótica com n graus de liberdade se a seqüência  $\{\hat{\epsilon}_t^2\}$  é não-correlacionada. Rejeitar a hipótese nula é equivalente a aceitar que o modelo é não-linear. Alternativamente, pode-se estimar a regressão:  $\hat{\epsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\epsilon}_{t-1}^2 + \dots + \alpha_n \hat{\epsilon}_{t-n}^2 + v_t$ .

Se não existir não-linearidade,  $\alpha_1$  até  $\alpha_n$  deverão ser estatisticamente iguais a zero. Com uma amostra de T resíduos, se não houver não-linearidades, a estatística teste  $TR^2$  converge para uma distribuição  $\chi^2$  com n graus de liberdade. Este teste tem um poder substancial para detectar várias formas de não-linearidade. Porém, a forma real da não-linearidade não é especificada pelo teste. Rejeitar a hipótese nula de linearidade não informa a natureza da não-linearidade presente nos dados.

## 7. EXPERIMENTO

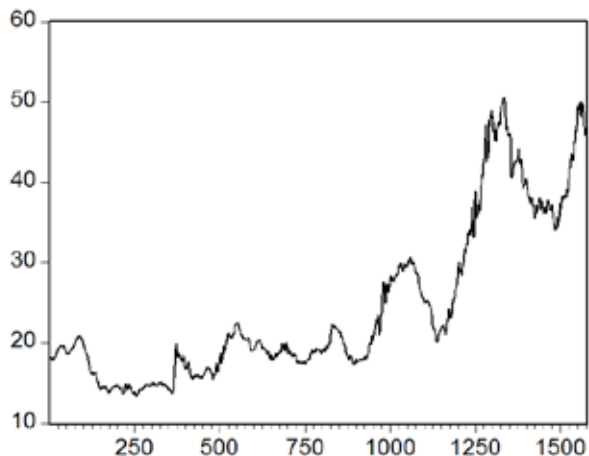
Para realizar a comparação entre as duas metodologias utilizamos a série de preços diários da saca de soja de 60Kg. O procedimento consiste basicamente em realizar previsões estáticas (ou seja, com atualização) utilizando um processo ARIMA-GARCH e transformar os dados previstos para obter os valores da série no nível. No caso da rede neural artificial treinada com o algoritmo de aprendizagem recorrente em tempo real (RNA-RTRL) os valores previstos também foram transformados para ter-se valores no nível da série e dessa forma realizar a comparação entre os métodos.

### 7.1. MODELAMENTO ARIMA-GARCH

O conjunto de dados utilizados é a série de

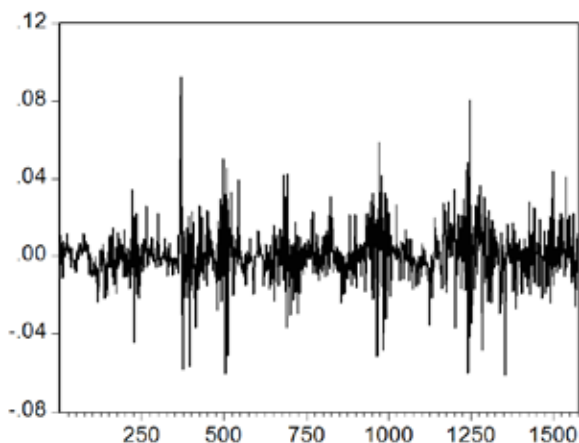
valores de preços diários da saca de 60Kg de soja no período de 29/07/1997 a 28/11/2003, totalizando 1575 valores, na Figura-2 é mostrada a representação de todos os dados.

Figura-2 Série de valores de preços diários da saca de 60Kg da soja.



Fonte: Série histórica indicador CEPEA/ESALQ (R\$ /sc 60 kg)

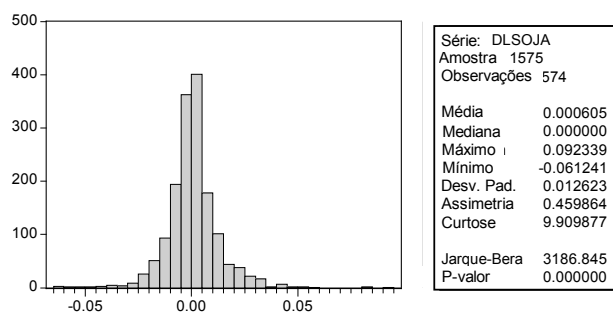
Figura-3 Série de retornos diários da saca de 60Kg da soja



Fonte: Autores

Na Figura-3 tem-se as estatísticas da série de retornos diários da saca de 60Kg de soja, chamada aqui de DLSoja, que utiliza-se para verificar a presença de efeitos ARCH.

