

PREVISÃO DO PREÇO DA SOJA: uma comparação entre os modelos ARIMA e redes neurais artificiais¹

Paulo Sergio Ceretta²
Marcelo Brutti Righi³
Sergio Guilherme Schlender⁴

1 - INTRODUÇÃO

Nas últimas décadas, a soja consolidou-se a principal e mais expressiva *commodity* de exportação no Brasil. Só em 2008, a safra da soja em grão atingiu 60 milhões de toneladas. Considerando todo o complexo - grão, farelo e óleo -, 39 milhões de toneladas foram exportadas no referido ano, quantidade que corresponde a 26% do total das exportações do agronegócio. Isoladamente, a soja em grão representou 62% da quantidade exportada, contribuindo com mais de 24 milhões de toneladas. Ou seja, 41% da produção nacional de soja em grão foi comercializada no mercado externo (CONAB, 2009).

Conseqüentemente, o mercado da soja, canalizado com as operações na Bolsa de Mercadorias e Futuros do Brasil (BM&F), e balizado com as cotações internacionais na Bolsa de Chicago (CBOT) e no porto de Rotterdam, influencia diretamente na estrutura econômica do País, pois, além de ocupar um papel destacado no cenário mundial - já que o Brasil é atualmente o segundo maior produto da *commodity* no mundo, seguido pela Argentina, China e Índia, sendo os Estados Unidos o maior produtor - também afeta elementos que ligam desde a produção até a logística no País (FAO, 2010; BECERRA, 2007; IICA, 2007).

Esse mercado tem como característica mais observada sua difícil previsibilidade de preço e de produção. Tal característica é decorrente de fatores que estão além do controle por parte do produtor, como variações na oferta e demanda,

além das intempéries climáticas e enfermidades durante a produção (ABITANTE, 2008).

Devido a esta característica, diversas técnicas de previsão são apresentadas para auxiliar a tomada de decisão por parte dos agentes envolvidos, a fim de administrar os riscos de mercado, mais especificamente, as oscilações dos preços que afetam a estrutura e a eficiência do mercado da soja (ABITANTE, 2008; BRESSAN, 2004).

Dentro dessas técnicas de previsão, encontram-se o modelo Autorregressivo Integrado de Médias Móveis (ARIMA), de Box; Jenkins; Reinsel (1994), e as Redes Neurais Artificiais (RNAs), sendo ambos instrumentos de segurança e sinalização dos preços do mercado distribuídos num tempo determinado. Enquanto o ARIMA é o modelo mais utilizado e conhecido, as Redes Neurais Artificiais têm como principal vantagem a facilidade da compreensão dos fenômenos com menos restrições, e vêm obtendo interesse crescente de pesquisa nas últimas décadas devido ao seu desempenho na previsão (SOBREIRO; ARAÚJO; NAGANO, 2009).

Assim sendo, este artigo compara o modelo ARIMA e as RNAs no que tange ao ajuste dos mesmos à amostra dos valores pertencentes à série temporal de preços da soja, visando a explicação e a previsão de seu comportamento ao longo do período determinado. Também será analisada a qualidade dos modelos fora da amostra, quando são realizadas estimativas um passo à frente.

2 - REFERENCIAL TEÓRICO

O referencial teórico do trabalho abordará os seguintes temas: a presença da soja no Brasil, que contextualizará a cultura e o mercado dessa *commodity* no País; a previsão de séries temporais financeiras, expondo breve explicação,

¹Registrado no CCTC, IE-52/2010.

²Administrador, Doutor, Universidade Federal de Santa Maria, (e-mail: ceretta@smail.ufsm.br).

³Graduando em Administração, Universidade Federal de Santa Maria (e-mail: marcelobrutti@hotmail.com).

⁴Graduando em Administração, Universidade Federal de Santa Maria (e-mail: sergio.schlender1@gmail.com).

argumentando também sobre os modelos estatísticos que serão empregados na metodologia.

2.1 - Importância da Soja no Brasil

A origem da soja remete ao ano de 1882, quando professores da Escola de Agronomia da Bahia realizaram os primeiros estudos de avaliação da cultura de produção que era utilizada pelos americanos. Nove anos após esses estudos, vários testes de culturas semelhantes aos conduzidos pelos professores baianos foram realizados no Instituto Agrônomo de Campinas, da Agência Paulista de Tecnologia dos Agronegócios (IAC/APTA) da Secretaria de Agricultura e Abastecimento do Estado de São Paulo. Nessa época, a soja era mais utilizada como forrageira do que como planta fornecedora de grãos para a indústria de farelos e óleos vegetais, pois pouco se utilizava a soja para consumo de animais naquele período.

