

PREVISÃO DE PREÇO FUTURO DO BOI GORDO NA BM&F: UMA COMPARAÇÃO ENTRE MODELOS DE SÉRIES TEMPORAIS E REDES NEURAIAS

Live cattle prices forecast at BM&F: a comparison between time series models and neural networks

Luiz Eduardo Gaio¹, Luiz Gonzaga de Castro Júnior², André Ribeiro de Oliveira³

RESUMO

A capacidade do homem de prever o futuro ainda é muito limitada. Apesar do imenso esforço de especialistas das mais diferentes áreas durante anos de desenvolvimento do conhecimento científico, as previsões sobre os mais variados eventos, como as condições climáticas num determinado tempo, a evolução do preço de uma *commodity* no futuro, continuam sujeitas a um grau de erro razoavelmente elevado. Assim, objetivou-se com este trabalho comparar modelos de previsão de preço, para o mercado de boi gordo na Bolsa de Mercadorias & Futuros (BM&F), utilizando modelos baseados em redes neurais e ferramentas estatísticas de modelagem de séries heteroscedásticas. As séries utilizadas correspondem aos fechamentos dos preços do boi gordo, no período de 01 de agosto de 1997 a 27 de maio de 2005, num total de 1943 observações. Os resultados evidenciaram a supremacia dos modelos baseados em redes neurais comparados com o modelo AR-EGARCH, uma vez que os valores de Erro Quadrático Médio e Raiz do Erro Quadrático Médio das previsões foram inferiores para as redes neurais.

Palavras-chave: previsão de preço, mercado futuro, redes neurais.

ABSTRACT

Human forecasting capacity is still very limited. In spite of the extreme efforts of specialists in several different areas for years developing scientific knowledge, forecasting various events, such as climatic conditions at a given time, the evolution of a commodity price in the future, remain subject liable to a considerably high degree of error. Therefore, this paper aims to compare forecast price models for the Live Cattle market at Brazilian Mercantile and Future Exchange (BM&F) using models based in Neural Networks and statistical tools of heteroscedastic times series. The data used correspond to the closing of the live cattle prices, in the period ranging from August 1997 to May 2005, totalizing 1946 observations. The results show the supremacy of neural networks models compared with the AR-EGARCH model, once the Mean Squared Error and the Mean Squared Error Root forecasted were smaller for the neural networks.

Key words: price porecast, future market, neural networks.

1 INTRODUÇÃO

A previsão econômica sempre foi um importante fator para a análise de mercado, reduzindo as incertezas na tomada de decisão por parte dos agentes envolvidos. Efetuar uma boa previsão não é um trabalho fácil, visto que os mercados estão expostos a distúrbios durante todo o tempo, que alteram o seu comportamento. Tais distúrbios não podem ser captados pelos modelos mais robustos de previsão de preço encontrados hoje na literatura acadêmica, devido ao seu grau de aleatoriedade.

A utilização de instrumentos que auxiliam a tomada de decisão dos membros da cadeia produtiva do boi gordo (produtores, compradores e investidores em geral), proporcionando-lhes informações confiáveis, é um importante fator para manter um maior controle sobre a

decisão, minimizando riscos de decisões erradas, que geram perdas futuras.

Conforme Bressan & Lima (2002), com o intuito de se administrar risco de mercado (oscilação de preço), o mercado futuro permite a permuta destes riscos com especuladores, apresentando-se como importante instrumento de segurança e sinalização de preços para todos os participantes da cadeia agroindustrial. Hull (1996), citado por Bressan & Lima (2002), sintetiza essa relação, definindo os contratos futuros como compromissos de compra ou venda de um determinado ativo numa data pré-estabelecida, a um preço (cotação) que reflete as forças de oferta e demanda que atuam naquele momento.

No Brasil, os estudos sobre o mercado futuro de boi gordo têm utilizado, principalmente, o enfoque da

¹Graduado em Administração pela Universidade Federal de Lavras/UFLA. Foi bolsista pelo PET/SESu e pesquisador pelo Centro de Inteligência em Mercados – Rua das Palmeiras, 471 – Jardim Eldorado – 37.200-000-Lavras, MG – lugaio@yahoo.com.br

²Doutor em Economia pela ESALq/USP, Professor do Departamento de Administração e Economia da Universidade Federal de Lavras/UFLA e coordenador do Centro de Inteligência em Mercados – Cx. P. 3037 – 37200-000 – Lavras/MG – lgcastro@ufla.br

³Mestrando em Administração pela Universidade Federal de Lavras/UFLA e pesquisador pelo Centro de Inteligência em Mercados – Cx. P. 3037 – 37200-000 – Lavras, MG – andre180282@yahoo.com.br

Recebido em 19/03/07 e aprovado em 12/07/07

efetividade do *hedge*, comportamento da base, análise da volatilidade e análise de causalidade. Porém, pesquisas voltadas para a previsão de preço futuro não são comumente realizadas, como é o caso de Bressan & Lima (2002). Por isso, focou-se este trabalho em testes de modelos de previsão de preço do boi gordo, com o intuito de preencher parte da lacuna existente sobre este assunto, sendo uma pesquisa de caráter empírico.

Assim, neste estudo objetivou-se efetuar uma comparação entre modelos de previsão de preços aplicados aos contratos futuros de boi gordo na Bolsa de Mercadorias & Futuros (BM&F), de forma a fornecer um indicativo robusto para análise e projeção do referido mercado. Os modelos aqui utilizados para comparação são os estatísticos baseados em séries temporais, como os modelos ARMA-GARCH, ou seja, os modelos ARMA (autorregressivos – médias móveis), aplicados à média e GARCH (heteroscedasticidade condicional autorregressiva generalizada) aplicados à variância, e os modelos baseados em Redes Neurais Artificiais.

2 REFERENCIAL TEÓRICO

2.1 Modelos de Redes Neurais

As redes neurais artificiais nada mais são do que o funcionamento de um neurônio biológico na sua forma artificial, implementadas por meio de componentes eletrônicos ou através de *software* de computadores. Segundo Freitas (2001) o funcionamento de uma rede neural baseia-se na simulação do funcionamento do cérebro humano, em que o conhecimento para a solução de determinado problema é obtido através de um processo de aprendizagem. Com o propósito de desenvolver o mesmo tipo de estrutura para um modelo computacional de comportamento inteligente, neurologistas e pesquisadores da inteligência artificial propuseram uma rede altamente interconectada de nódulos (ou neurônios).

O cérebro humano, com seus bilhões de neurônios interligados em rede, é bem distinto de um computador. As experiências e sensações, que são observadas através dos sentidos humanos, geram sinais elétricos que passam pelos neurônios criando elos de ligação entre eles. Estas ligações entre neurônios são novamente sensibilizadas sempre que uma outra experiência ou sensação similar ocorre, permitindo então, o seu reconhecimento. O aprendizado de algo novo é sempre feito em relação àquilo que já era conhecido, isto é, ele confirma ou modifica as ligações preexistentes

entre os neurônios. Na verdade, diferentes experiências ou sensações são armazenadas no cérebro com diferentes ligações, ou diferentes intensidades de ligações, entre neurônios.