Figura-4 Estatísticas dos retornos diários da saca de 60Kg da soja



Fonte: Autores

Dos resultados apresentados na Figura-4, rejeita-se a hipótese de normalidade pelo teste de Jarque-Bera e observa-se excesso de curtose o que indica a necessidade de especificação de um modelo ARCH para a série.

### TESTES DE RAIZ UNITÁRIA PARA A SÉRIE DLSoja

No caso do processo ARIMA-GARCH é necessário criar a série DLSoja (diferença do logaritmo da série). Ou seja,  $DLSoja = \ln(Soja_t) - \ln(Soja_{t-1})$ . Isso é feito porque a logaritmização dos valores observados é uma das transformações mais usuais para estacionarizar séries temporais (ou sucessões cronológicas) e permite tanto a linearização de tendências exponenciais, como a estabilização de variâncias.

Abaixo tem-se o resultado do teste de raiz unitária de Dickey-Fuller aumentado, também chamado de ADF. Este teste foi realizado utilizando o programa EViews.

Teste Estatístico	-15,87229	1% Critical	-3,4374
ADF	Value*	5% Critical	-2,8638
		10% Critical	-2,5680
		Value	

Pode-se concluir a partir dos resultados obtidos no teste ADF que rejeita-se a hipótese de presença de uma raiz unitária na série DLSoja.

O resultado do teste de Dickey-Pantula para uma e duas raízes unitárias é mostrado a seguir.

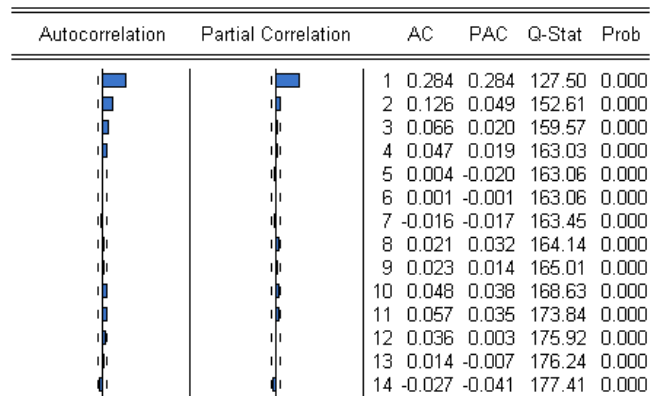
Hipóteses testadas	Estatística	valores críticos a 5%
$H_0 : 2RU$	-30.36328	-3,37
$H_a : 1RU$	-43.07428	-3,77
$H_0 : 1RU$	-38.76627	-3,42
$H_a : 0RU$	-21.56586	-3,80
	-19.13812	-4,16

Pode-se concluir a partir dos resultados obtidos que rejeita-se as hipóteses de



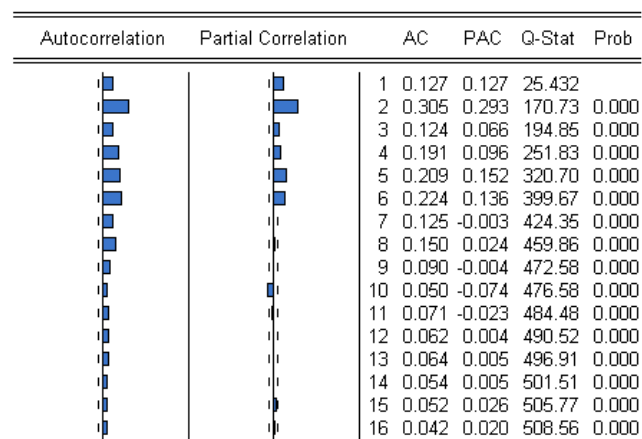
presença de duas e uma raiz unitária pelo teste de Dickey-Pantula (1987).

Figura-5 FAC e FACP da série de retornos diários (DLSOJA) da saca de 60Kg da soja



A Figura-3 representa a série de retornos diários da saca de 60Kg da soja, pode-se verificar na Figura-5 que a função de autocorrelação (Autocorrelation) apresenta comportamento declinante e a função de autocorrelação parcial (Partial Correlation) é truncada na primeira defasagem, isso é indicação de um modelo auto-regressivo de primeira ordem AR([1]).