O principal fator para o sucesso da cultura no País, se deve à semelhança do ecossistema do Sul do Brasil com o que é predominante no extremo sul dos Estados Unidos, condição que favoreceu o êxito na transferência das variedades do grão bem como as implementações tecnológicas de armazenamento e produção.

Com o tempo, observa-se o surgimento de um sistema cooperativista dinâmico e eficiente no País, apoiando fortemente a produção, a industrialização e a comercialização das safras, o que gerou o estabelecimento de uma bem articulada rede de pesquisa de soja envolvendo governo e iniciativa privada. Além desses fatores citados, pode-se ainda destacar melhorias nos sistemas viário, portuário e de comunicações, que facilitam o transporte e as exportações (MELO FILHO; RICHETT, 1998).

Tendo em vista tais melhorias, aliadas ao fato de a soja ser a maior e mais expressiva *commodity* no Brasil em termos de exportação (SILVA; SANTO; SILVA, 2003), as transações financeiras tornam-se relevantes para a economia do país, e decididas, desde 2008, de acordo com as negociações operadas na companhia aberta de capital brasileira BM&FBovespa, em São Paulo.

A BM&FBovespa é o principal meio para as transações de compra e venda de mercadorias ou ativos, incluindo as da soja e de outros negócios agropecuários. Além disso, a descoberta de preços para serem negociados e a proteção e

gestão dos agentes envolvidos (*hedge*) são funções primordiais da instituição (BECERRA, 2007).

Especificamente a respeito do mercado da soja no Brasil, a partir do momento em que a volatilidade dos preços das *commodities* começaram a ser balizados na Bolsa de Chicago (CBOT) e na comercialização em Rotterdam, a soja brasileira consegue consolidar sua competitividade no cenário internacional. Isso ocorre porque as cotações são transmitidas instantaneamente e sem defasagens para os preços recebidos pelos produtores no Brasil, dando ao produtor a obtenção de um “preço justo”. Além disso, a taxa de câmbio em negociação começa a funcionar como um seguro contra inflação e altas nos custos de produção (MARGARIDO; SOUSA, 1998; IICA, 2007).

Esse tipo de balizamento deriva da tentativa de diminuir as dificuldades na previsão de preços e produção, decorrentes de fatores como: variações na oferta e demanda (necessidades e exigências da demanda dos países consumidores, além da demanda do setor alimentício animal, já que a soja é também matéria-prima na avicultura e pecuária); custos atrelados à *basis* doméstico dos preços e respectiva cadeia produtiva; além das intempéries climáticas e enfermidades durante a produção (ABITANTE, 2008; MISSÃO, 2006; TORRES FILHO, 2004).

2.2 - Previsões de Séries Temporais

Entende-se por série temporal ou histórica, segundo Souza (1989) e Larrote e Cardoso (2001), a classe de fenômenos cujo processo observacional e consequente quantificação numérica geram uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico.

A análise de uma série temporal modela o fenômeno estudado para, a partir daí, descrever o comportamento (sistemático e com um padrão persistente) da série, fazer estimativas e, por último, avaliar quais os fatores que influenciaram o comportamento da série, buscando definir relações de causa e efeito entre duas ou mais séries (LARROTE; CARDOSO, 2001; PINDYCK; RUBENFIELD, 1991).

Mais especificamente, a análise de séries temporais financeiras está preocupada com a teoria e prática dos valores de ativos no decorrer do tempo. Essa é uma disciplina altamente empíri-

ca, mas, como em outros campos da ciência, a teoria forma a fundamentação para que seja possível realizar inferência (TSAY, 2005). O mesmo autor enfatiza que existe, entretanto, um elemento-chave que distingue a análise de séries temporais financeiras da análise das demais séries temporais. Tanto a teoria financeira como suas séries temporais empíricas contêm um elemento de incerteza. Como resultado dessa incerteza, a teoria e os métodos estatísticos ocupam um importante papel na análise de séries temporais financeiras.

Dentre os modelos estatísticos e econométricos mais utilizados, destacam-se o modelo ARIMA, de Box; Jenkins; Reinsel (1994), e o modelo de RNAs. O primeiro é popularmente utilizado para previsão de variáveis econômicas, mercadológicas e sociais, porém restringe-se ao uso preferencial de 100 ou mais observações, já que tais variáveis são de natureza não estacionária, ou seja, com média e variância não constantes ao longo da amostra (TSENG; TZENG, 2002; SOBREIRO; ARAÚJO; NAGANO, 2009; BRESAN, 2004). Já no segundo, são consideradas classes de modelos não-lineares que permitem identificar padrões em dados, compreendendo as interconexões entre as unidades de processamento não-lineares, chamadas neurônio (nesse caso as séries temporais), que incorporam padrão de comportamento para prever, da maneira mais eficiente possível, valores futuros dessa série (SOBREIRO; ARAÚJO; NAGANO, 2009; BRESAN, 2004).