Como o cérebro humano, as redes neurais são constituídas de um conjunto de unidades de processamento conectadas entre si, chamadas de neurônios artificiais ou, simplesmente, neurônios. Cada neurônio possui uma função de ativação, permitindo que, a partir de um valor recebido como entrada e do valor já armazenado internamente, seja gerado um valor de saída, que será propagado ao neurônio seguinte. Em uma rede neural, a cada conexão que liga um neurônio a outro é atribuído um peso, chamado de peso sináptico (em referência à sinapse do cérebro humano). Durante o processo de aprendizagem, os pesos dados a cada conexão vão se modificando até que se chegue a um conjunto de valores de pesos que, aplicados aos valores de entrada de cada neurônio, resultem em respostas satisfatórias ao problema estudado. Assim, pode-se dizer que os pesos sinápticos constituem o meio utilizado por uma rede neural para armazenar o conhecimento adquirido durante o processo de aprendizagem. É importante salientar que a maioria dos autores define o funcionamento das redes neurais com base nesse modelo neuronal, em que pesos são atribuídos a cada conexão entre os neurônios. No entanto, existem modelos de redes neurais sem pesos (RNSPs), que armazenam o conhecimento adquirido na memória dos neurônios, através de tabelas-verdade (BRAGA et al., 2000, citados por FREITAS, 2001).

O neurônio artificial foi projetado conforme um neurônio biológico dos seres vivos (Figura 1) e seu mecanismo pode ser ilustrado conforme a Figura 2. Assim, pode-se observar que um conjunto de entradas, aqui representadas por X_1 , X_2 e X_3 , são aplicadas ao neurônio artificial e serão multiplicadas por um peso correspondente W_i , gerando, portanto, entradas ponderadas de dados. Fazendo uma analogia com a realidade (Figura 1), estas ponderações representam a força das *sinapses*. Após esta entrada de dados ponderados, os mesmos são somados, resultando no *potencial de ativação* representado por v . Por conseguinte o valor de v é submetido a uma *função de ativação*, cujo resultado Y é o valor de saída do neurônio. Vale ressaltar que o valor v será comparado com o valor limite para ativação do neurônio; caso este valor alcance o valor limite de ativação ele se ativará, caso contrário ele ficará inativo.

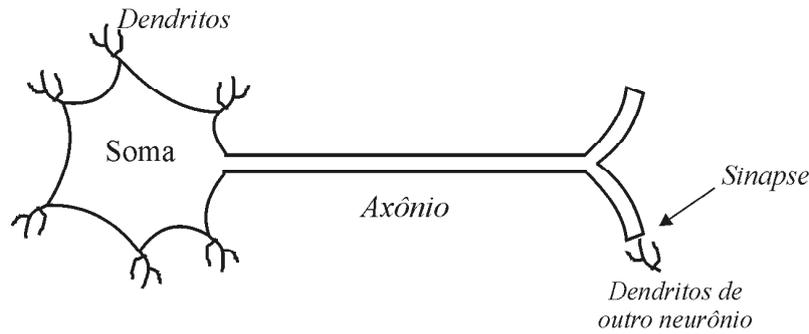


FIGURA 1 – Neurônio biológico, adaptado de Fausset (1994).

Fonte: Fausset (1994, p.35).

Nota-se que, além dos elementos mencionados há um outro elemento, representado por B (Figura 2). Esse elemento corresponde a um parâmetro, conhecido como *bias*, que não está presente em todos os tipos de neurônios artificiais. Seu objetivo é permitir uma calibragem do sinal de saída, sendo ele somado aos dados ponderados, resultando no potencial de ativação v . Expressando em forma matemática tem-se :

$$S = \sum_{i=1}^n W_i X_i \quad (1)$$

$$v = S + B \text{ e } Y = \varphi(v) \Rightarrow Y = \varphi(S + B) \quad (2)$$

Na literatura, pode-se encontrar uma outra denominação para tais *bias*, denominada *activation threshold*.

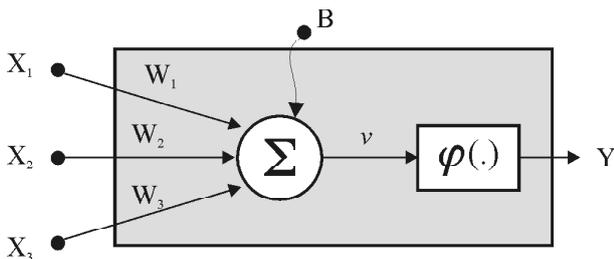


FIGURA 2 – Modelo de neurônio artificial.

Fonte: Freitas (2001, p. 22).

2.1.1 Função de Ativação

A função de ativação, aqui representada por $\varphi(\cdot)$, tem como objetivo processar o sinal recebido v para produzir

a saída final do neurônio artificial, ou seja, ela aplica do potencial de ativação v e gera Y (SAÍDA). Na literatura, podem-se identificar vários tipos de funções de ativação: uma simples função linear; uma função limiar (função degrau); uma função sigmoideal; função rampa; ou ainda uma função que simule mais precisamente as características não lineares do neurônio biológico. No caso de ser uma função linear, sua saída do neurônio é dada por:

$$\varphi(v) = \alpha.v \quad (3)$$

onde α é uma constante. Caso $\varphi(v)$ seja uma função limiar, tem-se:

$$\varphi(v) = \begin{cases} +\gamma & \text{se } v > 0 \\ -\gamma & \text{se } v \leq 0 \end{cases} \quad (4)$$

e a função rampa como:

$$\varphi(v) = \begin{cases} +\gamma & \text{se } v \geq +\gamma \\ \alpha.v & \text{se } -\gamma < v < +\gamma \\ -\gamma & \text{se } v \leq -\gamma \end{cases} \quad (5)$$

onde γ é o valor de limiar do neurônio artificial (constante). No caso de se querer imitar com maior precisão os neurônios biológicos, usam-se funções não lineares, chamadas funções “*squashing*”. Elas caracterizam-se por não deixarem Y (Saída) exceder certos valores limites, os quais, em geral, são menores que os valores v . A função “*squashing*” mais usada é a função logística ou sigmoideal, sendo representada por:

$$\varphi(v) = \frac{1}{1 + e^{-v/T}} \quad (6)$$

Freitas (2001) afirma que as funções sigmóides, com curvas de diferentes suavidades, podem ser obtidas variando-se o parâmetro T . Portanto, o parâmetro T determina a suavidade da curva.

Outra função de ativação “squashing” bastante usada é a tangente hiperbólica. Ela possui uma forma similar à da sigmóide e é frequentemente usada por biólogos para o modelo matemático de ativação das células nervosas. A função é expressa como:

$$\varphi(v) = \tanh(v) \quad (7)$$

Apesar da forma semelhante a da sigmóide, a tangente hiperbólica possui valores de Y bipolares, característica que é benéfica para certos tipos de redes neurais artificiais.

Abelém (1994) coloca que este modelo simples de neurônio artificial ignora diversas características do neurônio natural, tais como a não consideração dos atrasos de tempo que afetam a dinâmica do sistema – as entradas produzem saídas imediatas – e a não inclusão dos efeitos de sincronismo ou de modulação de frequência – característica que alguns pesquisadores consideram de fundamental importância. Apesar dessas limitações, as redes neurais artificiais formadas por simples neurônios artificiais possuem atributos semelhantes aos do sistema biológico, como a capacidade de aprendizado e generalização, podendo-se dizer que a essência do funcionamento do neurônio natural foi absorvida.

2.1.2 Arquitetura de Redes Neurais

A arquitetura da rede é muito importante para as capacidades de convergência e generalização das Redes Neurais Artificiais. A escolha de uma arquitetura adequada, nem tão grande e nem tão pequena, é a principal responsável pelo sucesso de uma aplicação. Para se obter uma boa habilidade de generalização, deve-se projetar a rede colocando dentro dela o máximo conhecimento possível sobre o problema e limitar o número de conexões adequadamente (ABALÉM, 1994).

De forma mais simplificada, definir a arquitetura de uma rede neural consiste em determinar a estrutura sobre a qual os neurônios da rede estarão organizados. A

arquitetura da rede é definida pelas seguintes características: número de camadas da rede, número de neurônios em cada camada, tipo de conexão entre os neurônios e a topologia da rede (BRAGA et al., 2000).