Figura-6 Correlograma dos resíduos quadrados da série DLSOJA



Especificando um modelo AR([1]) para representar a série de retornos diários do preço da saca de 60Kg da soja, verifica-se que

neste modelo existe correlação dos resíduos quadrados, conforme pode ser verificado pelo correlograma da Figura-6. Nas Figuras 5 e 6 os valores de Prob iguais a zero indicam que pode-se rejeitar a hipótese de que as correlações obtidas sejam iguais a zero.

Resultado do teste ARCH-LM para a série DLSOJA AR([1]), para uma defasagem:

ARCH Test:			
Obs*R-squared	25,36882	Prob	0,000000

Portanto, rejeita-se a hipótese nula “não existe efeitos ARCH” no modelo.

Para esse caso estima-se um modelo AR([1])-GARCH(1,1).

A seguir é apresentada a estimativa dos parâmetros do modelo AR([1])-GARCH(1,1), que pode ser escrito como:

$$DLSOJA_t = \phi_1 DLSOJA_{t-1} + \varepsilon_t, \quad \text{sendo que}$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0; h_t) \quad \text{e a variância é dada por}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}$$

As estatísticas do modelo: critérios de informação (AIC e BIC) e log de verosimilhança poderiam ser utilizados se estivesse selecionando modelos ARIMA-GARCH concorrentes.

### TESTES DE DIAGNÓSTICO

Abaixo são apresentados os teste de Ljung-Box (Q (K) e Q2(K)) e o teste ARCH-LM . Da tabela abaixo pode-se concluir que a série apresenta comportamento não-linear (teste de McLeod-Li), ou seja existe autocorrelação entre os resíduos quadrados.

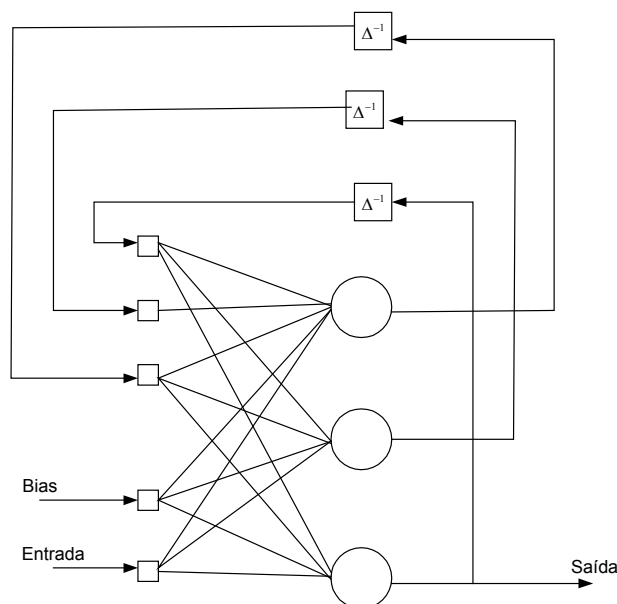
Estatística	AR([1])-GARCH(1,1)	P-value
Q(3)	14,659	0,001
Q(6)	25,956	0,000
Q(24)	47,975	0,002
Q2(3)	0,2489	0,883
Q2(6)	1,4315	0,921
Q2(24)	10,194	0,990
LM	4,533526	0,920087

Parâmetros	Coefficiente	Erro-padrão	Estatística z	Significância
$\phi_1$	0,344720	0,027821	12,39087	0,0000
$\alpha_0$	0,00000333	0,00000106	3,153767	0,0016
$\alpha_1$	0,160294	0,028825	5,561005	0,0000
$\beta$	0,831960	0,028134	29,57108	0,0000
Modelo	Log de Verosimilhança		AIC	BIC
AR([1])-GARCH(1,1)	4955,847		-6,336337	-6,322634

## 7.2. MODELAMENTO UTILIZANDO REDE NEURAL ARTIFICIAL (RNA-RTRL)

A especificação da rede neural apresenta as seguintes características: 2 neurônios na camada escondida, vetor de entrada de 10 valores, matriz inicial de pesos igual a zero, bias igual a 1, função de ativação do tipo sigmóide, taxa de aprendizagem de 0,005 e número de épocas igual a 20000. A representação da rede é mostrada na Figura-7, sendo que  $\Delta^{-1}$  é o operador de atraso unitário aplicado aos dois neurônios da camada escondida e à saída.