Existem vários estudos que, conforme Sobreiro; Araújo; Nagano (2009), procuram comparar as Redes Neurais com métodos econométricos como o ARIMA na previsão de séries temporais, que resultam na conclusão de que os resultados na Rede Neural são mais preditivos do que no ARIMA - como os trabalhos de Church e Curram (1996), Kohzadi et al. (1996), Cao; Leggio; Schniederjans (2005), Kumar (2005), Co e Boosarawongse (2007), Wang e Elhag (2007) e Pino et al. (2008). Porém, o mesmo autor adverte que tais qualidades ou vantagens da Rede Neural são dependentes do objeto de observação e seus dados, já que o melhor modelo no caso seria resultante da combinação entre RNA e ARIMA.

3 - ASPECTOS METODOLÓGICOS

A metodologia deste trabalho abordará

os seguintes temas: descrição dos dados e medidas de avaliação do ajuste, que apresentará o tipo de pesquisa aplicado, seu objetivo, os dados coletados e os critérios de avaliação para avaliar e comparar o desempenho do ajuste entre os modelos; e a apresentação do modelo ARIMA e das RNAs, que explanará os modelos propostos pelo trabalho, que servirão como base para posterior análise dos dados e para a conclusão.

3.1 - Descrições dos Dados e Medidas de Avaliação da Previsão

Neste estudo será efetuada uma pesquisa descritiva longitudinal. Vieira (2002) destaca que a pesquisa descritiva expõe as características de determinada população ou de determinado fenômeno, sem ter, necessariamente, o compromisso de explicar os fenômenos que descreve, embora sirva de base para tal explicação. Já para Gil (2002), as pesquisas descritivas têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre as variáveis.

Foram coletados dados referentes à série histórica de preços do mercado *spot* (à vista) da soja. A cotação desses dados está em moeda brasileira, isto é, em real, devido ao fato de que com o uso do real, tem-se uma percepção do contexto nacional do preço dessa *commodity*. A série de preços é diária, composta pelos preços de fechamento do indicador CEPEA/ESALQ/BM&FBovespa correspondentes ao período pertencente aos anos de 1997 a 2010, totalizando no período de ajuste dos modelos 3.114 observações (31/07/1997-10/02/2010), e no período de estimativas fora da amostra 30 observações (11/02/2010-26/03/2010) (CEPEA, 2010).

No período de ajuste dos modelos aos dados, o desempenho utilizará como critério de avaliação o coeficiente de determinação (R^2) e a Soma dos Quadrados do Resíduo. O uso dos coeficientes propostos se justifica pelo fato de que os modelos, ao serem analisados, possuem o mesmo número de variáveis exógenas e também o mesmo número de observações. O modelo que apresentar um maior coeficiente R^2 (situados entre zero e um, e mais próximos de um) será o modelo com maior poder de previsão. Não obstante, será realizado o teste F para confirmar a significância dos modelos estimados com base na amostra re-

presentante do preço da soja (GUJARATI, 2006).

3.2 - Modelo ARIMA

Segundo Tsay (2005) um modelo ARMA, basicamente, combina as ideias dos modelos auto-regressivos (AR) e por média móvel (MA) em uma forma compacta, de tal maneira que o número de parâmetros permaneça reduzido. Um modelo ARMA (p,q) genérico tem a forma, conforme [1].

$$P_t = \varphi_0 + \sum_{i=1}^p \varphi_i P_{t-i} + a_t + \sum_{i=1}^q \theta_i a_{t-i} \quad [1]$$

Onde φ_i e θ_i são constantes, (a_t) é uma série ruído branco e p e q são inteiros não-negativos. Os modelos AR e MA são casos especiais do modelo ARMA (p, q) . Usando o operador *backshift*, o modelo pode ser representado por [2].

$$\left(1 - \varphi_1 B^1 - \dots - \varphi_p B^p\right) P_t = \varphi_0 + \left(1 - \theta_1 B^1 - \dots - \theta_q B^q\right) a_t \quad [2]$$

O polinômio $1 - \varphi_1 B^1 - \dots - \varphi_p B^p$ é o polinômio AR do modelo. Similarmente, $1 - \theta_1 B^1 - \dots - \theta_q B^q$ é o polinômio MA. Tsay (2005) enfatiza que é preciso não existir fatores comuns entre os polinômios AR e MA, do contrário a ordem (p,q) do modelo pode ser reduzida. O mesmo autor afirma que o polinômio AR introduz a equação característica de um modelo ARMA.