Existem várias técnicas que tentam encontrar a arquitetura ideal através da diminuição de neurônios redundantes e do compartilhamento de pesos. Pode-se citar Hertz et al. (1991) e Regenes & Vithlani (1991), entre outras, porém, nenhuma conseguiu resultados convincentes para todas as aplicações. Existe também a possibilidade de fazer uma busca no espaço de todas as arquiteturas possíveis. Para isso, deve-se treinar cada uma das arquiteturas separadamente e então avaliá-las com uma função erro apropriada que incorpore o desempenho e o número de unidades. A busca também pode ser feita através de um *algoritmo genético*, tal que as boas arquiteturas encontradas em uma triagem sejam mantidas para serem combinadas com boas arquiteturas de outras triagens. Contudo, esses tipos de busca se mostram pouco práticos para aplicações que utilizam grandes redes neurais, onde cada treinamento exige um alto esforço computacional. A verdade é que a arquitetura ideal das redes neurais ainda continua sendo determinada principalmente pela experimentação manual e pela experiência e sensibilidade do projetista (ABELÉM, 1994).

Dentre as diversas arquiteturas da rede pode-se demonstrar as redes de camada única e as de múltiplas camadas. As redes de camada única se caracterizam por possuir apenas um neurônio entre cada entrada e cada saída; já para as de múltiplas camadas, há mais de um neurônio entre qualquer entrada e qualquer saída da rede. As Figuras 3a e 3b representam estes dois tipos de redes neurais artificiais.

Vale ressaltar que a utilização de camadas intermediárias (ou ocultas) permite às redes neurais artificiais implementar superfícies de decisão mais complexas. Essas camadas permitem que seus elementos se organizem de tal forma que cada elemento aprenda a reconhecer características diferentes do espaço de entrada; assim, o algoritmo de treinamento deve decidir que características devem ser extraídas do conjunto de treinamento. A desvantagem em utilizar camada escondida é que o aprendizado se torna muito mais difícil, pois o número de especificações para cada neurônio aumenta, juntamente com seus diferentes tipos e estruturas de entrada de dados, o que prolonga a convergência das funções de treinamento.

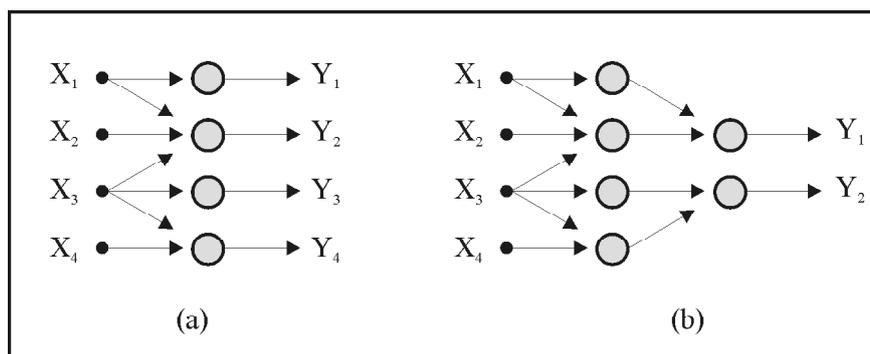


FIGURA 3 – Rede neural: (a) com camada única (b) com múltiplas camadas.

Fonte: Freitas (2001, p. 24).

Conforme Freitas (2001), quanto ao tipo de conexão, as redes neurais podem ser *feedforward* (acíclicas) ou recorrentes (cíclicas). Uma rede neural recorrente diferencia-se das redes neurais acíclicas por possuir pelo menos um *link* de realimentação (*feedforward*), pelo qual a saída de algum neurônio de uma camada i é utilizada como entrada para um neurônio de cada camada de ordem menor ou igual a i . Uma rede neural em que todas as conexões são cíclicas é chamada de auto-associativa.

2.1.3 Aprendizagem

A propriedade mais importante das redes neurais é a habilidade de aprender a partir de seu próprio ambiente e com isso melhorar seu desempenho. Isto é feito através de um processo iterativo de ajuste aplicado a seus pesos, o treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

O processo de aprendizagem nas redes neurais acontece, internamente, por meio do ajuste dos pesos sinápticos das conexões durante a exposição dos exemplos, em resposta à quantidade de erros gerados pela rede. Ou seja, a rede neural é capaz de modificar-se em função da necessidade de aprender a informação que lhe foi apresentada (TAFNER et al., 1998).

As redes neurais são treinadas para aprender a partir dos dados de entrada. Assim como o cérebro humano, elas aprendem a partir de experiências e não através de programação. Por este motivo, deve-se tomar bastante cuidado com a formação do conjunto de treinamento. Este conjunto deve ser gerado a partir de dados históricos, ou seja, a partir de experiências e fatos ocorridos no passado.

Como já mencionado anteriormente, a rede deve ser capaz de generalizar. Mas, ao mesmo tempo, deve-se

tomar cuidado para que não aconteça um “supertreinamento” e memorização dos dados. Se uma rede neural é submetida a um supertreinamento, ela perde a capacidade de reconhecer padrões fora do conjunto de treinamento. Para evitar essa situação, deve-se ter um conjunto de teste com dados diferentes do conjunto de treinamento, e a rede deve ser capaz de classificá-los corretamente, provando assim sua flexibilidade e capacidade de generalização.

Os três principais paradigmas de aprendizagem são apresentados a seguir:

a) Aprendizagem supervisionada, quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada. O ajuste dos pesos ocorre quando o sistema compara a saída da rede com a resposta desejada, previamente conhecida.

b) Aprendizagem não-supervisionada, quando não existe um agente externo indicando a resposta desejada para os padrões de entrada. A rede neural utiliza os neurônios como classificadores, e os dados de entrada como elementos de classificação. Esse tipo de rede trabalha essas entradas e se organiza de modo a classificá-las mediante algum critério de semelhança.

c) Aprendizagem híbrida, que mescla os conceitos apresentados anteriormente. Parte dos pesos é determinada através da aprendizagem supervisionada, enquanto outros são obtidos através da aprendizagem não-supervisionada.

Em um processo de aprendizagem, para alguns autores conhecidos como treinamento, os pesos dos neurônios são ajustados através de um algoritmo de aprendizagem. O algoritmo de aprendizagem é um conjunto pré-estabelecido de regras bem-definidas para resolução

de um problema de aprendizagem (HAYKIN, 2001). Ele tem como objetivo encontrar pesos para a rede que permitam que esta gere saídas compatíveis com as desejadas. Os algoritmos existentes diferem entre si pela forma como ocorre o ajuste dos pesos sinápticos dos neurônios, ou seja, pela regra de aprendizagem adotada.

Após o treinamento, é feito o teste com o primeiro padrão do conjunto de teste. Guarda-se o resultado obtido (a resposta da rede e o erro gerado) em um vetor, adiciona-se o padrão testado ao conjunto de treinamento e treina-se novamente a rede p vezes, agora com um padrão a mais. Depois desse treinamento, realiza-se o teste com o próximo padrão do conjunto de teste. Adiciona-se o novo resultado ao vetor de resultados, adiciona-se o padrão ao conjunto de treinamento voltando a treinar a rede p vezes, com o novo conjunto de treinamento. Este processo repete-se até que todos os padrões de teste tenham sido testados.