Figura-7 Rede recorrente totalmente conectada para formulação do algoritmo RTRL



A rede tem duas camadas distintas: uma camada de entrada conectada com realimentação e uma camada de processamento com nós computacionais (Haykin, 2001). Para realizar o treinamento dos pesos sinápticos da rede foi utilizado um subconjunto de validação de 100 valores. A soma quadrática dos erros é

$$SSE = \frac{1}{2} \sum_{t=0}^{T-1} \sum_{k=0}^{U-1} e_k^2(t)$$

dada por:  $\dots$ , sendo que T é o número total de dados da amostra, no caso do subconjunto de validação T=100, U é o número de nós da rede recorrente, nesse caso U=3.

## 8. RESULTADOS DAS PREVISÕES DOS MODELOS ARIMA-GARCH E RNA-RTRL

Para comparar as previsões do valor do preço da saca de soja utilizamos as estatísticas MAPE e TIC.

Erro Percentual Absoluto Médio (Mean Absolute Percentage Error - MAPE):

$$MAPE = \sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| / h$$

Coefficiente de Desigualdade de Theil (Theil Inequality Coefficient - TIC):

$$TIC = \frac{\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t^2 / h} + \sqrt{\sum_{t=T+1}^{T+h} y_t^2 / h}}$$

O coeficiente de desigualdade de Theil (TIC) sempre estará entre zero e um, sendo que zero indica um ajuste perfeito.

O critério MAPE é invariante com relação a escala, quanto menor este critério melhor o modelo. No caso do processo ARIMA-GARCH foi necessário criar a série DLSOJA (diferença do logaritmo da série). No entanto, em geral na maioria dos casos está-se mais interessado nas previsões das observações em termos da escala inicial, ou seja no nível, do que nos valores dos logaritmos.

Pode-se pensar que a questão é resolvida facilmente calculando o anti-logaritmo das previsões dos valores logaritmizados. Mas, procedendo-se deste modo, as previsões construídas não vão minimizar o erro quadrático médio. Com efeito, tem-se:

$$X_t(h) = E[X_{t+h} | X_t, X_{t-1}, \dots] \neq \exp[Y_t(h)] = \exp[E\{Y_{t+h} | Y_t, Y_{t-1}, \dots\}]$$

sendo que  $Y_t = \ln X_t$ . Admitindo-se que os  $\{\varepsilon_t\}$  da série logaritmizada têm distribuição Normal, o processo  $Y_t$  também é e, conseqüentemente,  $\{X_t\}$  tem para cada t distribuição Log-Normal. Atendendo às propriedades desta distribuição, pode-se provar (Nelson, 1973) que a previsão com horizonte h da série

original está relacionada com a previsão da série transformada do seguinte modo:

$$X_t(h) = \exp \left[ Y_t(h) + \frac{1}{2} Var \{ \varepsilon_t(h) \} \right] \text{ sendo que}$$

$\varepsilon_t(h)$  representa o erro de previsão em h passos da série logaritmizada. Para encontrar os limites do intervalo de confiança na previsão de  $X_{t+h}$  é suficiente calcular o anti-logaritmo dos correspondentes limites de confiança na previsão de  $Y_{t+h}$ , para analisar o que se passa com outras transformações pode-se consultar Pankratz (1991) e Pankratz e Dudley (1987).

Portanto, a partir do valor previsto da série DLSOJA, para obter-se o valor da série no nível realiza-se a seguinte transformação:

$$\ln(SOJA_t) - \ln(SOJA_{t-1}) = \text{Previsão\_ARIMA\_GARCH}$$

$$\ln(SOJA_t) = \text{Previsão\_ARIMA\_GARCH} + \ln(SOJA_{t-1})$$

$$SOJA_t = \exp \left\{ \text{Previsão\_ARIMA\_GARCH} + \ln(SOJA_{t-1}) + \frac{1}{2} Var \{ \varepsilon_t(h) \} \right\}$$

Sendo que  $SOJA_t$  indica o preço da saca de soja de 60Kg no instante t.