Se todas as soluções da equação característica são menores que 1 (um) em valor absoluto, então o modelo ARMA é fracamente estacionário. Porém, séries de preços tendem a ser não-estacionárias. Essa não-estacionariedade se deve principalmente ao fato de que não há um nível fixado para o preço (TSAY, 2005). Na literatura sobre séries temporais, uma série não estacionária possui raiz unitária. Os testes mais utilizados para verificar a presença de raiz unitária são Dickey-Fuller Aumentado (ADF) e KPSS.

Se o modelo ARMA, anteriormente definido, permitir ao polinômio AR possuir raiz unitária, então ele se torna o modelo autorregressivo integrado com média-móvel (ARIMA). O modelo ARIMA tem forte memória, pois os coeficientes na representação MA não decaem para zero com o tempo, implicando que os choques passados a_{t-i}

do modelo tenham um efeito permanente na série (TSAY, 2005).

Uma abordagem convencional para lidar com a não-estacionariedade é usar diferenciação. Uma série temporal é dita ser um processo ARIMA $(p,1,q)$ se a série $\Delta P_t = P_t - P_{t-1}$ segue um modelo estacionário e inversível ARMA (p,q) (TSAY, 2005). A ideia de transformar séries não-estacionárias em estacionárias através de suas diferenças é conhecida na literatura com diferenciação. Mais formalmente, ΔP_t é referida como a série das primeiras diferenças de P_t . Não obstante, em alguns campos científicos, uma série temporal pode conter múltiplas raízes unitárias e, portanto, precisa ser diferenciada múltiplas vezes para se tornar estacionária.

Sendo assim, Gujarati (2006) indica que a identificação do modelo mais apropriado à série temporal é realizada por meio de funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Consequentemente, de maneira genérica, o modelo ARIMA pode ser aplicado em três fases: identificação do modelo, ou seja, verificação da aplicação do modelo; estimação dos parâmetros; e verificação de seu desempenho por meio das medidas de erro agregado (SOBREIRO; ARAÚJO; NAGANO, 2009).

3.3 - Redes Neurais Artificiais

Os Modelos de RNAs se diferenciam dos modelos tradicionais de previsão por serem modelos não-paramétricos, envolvendo algoritmos de aprendizado. Tais algoritmos buscam imitar a estrutura de interconexões do cérebro humano, com o intuito de incorporar o padrão de comportamento de uma série temporal de modo a prever, da maneira mais eficiente possível, valores futuros dessa série (TURBAN, 1993).

Esses modelos podem ser interpretados como uma classe de modelos de regressão não-linear, que permite ao analista ajustar grande número de parâmetros e testar diferentes configurações para um ajuste (CHATFIELD, 1996). Essas diferentes configurações podem ser interpretadas como formas funcionais alternativas para um ajuste adequado da rede em termos do número de camadas da rede e do número de neurônios em cada camada (FERNANDES; PORTUGAL; NAVAU, 1995).

A sistematização das RNAs é realizada

conforme o padrão de conexão entre camada das redes, os números dos neurônios em cada camada, da capacidade da aprendizagem e as funções de ativação, ou seja, uma rede neural é um conjunto de entrelaçamentos, no qual cada conexão apresenta um peso associado a fim de ajustar os *inputs* (entradas) aos *outputs* (saída) (SOBREIRO; ARAÚJO; NAGANO, 2009). Adicionalmente, é válido salientar que a camada localizada entre a camada de entrada e a de saída recebe a denominação de camada escondida (Figura 1).

Conforme a figura 1, a camada de entrada (*I*) fornece os *inputs* (dados ou informações) sem nenhum procedimento computacional. A camada oculta (*H*) recebe as informações da camada de insumos, ou de outra camada oculta, e fornece informações para outra camada oculta ou para a camada de saída. A camada de saída (*O*) recebe informações da camada oculta e fornece o produto da rede (BRESSAN, 2004).

Bressan (2004) evidencia que a estrutura das RNAs busca reproduzir a estrutura do neurônio humano, tendo, assim, três componentes básicos: um elemento somatório Σ , uma função de ativação f e as ponderações w_i . Conforme Sobreiro; Araújo; Nagano (2009), para o funcionamento das RNAs, cada neurônio da camada escondida deve receber um sinal de ativação que corresponde à soma ponderada dos dados da camada de entrada ou de outra camada, e processar pela função de ativação, fornecendo à próxima camada um novo sinal de ativação. O neurônio artificial soma os x_i *inputs* ponderados por w_i , que passa pela função de ativação f , resultando no produto y_j , que é o valor do neurônio j da camada escondida ou de saída. A ponderação dos sinais de ativação pode ser expressa matematicamente por $y_i = f(\sum_{i=0}^N x_i w_i)^p$.