Conforme Mello (2004), os passos para o treinamento da rede, listados, são demonstrados no fluxograma na Figura 4, de acordo com as seguintes nomenclaturas:

- i = número de variáveis de entrada;
- t = padrão de entrada atual para teste;
- TR = número de padrões de entrada para treinamento;
- TT = número de padrões de entrada para teste;

Por fim, para se analisar o desempenho dos modelos de redes neurais, utilizam-se os Erros Quadráticos Médios (EQM), dados pela seguinte expressão matemática:

$$EQM = \frac{1}{n} \sum_{i=1}^n (P_i - \hat{P}_i)^2$$

onde P_i é o preço do boi gordo observado, \hat{P}_i é o preço do boi gordo estimado pelo modelo e n é o número de observações.

2.1.4 Redes Neurais Artificiais Multicamadas

A capacidade limitada de representação das Redes Neurais Artificiais com uma única camada foi superada na década de 60, a partir do perfeito entendimento do problema da separabilidade linear (MINSK & PAPERT, 1988). Contudo, apesar de se ter descoberto como aumentar a capacidade de representação da rede com o acréscimo de novas camadas, chamada de redes MLP (*Multilayer Perceptron*), faltava ainda uma forma de treinar a rede. Essa forma só foi descoberta em meados da década de 70, com a criação de vários algoritmos, entre eles o de retropropagação (*backpropagation*). O *backpropagation* se tornou um dos algoritmos mais populares para o treinamento, sendo em parte responsável pelo ressurgimento do interesse pela área de Redes Neurais Artificiais.

As redes conhecidas como MLP (*Multilayer Perceptron*) são assim chamadas pela existência de diferentes camadas (*layers*) intermediárias, trabalhando em uma forma acíclica (*feedforward*). A presença das camadas intermediárias nos modelos de redes MLP permitem a solução de problemas mais complexos.

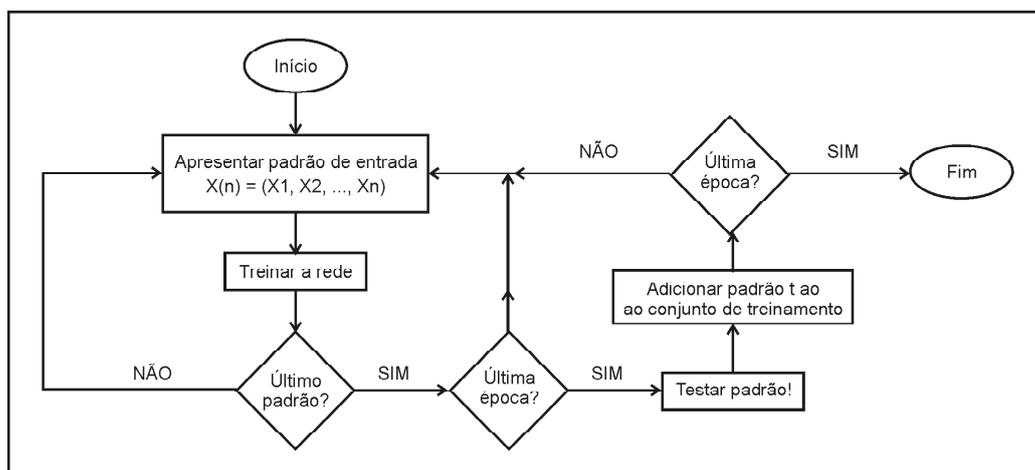


FIGURA 4 – Fluxograma do processo de treinamento da rede.

Fonte: Mello (2004, p. 32).

O algoritmo *backpropagation* segue o paradigma de aprendizado supervisionado, em que são fornecidos valores de entrada às redes neurais com as respectivas saídas desejadas. Através do algoritmo *backpropagation*, o processo de aprendizado é realizado em duas etapas. A primeira é a etapa *forward*, em que os valores de saída da rede são calculados a partir dos valores de entrada fornecidos. Na segunda etapa, a *backward*, os pesos associados a cada conexão são atualizados conforme as diferenças entre os valores de saída obtidos e os valores desejados, da última camada até a camada de entrada (FREITAS, 2001).

Conforme Corrêa & Portugal (1998) o algoritmo de *backpropagation* ajusta então as ponderações, de modo a minimizar os erros de previsão dentro da faixa teste. Baseado nesse erro, entre observado e estimado, os pesos são ajustados de modo a se obter um erro mínimo. O algoritmo de *backpropagation* é formalizado nas seguintes equações:

$$\Delta w_{ij,k}^n = \eta \cdot \delta_{j,k} \cdot y_{i,k-1}$$

$$w_{ij,k}^{n+1} = w_{ij,k}^n + \Delta w_{ij,k}^n$$

$$\delta_{j,k} = \left(\frac{dy}{dz} \right)_{i,k} \cdot (y_{i,k} - y_{i,k}^T)$$

com $i=1,2, \dots, I$; $j=1,2, \dots, J$, e $k=1, \dots, K$.

Onde: η é o coeficiente de aprendizado ($0,01 < \eta < 1,00$);

$y_{i,k}$ é a saída do neurônio i na camada k ;

$y_{i,k}^T$ é o valor desejado para $y_{i,k}$;

$w_{ij,k}^n$ é a ponderação ligando o neurônio i na camada $k-1$ ao neurônio j na camada k no passo n .

Assim, as previsões de valores futuros de séries temporais geradas por uma combinação de processos determinísticos e estocásticos não é uma tarefa simples, devendo ser corretamente analisadas e bem treinadas a fim de que forneçam boas previsões.

2.2 Modelos de Séries Temporais

2.2.1 Modelos ARIMA

Os modelos ARIMA (Auto-regressivo – Integrados – Médias móveis), inicialmente formulados por Box & Jenkins (1976), baseiam-se na idéia de que uma série temporal não-estacionária pode ser modelada a partir de diferenciações e da inclusão de componentes auto-

regressivos e médias móveis. Os modelos ARIMA possuem recursos que possibilitam a caracterização estatística de séries temporais e podem caracterizar, simultaneamente, num único modelo, três aspectos importantes: a autocorrelação (AR), a ordem de integração (I) e a dependência temporal de médias móveis (MA) (BITENCOURT, 2005). Uma das inovações dos modelos ARIMA é o fato de conter a ordem de integração (I) e trabalhar com séries não-estacionárias, ao contrário dos modelos ARMA, em que a série deve ser estacionária.

Baseando-se, então, na metodologia de Box & Jenkins (1976), o modelo auto-regressivo de ordem p , AR(p), pode ser representado por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (8)$$

E o modelo de Média Móvel de ordem q , MA(q), representa-se como:

$$Y_t = \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + e_t \quad (9)$$

Portanto, segundo a generalização de Box & Jenkins (1976), a partir de uma série temporal diferenciada de ordem (I) tem-se um modelo ARMA (p,q) dado por:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + e_t \quad (10)$$

onde Y_t representa a série no tempo t , e_t é um erro aleatório i.i.d (independente e identicamente distribuído), com $N \sim (0,1)$.

2.2.2 Modelos de volatilidade

A modelagem da volatilidade teve seus primórdios apresentados por Engle (1982), em que um processo denominado ARCH (*autoregressive conditional heteroskedasticity*) expressa a variância condicional como uma defasagem distribuída do quadrado dos retornos passados. A idéia básica é que o retorno na série é não-correlacionado serialmente, mas a volatilidade (variância condicional) depende de retornos passados por meio de uma função quadrática. Pode-se definir um modelo ARCH (p) por:

$$X_t = \sqrt{\sigma_t^2} \varepsilon_t \quad (11)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 \quad (12)$$

onde X_t é o retorno, ε_t é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média zero e variância um.

Morettin & Toloi (2004) colocam que, na prática, usualmente supomos $\varepsilon_t \sim N(0,1)$ ou $\varepsilon_t \sim t_\nu$ (t de Student com ν graus de liberdade).