No caso dos valores previstos pela rede neural tem-se que os valores para a série no nível são obtidos da seguinte forma:

NDSLJOJA é a normalização que foi realizada na série DLSOJA para que esta série

ficasse no intervalo [0,1], isso é necessário devido a utilização da função sigmóide.

$$NDSLJOJA_t = \frac{DLSOJA_t - \min \{ DLSOJA_t \}}{\max \{ DLSOJA_t \} - \min \{ DLSOJA_t \}}$$

$$DLSOJA_t = NDSLJOJA_t [\max \{ DLSOJA_t \} - \min \{ DLSOJA_t \}] + \min \{ DLSOJA_t \}$$

$$\ln(SOJA_t) - \ln(SOJA_{t-1}) = NDSLJOJA_t [\max \{ DLSOJA_t \} - \min \{ DLSOJA_t \}] + \min \{ DLSOJA_t \}$$

$$\ln(SOJA_t) = NDSLJOJA_t [\max \{ DLSOJA_t \} - \min \{ DLSOJA_t \}] + \min \{ DLSOJA_t \} + \ln(SOJA_{t-1})$$

$$SOJA_t = \exp \{ NDSLJOJA_t [\max \{ DLSOJA_t \} - \min \{ DLSOJA_t \}] + \min \{ DLSOJA_t \} + \ln(SOJA_{t-1}) \}$$

Sendo que  $\max \{ DLSOJA_t \}$  e  $\min \{ DLSOJA_t \}$  representam os valores máximo e mínimo da série DLSOJA, respectivamente. Na figura abaixo é mostrado o resultado obtido pela rede neural e pelo modelamento ARIMA-GARCH da série do preço da saca de 60Kg soja.

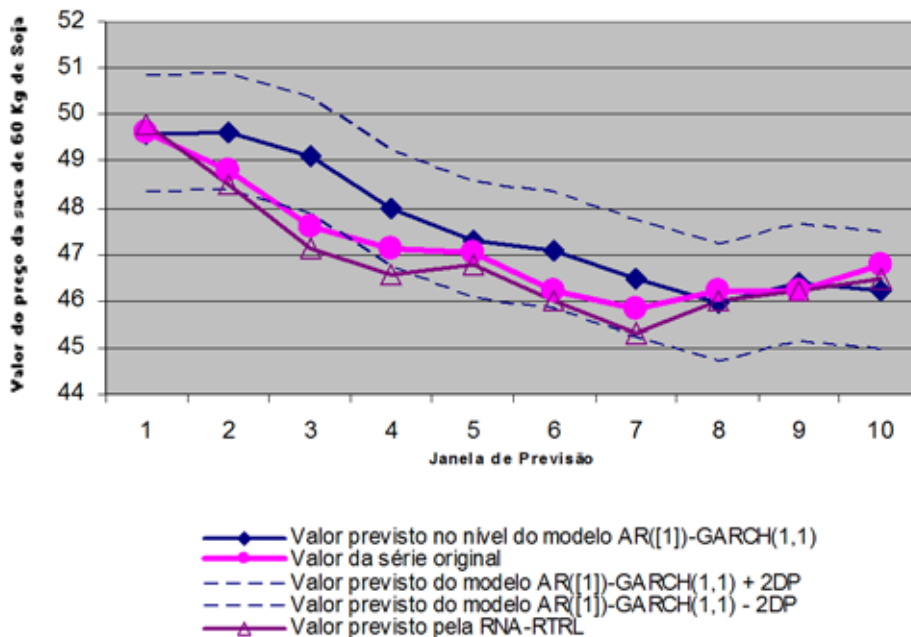
As estatísticas MAPE e TIC obtidas pela RNA e pelo modelo AR([1])-GARCH(1,1) para a série da saca de 60Kg de soja no nível são mostradas a seguir.

**Estatísticas de Previsão – Série do preço da saca de 60Kg de Soja**

	ARIMA-GARCH	Rede Neural	
TIC	0,007775	TIC	0,003610
MAPE	0,012857	MAPE	0,006435

No caso da série do preço da soja a rede neural 10-2R-1 forneceu uma previsão mais precisa que o modelo AR([1])-GARCH(1,1).

Figura-8 Valores da série do preço da saca de 60 Kg de soja no nível e a curva de previsão do modelo ARIMA-GARCH e da rede neural artificial RNA-RTRL.