Corrêa e Portugal (1998) enfatizam que na modelagem das RNA, a forma típica da função de ativação é $f = 1/(1 + e^{-z})$. Onde $z = \sum_{i=0}^N x_i w_i$; e é o número neperiano; x_i é o valor dos *inputs* da camada de entrada ou escondida; w_i são as ponderações. Gujarati (2006) argumenta que tal função é utilizada por respeitar a propriedade de seus valores dentro do intervalo de zero e um. Como pode ser observado, o somatório vai de 0 a N . Isso porque x_0 constitui o viés e assume o valor 1. O viés é adicionado para dar capacidade do neurônio em tender para um nível específico (BRESSAN, 2004). A função de erro para os neurônios de saída é definida por [3].

$$E(W) = 1/2 \sum_k (y_k - o_k)^2 \quad [3]$$

Onde $E(W)$ representa a função de erro; y_k é o valor desejado para a camada de saída k ; o_k representa o valor obtido na camada de saída k . A função de erro deverá ser minimizada para que as RNAs obtenham o melhor desempenho possível.

Como consequência, vários algoritmos têm sido desenvolvidos para minimizar o valor obtido pela função de erro. Conforme constatado na literatura especializada, o algoritmo *backpropagation*, pertencente à classe de técnicas de aprendizagem supervisionada, é o mais utilizado nas aplicações de previsão. Esse algoritmo permite realizar as transformações necessárias à transmissão dos dados para a rede, possibilitando a formação de previsões consistentes com o comportamento da série em estudo (FITZPATRICK, 2010).

Sendo assim, Sobreiro; Araújo; Nagano (2009) evidenciam que o algoritmo é compreendido como a realização de dois procedimentos: 1) *forward*, onde as RNAs são apresentadas aos dados de entradas, que se propagam camada por camada, mas nenhum peso da RNA é alterado; 2) *backward*, no qual os pesos são ajustados conforme a função de erro, ou seja, os resultados obtidos são subtraídos dos desejados, visando produzir os erros. Assim, os pesos da RNA são atualizados para que a resposta obtida se aproxime o máximo possível da desejada. Esse processo é repetido várias vezes com o propósito de se alcançar o menor erro possível.

Para aprendizado, a amostra é dividida em duas faixas. Na primeira, denominada treinamento, a rede capta o padrão de comportamento da série para, no teste, verificar se o erro é minimizado. O treinamento deve ser suficientemente grande para captar o padrão de comportamento da série, e o teste deve ser representativo das condições futuras que a rede vai tentar captar para realização da previsão (FITZPATRICK, 2010).

4 - ANÁLISE DOS RESULTADOS

A análise dos resultados obtidos por meio deste estudo será exposta da seguinte forma: 1) análise do modelo ARIMA estimado; 2) análise do modelo de RNAs estimado; e 3) com comparação entre os modelos ARIMA e de RNAs no

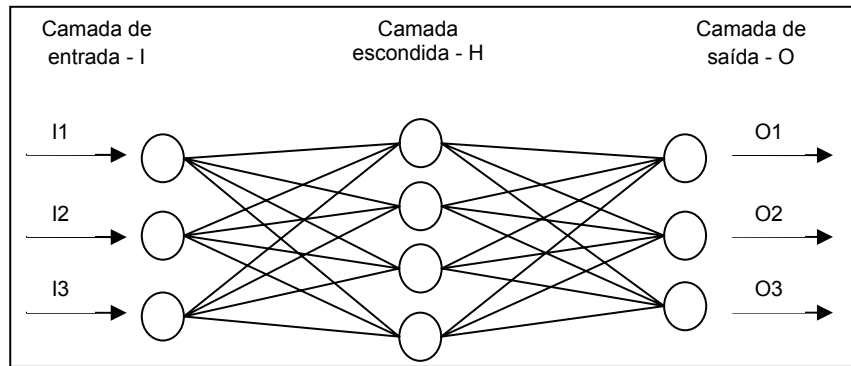


Figura 1 - Camadas de uma Rede Neural Artificial
Fonte: Wang; Elhag (2007).

que tange à qualidade do ajuste dos mesmos para a amostra pesquisada.