Conforme Mol (2003), este modelo de variância condicional possui algumas propriedades desejáveis. Em primeiro lugar, por meio da técnica de decomposição de erros de predição, é possível construir a função de verossimilhança, tornando possível a estimação dos parâmetros pelo método de máxima verossimilhança. Essa propriedade é importante porque estes estimadores possuem distribuições conhecidas que viabilizam a execução de testes de hipóteses diversos. Além disso, é possível provar que esse modelo implica em uma distribuição não-condicional com “caudas pesadas” para os retornos.

A proposição original, elaborada por Engle (1982), mereceu extensos debates e diversos aperfeiçoamentos ao longo dos anos. A primeira, e mais significativa, foi introduzida por Bollerslev (1986) ao propor que a volatilidade condicionada fosse função não apenas dos quadrados dos erros passados (X_t^2), como também dos seus próprios valores (σ_{t-1}^2). Os modelos assim construídos foram denominados *Generalized ARCH* (GARCH). Em termos matemáticos, um modelo GARCH (p,q) pode ser expresso como:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + v_t \quad (13)$$

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^p \left(a_i \left(\left| \frac{X_{t-i}}{\sigma_{t-i}} \right| - E \left| \frac{X_{t-i}}{\sigma_{t-i}} \right| \right) + \gamma_i \frac{X_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^q \left(b_j \ln(\sigma_{t-j}^2) \right) \quad (14)$$

onde o parâmetro γ_i permite um efeito assimétrico; se $\gamma_i=0$, é um indicativo de ausência de assimetria na volatilidade. Neste sentido, um procedimento de teste para o efeito da assimetria na série é checar a significância de γ_i no modelo. Se é estatisticamente diferente de zero, evidencia-se um impacto diferenciado de choques negativos e positivos na volatilidade. Se < 0 , há presença do “efeito alavancagem”. Nesse modelo, a persistência de choques na volatilidade é medida pelo parâmetro b.

onde ω é a constante, v_t é um ruído branco [$N \sim (0,1)$]

A fim de se garantir que a variância condicional não seja negativa, e também a estacionariedade do processo, tem-se que $\omega > 0$; $\alpha_i \geq 0$, para $i = 1, \dots, p$;

$\beta_j \geq 0$, para $j = 1, \dots, q$; e $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$

É interessante ressaltar que, apesar de o modelo GARCH(p,q) captar corretamente diversas características observadas nas séries históricas de finanças, tais como a leptocurtose e o agrupamento de volatilidade, ele não capta o efeito de alavancagem, uma vez que a variância condicional é função apenas das magnitudes das inovações e não dos seus sinais (BROOKS, 2002). A partir daí, surgiu a idéia de se criar modelos que tenham a capacidade de captar a assimetria das respostas, tais como os modelos EGARCH, TARCH e APARCH.

O modelo EGARCH (*Exponential GARCH*), proposto por Nelson (1991), foi uma inovação aos modelos de volatilidade uma vez que os choques na variância passam a ter efeito exponencial e não quadrático. O mesmo apresenta duas vantagens em relação ao modelo GARCH: a incorporação dos efeitos assimétricos de mercado aos modelos auto-regressivos de volatilidade condicional, e a não imposição artificial de restrições aos parâmetros da equação, dada a sua formulação em termos logarítmicos.

Na forma simplificada o modelo EGARCH(p,q) pode ser exposto:

Segundo Bollerslev & Wooldridge (1992), estimadores de máxima verossimilhança dos parâmetros de um modelo GARCH podem ser utilizados, assumindo-se inovações gaussianas, ainda que a verdadeira distribuição não seja gaussiana. Para que se possa conseguir a consistência necessária nesse processo, utiliza-se a matriz de variância-covariância corrigida, proposta pelos autores. Neste trabalho, adotou-se a correção proposta por Bollerslev & Wooldridge (1992), uma vez que as distribuições de séries de retorno apresentam-se de forma leptocúrticas.

Um modelo mais simples, para a captação do comportamento assimétrico da volatilidade nas séries financeiras, foi apresentado por Glosten et al. (1993) e posteriormente implementado por Zakoian (1994), denominado por TARARCH (*Threshold ARCH*). Esse novo modelo é um caso particular do modelo ARCH não-linear, e a volatilidade agora segue a forma funcional:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \gamma X_{t-i}^2 d_{t-1} + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + v_t \quad (15)$$

onde: d_{t-1} é uma variável dummy que assume o valor igual a 1 se $X_{t-1} < 0$ (más notícias), e valor igual a 0 se $X_{t-1} > 0$ (boas notícias).

Destacando que, como no modelo GARCH, as condições para não negatividade e estacionariedade são: $\omega \geq 0$; $\alpha_i \geq 0$, para $i = 1, \dots, p$; $\beta_j \geq 0$, para $j = 1, \dots, q$; e $\sum \alpha_i + \gamma \geq 0$ (BROOKS, 2002).

Por fim tem-se o modelo ARCH com potência assimétrica (APARCH), proposto por Ding et al. (1993). O modelo APARCH(p,q) pode ser representado por:

$$\sigma_t^\delta = \omega + \sum_{i=1}^p \alpha_i (|X_{t-i}| - \gamma_i X_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta \quad (16)$$

onde $\delta \geq 0$ e $-1 < \gamma_i < 1$. Esse modelo concilia a flexibilidade de um expoente variante com o coeficiente de assimetria; dessa forma, esse modelo é capaz de detectar impactos assimétricos de choques sobre a volatilidade. Especificamente, se $\gamma_i > 0$, verifica-se o efeito alavancagem, isto é, choques negativos têm um impacto maior sobre a volatilidade da série que choques positivos. Se $\gamma_i < 0$, verifica-se o contrário. Esse efeito ocorre porque a distribuição dos retornos admite, usualmente, uma característica de caudas pesadas (BAUWENS & GIOT, 2001).

3 METODOLOGIA

3.1 Dados

Neste estudo foram utilizados dados secundários referentes às cotações diárias de fechamento do indicador de preço Esalq/BM&F dos contratos futuros de Boi Gordo, negociados na Bolsa de Mercadorias & Futuros (BM&F), observadas no período de 01 de agosto de 1997 a 27 de maio de 2005, num total de 1943 observações. Os dados foram coletados junto ao site do CEPEA/Esalq (www.cepea.esalq.usp.br). A escolha

da data inicial e final para a amostra foi arbitrada pelos pesquisadores, não tendo, portanto, nenhum significado especial.

O indicador utilizado atende às especificações do Contrato Futuro de Boi Gordo negociado na BM&F, ou seja, tratam-se de bovinos machos, castrados, bem acabados (carcaça convexa), em pasto ou confinamento, com peso vivo entre o mínimo de 450kg e o máximo de 550kg e idade de 42 meses. O indicador é uma média ponderada dos preços da arroba do boi gordo pagos pelos frigoríficos nas principais regiões produtoras do Estado de São Paulo (BRESSAN & LIMA, 2002).

3.2 Procedimentos Metodológicos

O primeiro passo da investigação empírica se faz a partir da escolha das variáveis que servirão de entrada para a rede neural. Serão estas variáveis que farão com que a rede gere a saída mais precisa para o determinado trabalho, portanto, a escolha delas não deve ser aleatória e sim, baseando-se em escolhas precisas.