## 9. CONCLUSÃO

Séries temporais de ciclos econômicos, preços de ações, crescimento econômico, taxas de câmbio e curva de Phillips são não-lineares ou caóticas. Portanto, é necessário que métodos não-lineares sejam utilizados nas análises dessas séries. Por causa da sua capacidade de aprender, tem crescido a quantidade de aplicações das redes neurais para a estimação de parâmetros. Pudemos verificar nesse trabalho que uma rede neural treinada com o algoritmo de aprendizagem recorrente em tempo real (RTRL) apresentou melhores resultados de previsão quando comparados com aqueles obtidos pelo modelamento econométrico tradicional utilizando um processo ARIMA-GARCH. Uma diferença importante que deve ser destacada entre a RNA-RTRL e o modelo ARIMA-GARCH é a capacidade da RNA capturar padrões não-lineares subjacentes ao comportamento da série, enquanto que isso não acontece na modelagem ARIMA-GARCH. No entanto, comparando com as técnicas econométricas tradicionais, como a modelagem ARIMA-GARCH, fica clara a falta de poder estatístico nos valores obtidos pela RNA. Esta fraqueza das redes neurais, no que diz respeito à falta de procedimentos estabelecidos para realizar testes de significância para as variáveis é um ponto que tem sido explorado por muitos autores, alguns trabalhos que sugerimos são: (Refenes et al., 1997), (Hastie et al., 2001) e (Vapnik, 1999).

## 10. REFERÊNCIAS

ABU-MOSTAFA, Y. S.; ATIYA, A. F.; MAGDON-ISMAIL, M.; WHITE, H. Introduction to the special issue on neural networks in financial engineering. *IEEE Transactions on Neural Networks* 12: 653-655, 2001.

ANDRADE, M. M. Como Preparar Trabalhos para Cursos de Pós-Graduação. São Paulo: Atlas, 1999.

AUSSEM, A.; MURTAGH, F.; SARAZIN, M. Dynamical recurrent neural networks – towards environmental time series prediction, Cogprints, Reino Unido, 1995. Disponível em: <http://cogprints.org/perl/oai2>. Acesso em: 10 jan 2007.

AZOFF, E. M. Reducing Error in Neural Network Time Series Forecasting. *Neural Computing and Applications*, vol. 1, pp. 240-247, 1993.

AZOFF, E. M. Neural network time series forecasting of financial markets. Hoboken: John Wiley & Sons Ltd., 1994.

BOLLERSLEV, T. Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, n. 31, p. 307-327, 1986.

BOLLERSLEV, T.; CHOU, R. e KRONER, K. ARCH modeling in finance: a review of the theory and empirical evidence, *Journal of Econometrics*, n. 52, p. 5-59, 1992.

BOLLERSLEV, T.; ENGLE, R. F.; E NELSON, D. B. ARCH models, *Handbook of Econometrics*, n. 63, p. 37-60, 1994.

BROCK, W. A.; DECHERT, W. D.; SCHEINKMAN, J. A. A Test for Independence Based on the Correlation Dimension. Work. pap., Department of Economics, University of Wisconsin at Madison, University of Houston, and University of Chicago, 1987.

CASDAGLI, M. Nonlinear Prediction of chaotic time series. *Physica D*, vol. 35, pp. 335-356, 1989.

CASDAGLI, M.; EUBANK, S. Nonlinear Modeling and Forecasting, Santa Fe Institute Proceedings vol. XII, Addison-Wesley, 1992.

CHENG, B.; TITTERINGTON, D. M., Neural Networks: A Preview from Statistical Perspective. *Statistical Science*, 1994.

CLEMENTS, M. P.; HENDRY, D. F. Forecasting Non-stationary Economic Time Series. MIT Press: Cambridge, MA, 1999.

DICKEY, D. e PANTULA, S. Determining the order of Differencing in Autoregressive Processes. *Journal of Business and Economic Statistics*, n.15, p. 455-461, 1987.

ELMAN, J. L. Finding structure in time (CRL Technical Report 8801). La Jolla: University of California, San Diego, Center for Research in Language, 1988.

ENDERS, W. Applied econometric time series. Hoboken: John Wiley & Sons, 2003.

ENGLE, R. F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, n. 50, p. 987-1007, Julho 1982.