4.1 - Análise do Ajuste do Modelo ARIMA

Por meio de teste ADF de presença de raiz unitária, foi confirmada a não-estacionariedade da série de preços da soja. Dessa forma, a série precisou ser diferenciada uma vez, representando, assim, os retornos obtidos sobre os preços de negociação. A série diferenciada, no teste ADF de presença de raiz unitária, apresentou estacionariedade, o que confirmou a validade da aplicação da modelagem ARIMA. A análise da FAC e FACP indicou a utilização de um modelo ARIMA (2,1,0). A tabela 1 apresenta os coeficientes do modelo, bem como sua significância.

Por meio da tabela 1 é possível perceber que o retorno adquirido sobre o preço da soja é impactado em 0,282 pelo retorno obtido no dia de negociação anterior, pois o coeficiente autorregressivo apresentou elevada significância estatística. O coeficiente do parâmetro autorregressivo com dois dias de defasagem não exibiu significância ao nível de 5%.

Prosseguindo com a análise dos dados, a tabela 2 exhibe as estatísticas referentes ao teste *F*, representando a significância do modelo ARIMA estimado. Essa mesma tabela expõe, ainda, o coeficiente de determinação R^2 resultante da estimação do modelo ARIMA, que representa quão bem os parâmetros estimados se ajustaram aos dados da amostra estudada.

O teste *F* e o coeficiente de determina-

ção R^2 , apresentados na tabela 2, evidenciam que o modelo ARIMA estimado obteve significância estatística e um ajuste de aproximadamente 7,9% com relação à amostra pesquisada. O valor encontrado para o coeficiente de determinação R^2 não causa estranheza, uma vez que ajustes relativamente baixos são comuns em estudos que envolvem séries temporais financeiras, devido, em grande parte, ao fato de que o mercado financeiro, como um todo, apresenta determinado grau de eficiência, dificultando assim a previsão dos preços de ativos. A soja não é uma exceção.

4.2 - Análise do Ajuste das Redes Neurais Artificiais (RNAs)

Com objetivo de comparar diretamente os resultados de RNA com os do modelo ARIMA (2,1,0), a camada de entrada foi formada por duas variáveis explicativas: o retorno obtido sobre os preços de negociação da soja defasados em um e em dois dias (os mesmos do modelo ARIMA 2,1,0). A camada escondida foi formada por dois neurônios, sendo que o primeiro neurônio foi concebido para captar relações lineares, enquanto o segundo foi especificado para absorver relações quadráticas. A camada de saída foi formada por uma relação linear entre os neurônios da camada escondida anterior e o retorno obtido sobre os preços de negociação da soja no instante *t*. A figura 2 ilustra as relações que formaram a Rede Neural Artificial estimada.

Prosseguindo com a análise da Rede Neural Artificial estimado. A tabela 3 evidencia os

TABELA 1 - Coeficientes do modelo ARIMA (2,1,0) Estimado para a Série Temporal do Preço da Soja¹, Período 1997 a 2010

Parâmetro	Coeficiente	Erro padrão	Teste t	p-valor
Constante	0,015	0,021	0,727	0,467
ΔP_{t-1}	0,282	0,018	15,720	0,000
ΔP_{t-2}	0,001	0,018	0,057	0,954

¹Cotação diária da soja (60 kg) obtida em CEPEA (2010).

Fonte: Dados da pesquisa.

TABELA 2 - Teste F e R^2 do Modelo ARIMA (2,1,0), para a Série Temporal do Preço da Soja¹, Período 1997 a 2010

Item	Grau de Liberdade	Soma dos quadrados	Média dos quadrados	Teste F	p-valor
Regressão	2	370,499	185,250	134,502	0,000
Resíduo	3111	4284,787	1,377		
Total	3113	4655,286			
R^2	0,079				

¹Cotação diária da soja (60 kg) obtida em CEPEA (2010).

Fonte: Dados da pesquisa.

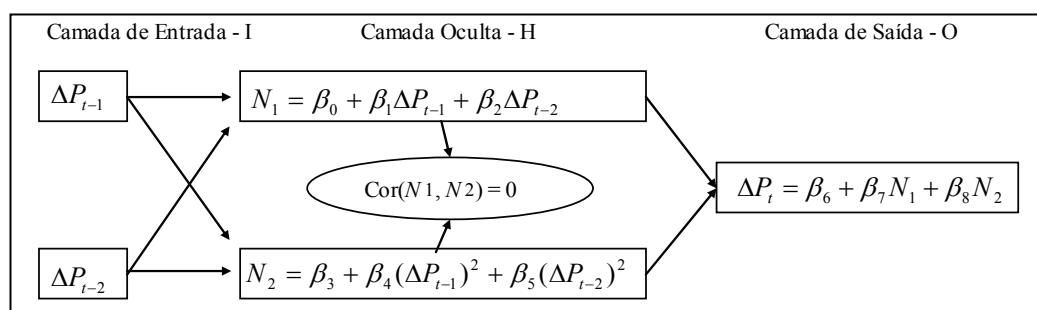


Figura 2 - Estrutura da Rede Neural Artificial Estimada, para a Série Temporal do Preço da Soja¹, Período 1997 a 2010

¹Cotação diária da soja (60 kg) obtida em CEPEA (2010).