Assim, como forma de escolha de variáveis que servirão como entrada para a rede neural utiliza-se a metodologia estatística de séries temporais, para sua definição. Segundo Morettim & Tolo (2004), uma série temporal é composta pelos seguintes fatores:

$$Z_t = T_t + S_t + a_t$$

onde, T_t representa a tendência, S_t representa o componente sazonal e a_t o erro aleatório. Substituído os componentes de tendência e sazonalidade na séries Z_t temos:

$$Z_t = \sum_{j=0}^m \beta_j t^j + \sum_{j=1}^{12} \alpha_j d_{jt} + a_t$$

A confirmação dos componentes de tendência e sazonalidade se faz a partir de testes, como é o caso do teste do sinal (Cox-Stuart) e o teste de Kruskal-Wallis que verificam a presença de tendência e sazonalidade, respectivamente. Para maiores detalhes sobre a metodologia dos testes, consulte Morettim & Tolo (2004).

Para se efetuar a modelagem dos preços utilizando as redes neurais, os dados devem estar normalizados, ou seja, as redes neurais artificiais exigem que suas entradas de dados estejam em um intervalo entre 1 e -1; há casos específicos em que se utilizam entradas de dados oscilando entre 1 e 0 mas, para este trabalho, utilizou-se o intervalo [1, -1], pela característica das funções de ativação e o tipo de arquitetura da rede.

Para tanto, a normalização dos dados é efetuada a partir da razão entre o preço no instante t e a maior cotação apresentada na série.

Devidamente tratados os dados de entrada da rede neural artificial, parte-se para a fase de modelagem, construção e treinamento da mesma.

A fim de se comparar a capacidade preditiva dos modelos baseados em redes neurais artificiais, utilizaram-se como fonte os modelos estatísticos de séries temporais não lineares com a presença de heteroscedasticidade condicional. Utilizaram-se mais especificamente os modelos ARIMA aplicados à média, e modelos GARCH, EGARCH, TARCH e APARCH para a variância, uma vez que estes modelos são amplamente utilizados e vêm apresentando bons resultados.

É interessante ressaltar que este trabalho tem como objetivo treinar uma rede neural artificial para a previsão de preços futuros; portanto, para se ter uma idéia do quanto uma previsão é “precisa”, é necessário haver uma forma de estimar o quanto se está errando, em média. Estes valores de imprecisão geralmente se originam da estatística, como é ilustrado, por exemplo, em Wonnacott & Wonnacott (1990). Esse erro é calculado de duas formas distintas neste trabalho:

a) Erro absoluto médio

$$EAM = \frac{\sum_{n=1}^n |\sigma_{in}^2 - \sigma_{pn}^2|}{n}$$

b) Raiz do erro quadrático médio

$$REQM = \left[\frac{\sum_{n=1}^n (\sigma_{in}^2 - \sigma_{pn}^2)^2}{n} \right]^{1/2}$$

Onde:

EAM é o erro absoluto médio nas previsões para n ;

REQM é a raiz do erro quadrático médio nas previsões para n ;

σ_{in}^2 é o retorno instantâneo em relação a observação n ;

σ_{pn}^2 é o retorno previsto em relação a observação n ; e

Assim, o modelo que apresentar os menores valores do EAM e REQM representa aquele que possui a melhor capacidade preditiva, pois seus erros de previsão são inferiores.

4 RESULTADOS E DISCUSSÃO

Para a série do boi gordo (Figura 5), pode-se verificar a presença apenas do componente de tendência, mediante os testes do sinal de Cox-Stuart e Kruskal-Wallis, não apresentados neste trabalho.

Assim, a série de tendência será um dos parâmetros de entrada no modelo de redes neurais.

Para a escolha de outras variáveis de entrada utiliza-se a função de autocorrelação e autocorrelação parcial, conforme exposto na Figura 6.

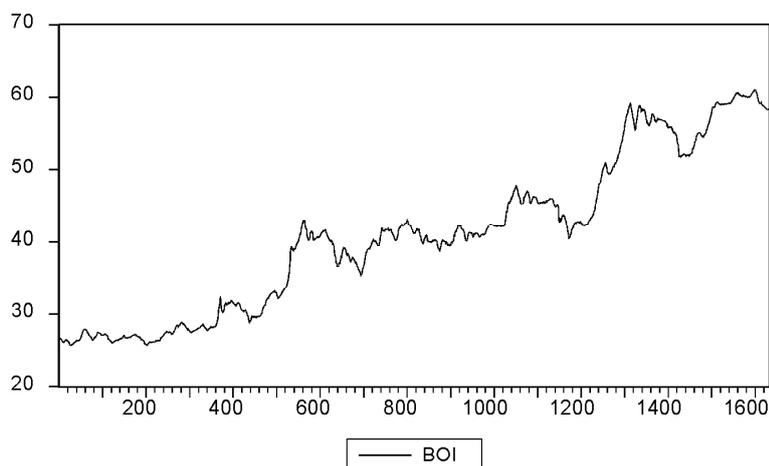


FIGURA 5 – Série de preço do boi gordo futuro na Bolsa de Mercadorias & Futuros (BM&F), Período de 01 de agosto de 1997 a 27 de maio de 2005.

Fonte: BM&F.

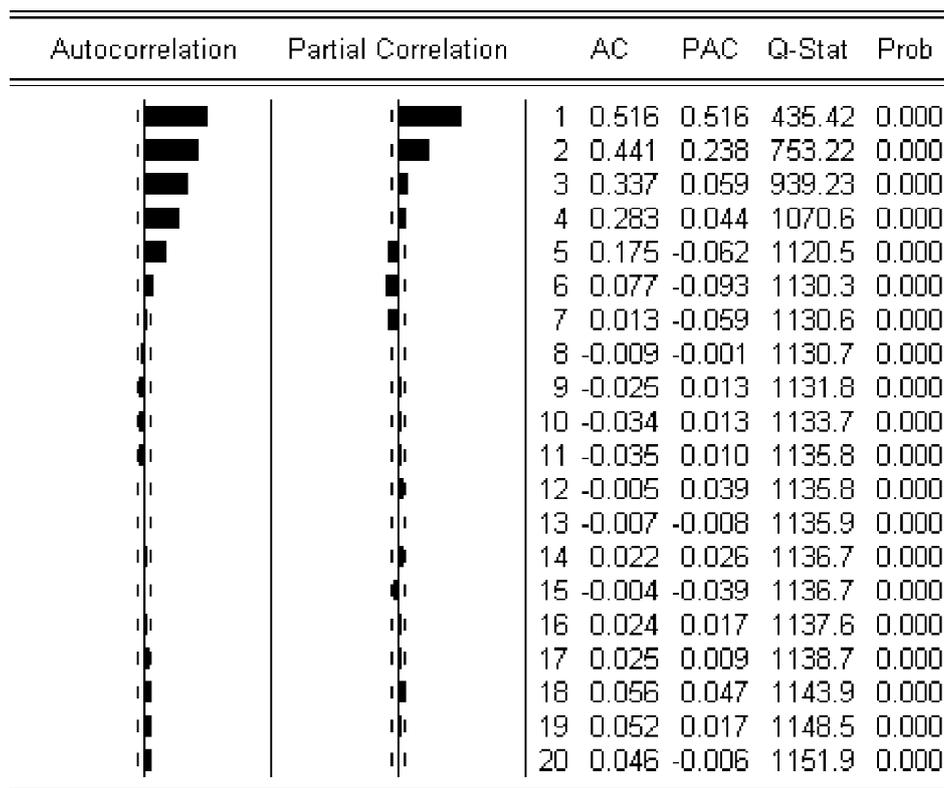


FIGURA 6 – Função de autocorrelação e autocorrelação parcial para a série de boi gordo futuro na BM&F.
Fonte: Dados da pesquisa.