- FISHMAN, M. B.; BARR, D. S.; LOICK, W. J. Using Neural Nets in Market Analysis. *Technical Analysis of Stocks and Commodities*, vol. 9, no. 4, 1991.
- GATELY, E. *Neural Networks for financial forecasting*. Wiley, 1996.
- HASTIE, T.; TIBSHIRANI, R.; e FRIEDMAN, J. *The elements of statistical learning*, New York, Springer-Verlag, 2001.
- HAYKIN, S. *Redes neurais: princípios e prática*. Porto Alegre: Bookman, 2001.
- HAWLEY, D. D.; JOHNSON, J. D. e RAINA, D. *Artificial Neural Systems: A New Tool for Financial Decision-Making*. *Financial Analysis Journal*, 1990.
- HSIEH, D. A. Testing for nonlinear dependence in daily foreign exchange rates, *Journal of Business*, v. 3, n. 62, p. 339–368, 1989.
- HWANG, J.T.G. e DING, A. A. Prediction intervals for artificial neural networks. *Journal of the American Statistical Association* 92: 109–125, 1997.
- KUAN, C. M. e WHITE, H. *Artificial Neural Networks: An Econometric Perspective*. *Econometric Reviews* 13, pp. 1-91, 1994.
- LAPEDES, A. e FARBER, R. *Nonlinear Signal Processing Using Neural Networks, Prediction and System Modelling*. Los Alamos Report LA-UR-87-2662, Los Alamos National Laboratory, 1987.
- MARTINS, G. A. *Manual para Elaboração de Monografias e Dissertações*. São Paulo: Atlas, 2002.
- MCLEOD, A. e LI, W. Diagnostic checking of ARMA time series models using squared residuals autocorrelations. *Journal of Time Series Analysis*, n. 4, p. 269–273, 1983.
- MCNELIS, P. D. *Financial forecasting with neural networks*. Department of Economics, Georgetown University, Outubro 2002.
- MEDEIROS M., TERÄSVIRTA, T. e RECH, G. *Building Neural Network Models for Time Series: A Statistical Approach*, *Journal of Forecasting* 25, pp. 49–75, 2006.
- NELSON, C. R. *Applied Time Series Analysis for Managerial Forecast*, Holden-Day, San Francisco, 1973
- OLIVEIRA, M. A. e FÁVERO, L. P. L. *Uma Breve Descrição de algumas Técnicas para Análise de Séries Temporais: Séries de Fourier, Wavelets, Arima, Modelos Estruturais para Séries de Tempo e Redes Neurais*, VI SEMINÁRIOS EM ADMINISTRAÇÃO – FEA/USP, 2002.
- PANKRATZ, A. *Forecasting with Dynamic Regression Models*, Wiley, Nova Iorque, 1991.
- PANKRATZ, A. e DUDLEY, U. *Forecasts of Power-Transformed Series*, *Journal of Forecasting*, 6, 239-248, 1987.
- REFENES, A. N., BURGESS, A. N. e BENTZ, Y. *Neural networks in financial engineering: a study in methodology*, *IEEE Transactions on Neural Networks*, v. 8, n. 6, Nov. 1997.
- RIPLEY, B. D., *Statistical Aspects of Neural Networks*. Chapman e Hall, 1993.
- RIPLEY, B. D., *Neural Networks and Related Methods of Classification*, *Journal of the Royal Statistical Society, Series B*, 56, 3, pp. 409-456, 1994.
- RIPLEY, B. D., *Pattern Recognition and Neural Networks*. Cambridge Press, 1996.
- SZILAS, N. *Les réseaux récurrents supervisés: une revue critique*, fev. 1997.
- TAYLOR, J. W. *A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns*, Saïd Business School, University of Oxford, *Journal of Forecasting*, Vol. 19, pp. 299-311, 2000.
- TERÄSVIRTA, T. e GRANGER, C. *Modeling Nonlinear Economic Relationships*. Oxford University Press, 1993.
- TERÄSVIRTA, T., LIN, C.-F. e GRANGER C.W.J. *Power of the Neural Network Linearity Test*, *Journal of Time Series Analysis*, 14: 209-220, 1993.
- TERÄSVIRTA, T., VAN DIJK, D. e MEDEIROS M. *Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination*. *International Journal of Forecasting*, 2005.
- VAPNIK, V. N. *Statistical learning theory*, New York, Wiley, 1999.
- WILLIAMS, R. J. e ZIPSER, D. *A learning algorithm for continually running fully recurrent neural networks*. *Neural Computation*, vol. 1, p. 270-280, 1989.
- WILLIAMS, R. J. e ZIPSER, D. *Gradient-based learning algorithms for recurrent networks and their computational complexity*, 1995.