Fonte: Dados da pesquisa.

coeficientes dos parâmetros das variáveis presentes na relação de formação do retorno obtido sobre os preços de negociação da soja no instante t (Constante, N_1 e N_2), bem como seus erros padrões, teste t , significância estatística, e impacto padronizado (BETA).

A partir da tabela 3, observa-se também que ambos os neurônios (N_1 e N_2) apresentaram impacto estatisticamente significativo na explicação do retorno obtido sobre os preços de negociação da soja no instante t . Entretanto, o primeiro neurônio (N_1), que captou as relações lineares, apresentou impacto 4,8 (0,082/0,017) vezes maior do que aquele proporcionado pelo segundo neurônio (N_2), que era responsável por absorver as relações quadráticas.

Corroborando a análise do modelo de Rede Neural Artificial estimado, a tabela 4 apre-

senta as estatísticas referentes ao teste F , representando a significância do modelo de RNAs estimadas. Não obstante, a mesma tabela exibe o coeficiente de determinação R^2 resultante da estimativa do modelo RNA, que representa quão bem os coeficientes estimados se ajustaram aos dados da amostra estudada.

Conforme os dados apresentados na tabela 4, o teste F e o coeficiente de determinação R^2 evidenciam que o modelo de Rede Neural Artificial estimado apresentou significância estatística e alcançou um ajuste de aproximadamente 8,3% com relação à amostra estudada. A validade de tais resultados fundamentou-se no fato de que os neurônios da RNA proposta foram construídos como sendo ortogonais, isto é, a correlação entre N_1 e N_2 é igual a zero. Como referido anteriormente, o valor encontrado é considerado

TABELA 3 - Coeficientes da RNA com um Neurônio Linear e com um Neurônio Quadrático na Camada Escondida

Parâmetro	Coeficiente	Erro padrão	Teste t	p-valor	BETA
Constante	-0,018	0,023	-0,760	0,047	-
N^1	0,074	0,004	16,434	0,000	0,082
N^2	1,742	0,449	3,881	0,000	0,017

Fonte: CEPEA (2010).

TABELA 4 - Teste F e R^2 do Modelo de RNAs

Item	Graus de liberdade	Soma dos quadrados	Média dos quadrados	Teste F	p-Valor
Regressão	2	390,843	195,421	142,564	0,000
Resíduo	3111	4264,443	1,371		
Total	31113	4655,286			
R^2	0,083				

Fonte: CEPEA (2010).

comum, devido à dificuldade de previsibilidade em séries temporais financeiras.

4.3 - Comparação dos Ajustes dos Modelos ARIMA e RNA

A comparação entre os dois modelos será feita corroborando as análises dos modelos ARIMA e RNA descritas anteriormente. Observou-se uma ligeira superioridade na qualidade do ajuste da RNA estimada, com 8,3% dos dados explicados contra 7,9% do modelo ARIMA. Essa análise também é válida sob a ótica da Soma dos Quadrados dos Resíduos, em que a RNA apresentou somatório de 4264,443 contra 4284,787 do ARIMA(2,1,0). Na tabela 5 é apresentado um conjunto de estatísticas auxiliares para comparação dos modelos.

No quesito ajuste dos modelos, os valores da tabela 5 destacam o modelo de RNA(2) como o de melhor qualidade em todos os critérios de informação e também no valor da verossimilhança, quanto menor, melhor. No diagnóstico dos resíduos após o ajuste, observa-se que tanto o modelo ARIMA (2,1,0) quanto RNA(2) obtiveram êxito em não apresentar dependência séria.

A superioridade no ajuste da RNA(2) pode ser explicada pelo fato de que os neurônios da camada escondida da RNA foram concebidos a captar tanto relações lineares como quadráticas do retorno obtido sobre os preços

de negociação da soja defasados em um e dois dias na tentativa de prever o retorno conquistado no instante t , enquanto na modelagem ARIMA a fundamentação compreende apenas relações lineares, e, portanto, é mais restritiva em relação à RNA.