Conforme exposto na Figura 6, pode-se observar que a função de autocorrelação parcial apresenta apenas dois *lags* fora do intervalo de confiança, caracterizando a necessidade de duas variáveis defasadas no modelo de rede neural. Portanto, serão incluídas duas variáveis defasadas no modelo de rede neural, sendo elas representadas por:

- Boi_{t-1} - Preço do boi gordo no tempo t-1;
- Boi_{t-2} - Preço do boi gordo no tempo t-2;

As variáveis que compõem o modelo foram normalizadas no intervalo de [0,1].

A arquitetura da rede pode ser representada conforme a Figura 7. Utilizou-se a rede MLP (*Multilayer Perceptron*), com uma camada intermediária com dois neurônios e uma camada de saída com apenas um neurônio. De acordo com Fausett (1994), duas ou mais camadas ocultas podem beneficiar determinadas

aplicações, mas uma única camada oculta é considerada suficiente. Os parâmetros que compõem a camada de entrada são, T_t , Boi_{t-1} e Boi_{t-2} . As conexões entre os neurônios são do tipo *feedforward* ou acíclica com a rede completamente conectada.

Além do número de neurônios da camada oculta, a taxa de aprendizagem, a taxa de *momentum* e o número de épocas, também foram sendo modificadas durante a realização de inúmeros testes na busca da combinação que produzisse o melhor resultado.

A função tangente hiperbólica (*tanh*) foi definida como a função de ativação para os neurônios da camada oculta. Ela assume valores em um intervalo contínuo entre -1 e 1. Sua fórmula é dada por:

$$\tanh(x) = \frac{1 - \exp(-x)}{1 + \exp(-x)}$$

O algoritmo de treinamento utilizado foi o algoritmo Levenberg-Marquardt. Conforme Freitas (2001), este algoritmo, desenvolvido para acelerar o tempo de treinamento, é uma variação do algoritmo *back-propagation*. Dessa forma, para o treinamento da rede neural, usou-se a função de treinamento *trainlm*, disponível no software Matlab.

É importante ressaltar que a combinação que resulta na melhor previsão para uma commodity não é a mesma combinação que gera o melhor resultado para outra. Portanto, é necessário configurar uma rede neural, com parâmetros diferentes, para cada commodity.

Para que se tenha um bom treinamento da rede, o número de épocas deve ser bem estabelecido; em outras palavras, esse número corresponde à quantidade de vezes em que a série em treinamento é comparada com a série original. Por isso, para este estudo testou-se diversas tentativas para as diversas épocas, testando elas entre 100 e 1000 repetições, ou seja, os dados de entrada e saída foram apresentados cerca de 1000 vezes para que a rede fosse treinada.

A Tabela 1 demonstra alguns resultados de diferentes modelos de redes neurais testados para a série de boi gordo futuro na BM&F.

Pode-se observar que, a partir dos valores do erro quadrático médio de treinamento, o primeiro modelo foi o que demonstrou melhor ajustamento à série, pois seu erro

foi reduzido. Porém observa-se também que os três modelos testados foram satisfatórios, visto que seus erros ficaram próximos de zero.

A Tabela 2 apresenta os resultados do ajustamento dos modelos de séries temporais aplicados a séries de boi gordo.

Portanto, conforme as Tabelas 2 e 3, observa-se que os modelos de séries temporais ajustados à série de boi gordo são satisfatórios; porém, dentre eles, o modelo AR(2)-EGARCH foi o que melhor demonstrou representar a realidade da série, tendo em vista os critérios de qualidade de ajuste, sendo ele então utilizado para previsão de preço do boi gordo.

É interessante salientar que a amostra foi dividida em dois grandes grupos, um para treinamento da rede, conforme os valores expostos na Tabela 1, com um total de 1581 observações, e o outro para a simulação e teste da rede, com um total de 50 observações.

A Figura 8 demonstra a simulação de preço futuro do boi gordo, na Bolsa de Mercadorias & Futuros para 50 dias após a amostra de treinamento, ou seja, estes valores correspondem às previsões efetuadas tanto pelos modelos de redes neurais quanto pelos modelos estatísticos de séries temporais. O modelo estatístico ajustado corresponde a um AR(2)-EGARCH(1,1), em outras palavras, um modelo autorregressivo de 2ª ordem e um EGARCH de 1ª ordem para a volatilidade defasada e 1ª ordem para o erro defasado.

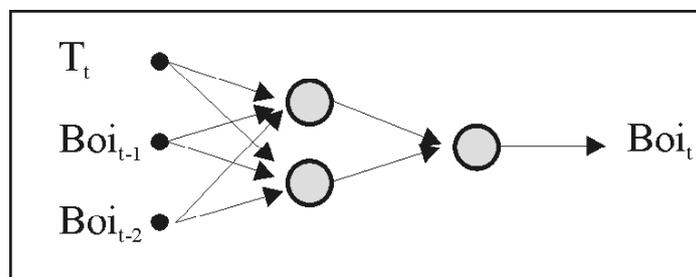


FIGURA 7 – Arquitetura da rede utilizada no trabalho.

Fonte: Dados da pesquisa.

TABELA 1 – Resultado da modelagem de redes neurais para o preço futuro do boi gordo.

Neurônio	Épocas	Momentum	Taxa de Aprendizagem	Erro de treinamento
2	1000	0.5	0.1	7.98106E-06
3	300	0.2	0.5	7.5552E-06
5	500	0.3	0.1	7.08E-06

Fonte: Dados da pesquisa.

TABELA 2 – Resultados da estimação dos modelos para a série de boi gordo.

Estimativas				
Especificação				
Média	AR	AR	AR	AR
Constante	-	-	-	-
AR(1)	0,41(16,1)	0,41(16,6)	0,41(15,9)	0,40(17,77)
AR(2)	0,29(12,9)	0,28(13,3)	0,29(12,9)	0,28(14,56)
Variância	GARCH	EGARCH	TARCH	APARCH
ω	0,01(11,4)	-1,49(-15,7)	0,02(12,0)	0,01(2,59)
ε^2_{t-1}	0,27(23,7)		0,23(10,7)	0,26(24,0)
σ^2_{t-1}	0,72(56,4)		0,70(51,6)	0,17(7,12)
$d_{t-1}\varepsilon^2_{t-1}$			0,13(6,34)	
$ \varepsilon_{t-1} / \sigma_{t-1} $		0,47(27,9)		
$\varepsilon_{t-1}/\sigma_{t-1}$		-0,05(-5,55)		
$\ln(\sigma^2_{t-1})$		0,89(118,5)		
γ_1				0,74(63,04)
δ				1,02(10,00)

Estatística t entre parênteses

Fonte: Dados da pesquisa.

TABELA 3 – Medidas da qualidade do ajuste dos modelos estimados para a série de boi gordo.

Qualidade do ajuste				
Indicadores	GARCH	EGARCH	TARCH	APARCH
ln(L)	8444,1	8463,7*	8448,9	8272,2
Akaike	-8,70	-8,72*	-8,70	-8,69
Shwarz	-8,70	-8,73*	-8,71	-8,67
Erro padrão	0,0037	0,0036*	0,0037	0,0037
ARCH(1)	0,922	0,980	0,933	0,943
ARCH(5)	0,870	0,881	0,801	0,861
ARCH(10)	0,984	0,989	0,971	0,981

Notas:

1. *denota o melhor modelo segundo cada critério.
2. ln(L) denota a máxima log-verossimilhança do modelo estimado.
3. Erro padrão denota o erro padrão do modelo.
4. ARCH(lag) denota o teste ARCH-LM efetuado até o lag entre parênteses.

Fonte: Dados da pesquisa.

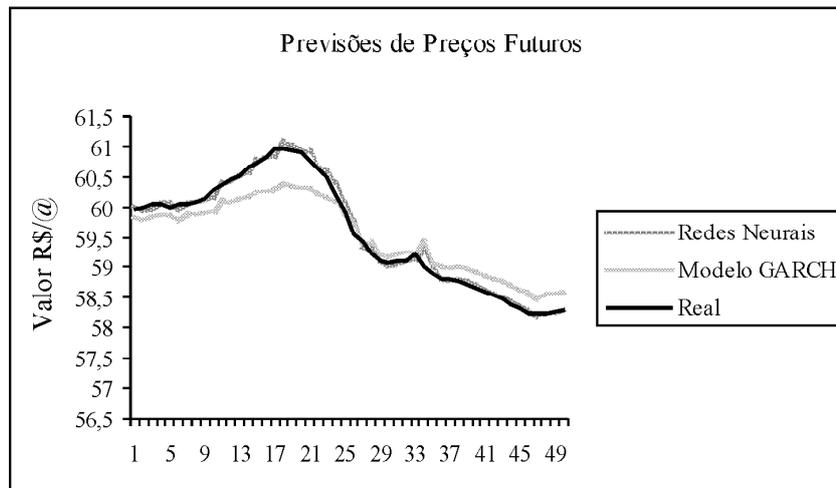


FIGURA 8 – Comparação entre os preços do boi gordo futuro na BM&F e as previsões com os modelos de redes neurais e métodos estatísticos (EGARCH).

Fonte: Dados da pesquisa.

A Tabela 4 demonstra o desempenho e previsão dos modelos baseados em redes neurais e o modelo de série temporal AR-EGARCH, através dos critérios de Erro Quadrático Médio (EQM) e Raiz do Erro Quadrático Médio (REQM) demonstrado no item 4.

TABELA 4 – Avaliação de desempenho de previsão dos modelos.

	EQM	REQM
AR-EGARCH	0.084304	0.258749
Redes Neurais	0.007051	0.062432

Fonte: Dados da pesquisa.

A partir da Figura 8 e da Tabela 4, fica evidente que os modelos de redes neurais, uma vez bem ajustados, apresentam melhor capacidade preditiva, conforme visualizados pelos EQM e REQM, para a série de preço futuro do boi gordo na BM&F. Esta constatação está de acordo com a literatura atual, que mostra a eficiência das redes neurais aplicadas às previsões de valores futuros.

Os resultados apresentados neste trabalho sugerem que a utilização de modelos para previsão de preço futuro é um valioso instrumento para investidores dos mercados derivativos que queiram mensurar seus riscos e até, quem sabe, decidir em optar por estratégias que possam reduzir seus riscos e maximizar seus retornos.

5 CONCLUSÃO

A utilização de previsão de preço nas informações semanais do boi gordo é uma excelente ferramenta para a geração de informações e redução dos riscos. Diante das dificuldades encontradas pelos investidores de mercado a respeito dos acontecimentos futuros e mensuração do risco de investimentos, esta metodologia proporciona uma maior facilidade para eles obterem, de forma dinâmica, uma previsão computacional para as possíveis oscilações durante a semana no mercado futuro de boi gordo, contribuindo para que se tenha uma maior precisão dos investimentos e segurança de seus riscos.

É importante ressaltar que, de qualquer forma, os modelos utilizados apresentam fortes avanços metodológicos no tratamento dos preços, uma vez que conseguem contornar os problemas de clusters de volatilidades, excesso de custos e assimetria nos choques das informações, assumindo um rigor metodológico acima das técnicas usuais. Os estudos sobre previsão apresentam-se como elemento primário na decisão de investimentos em mercados de derivativos.

A utilização destes modelos de previsão de preço futuro pode ser uma importante ferramenta de gerenciamento de risco a ser adotada pelos investidores (*hedgers*, especuladores, etc.), o que poderá incrementar o volume de negociações de contratos futuros pela redução da incerteza, aumentando assim a liquidez dos contratos negociados, o que é benéfico para o mercado como um todo.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABELÉM, A. J. G. **Redes neurais artificiais na previsão de séries temporais**. 1994. 100 f. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica, Rio de Janeiro, 1994.
- BAUWENS, L.; GIOT, P. **Econometric modelling of stock market intraday activity**. Dordrecht: Kluwer, 2001.
- BITENCOURT, W. A. **Criação e distribuição de informações ao mercado futuro da commodity boi gordo**. 2005. 44 p. Monografia (Graduação em Administração) – Universidade Federal de Lavras, Lavras, 2005.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, [S.l.], v. 31, p. 307-327, 1986.
- BOLLERSLEV, T.; WOOLDRIDGE, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. **Econometric Reviews**, [S.l.], v. 11, n. 2, p. 143-172, 1992.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Holden-Day, 1976.
- BRAGA, A.; LUDEMIR, T.; CARVALHO, A. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.
- BRESSAN, A. A.; LIMA, J. E. Modelos de previsão de preços aplicados aos contratos futuros de boi gordo na BM&F. **Nova Economia**, [S.l.], v. 12, n. 1, p. 117-140, 2002.
- BROOKS, C. **Introductory econometrics for finance**. Cambridge, UK: Cambridge University, 2002.
- CORRÊA, W. L. R.; PORTUGAL, M. S. Previsão de séries de tempo na presença de mudança estrutural: redes neurais artificiais e modelos estruturais. **Economia Aplicada**, São Paulo, v. 2, n. 3, p. 487-514, 1998.
- DING, Z.; GRANGER, C. W. J.; ENGLE, R. F. A long memory property of stock market returns and a new model. **Journal of Empirical Finance**, [S.l.], p. 83-106, 1993.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, [S.l.], v. 50, n. 4, p. 987-1007, 1982.
- FAUSSET, L. **Fundamentals of neural networks architectures, algorithms and applications**. [S.l.]: Prentice Hall, 1994.
- FREITAS, S. O. de. **Utilização de modelo baseado em redes neurais para a precificação de opções**. 2001. 50 f. Dissertação (Mestrado em Administração) – Universidade Federal de Minas Gerais, Belo Horizonte, 2001.
- GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. Relationship between the expected value and volatility of the nominal excess return on stock. **Journal of Finance**, [S.l.], v. 48, p. 1779-1801, 1993.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.
- HERTZ, J.; KROGH, A.; PALMER, R. G. **Introduction to the theory of neural computation**. Redwood City: A. Wesley, 1991.
- MELLO, A. T. **Aplicação de redes neurais artificiais no processo de precificação de ações**. 2004. 49 f. Monografia (Graduação em Ciência da Computação) – Universidade Federal de Pelotas, Pelotas, 2004.
- MINSK, M. L.; PAPERT, S. A. **Perceptrons: an introduction to computational geometry**. 3. ed. Massachusetts: M.I.T., 1988.
- MOL, A. L. R. **Value at Risk como medida de risco da volatilidade dos ajustes diários em mercados futuros de café**. 2003. 102 p. Dissertação (Mestrado em Administração) – Universidade Federal de Lavras, Lavras, 2003.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: E. Blücher, 2004.
- NELSON, D. B. Conditional heteroskedasticity in asset returns. **Econometrica**, [S.l.], v. 59, p. 347-370, 1991.
- REGENES, A. N.; VITHLANI, S. Constructive learning by specialisation. In: ICANN-91, 1991, Helsinki. **Anais... Helsinki**: [s.n.], 1991.
- TAFNER, M. A. Redes neurais artificiais: aprendizado e plasticidade. **Revista Cérebro & Mente**, [S.l.], n. 5, maio 1998.

WONNACOTT, T.; WONNACOTT, R. **Introductory statistics for business and economics**. New York: J. Wiley & Sons, 1990.

ZAKOIAN, J. M. Threshold heteroskedasticity models. **Journal of Economic Dynamics and Control**, [S.l.], v. 18, p. 931-955, 1994.