4.4 - Estimativas Fora da Amostra dos Modelos ARIMA e RNA

Na figura 3 são apresentados graficamente, em forma de linha, os valores estimados pelos modelos AR(2,1,0) e RNA(2) 1 passo a frente. Pode-se perceber que as estimativas dos modelos são valores muito próximos entre si. A proximidade das estimativas indica que os modelos utilizados para realizar as estimativas capturaram informações similares e que foram bem ajustados aos dados.

Mais precisamente, foram utilizadas diversas métricas para realizar a comparação formal entre o modelo ARIMA(2,1,0) com modelo de Redes Neurais Artificiais RNA(2). Os resultados dessas métricas são apresentados na tabela 6.

Dentre as diversas métricas apresentadas na tabela 6, deve-se destacar que em termos de ME, MPE, MAPE o modelo RNA apresenta uma pequena vantagem; já para o MSE e para o MAE não há uma clara distinção entre os modelos. Porém, a métrica chamada U de Theil,

TABELA 5 - Métricas Auxiliares no Ajuste dos Modelos e Diagnóstico dos Resíduos

Ajuste dos modelos	ARIMA (2,1,0)	RNA (2)
Log da verossimilhança	-4,915,511	-4.908,101
Critério de Akaike	9,837,022	9.822,202
Critério de Schwarz	9,855,153	9.840,333
Critério Hannan-Quinn	9,843,531	9.828,711
Durbin-Watson	1,999	2,003
Diagnóstico dos resíduos (p-valor)		
Ljung-Box Q' (3 defasagens)	0,168 (0,983)	0,214 (0,975)
Ljung-Box Q' (6 defasagens)	8,167 (0,226)	5,833 (0,442)
Ljung-Box Q' (12 defasagens)	17,723 (0,124)	14,717 (0,257)

Fonte: CEPEA (2010).

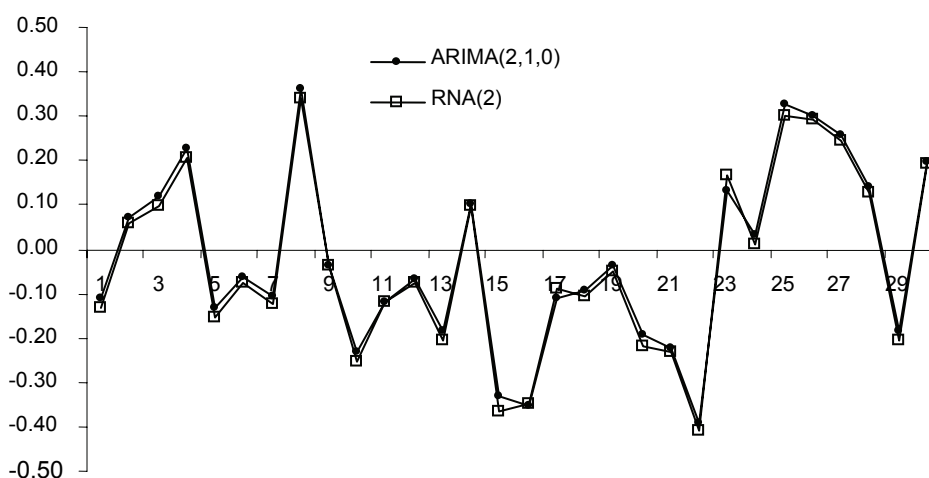


Figura 3 - Estimativas Fora da Amostra ARIMA(2,1,0) e RNA(2).
Fonte: CEPEA (2010).

TABELA 6 - Estatísticas de Avaliação das Estimativas Fora da Amostra, Valores Correspondentes a 30 Estimativas 1 Passo a Frente, Período 11/02/2010 a 26/03/2010

Estatística	ARIMA (2,1,0)	RNA (2)
Erro Médio (ME)	-0.100	-0.088
Erro Médio Quadrado (MSE)	0.472	0.473
Erro Médio Absoluto (MAE)	0.581	0.582
Erro Percentual Médio (MPE)	98.201	96.063
Erro Percentual Médio Absoluto (MAPE)	-19.393	-14.584
U deTheil (U)	1.019	1.033

Fonte: CEPEA (2010).

que avalia o desempenho das estimativas em relação a uma previsão ingênua a vantagem, fica com o modelo ARIMA(2,1,0). Deve-se destacar, também, que todos os indicadores são muito próximos entre si, e que, de certa forma, estão classificando as variações de preços da soja, como uma série imprevisível, pelo menos em períodos fora da

amostra considerado.

5 - CONCLUSÃO

Neste estudo, foram comparados dois modelos de ajuste e previsão de séries temporais,

visando à aproximação desses modelos ao comportamento do preço da soja no período que se estende de 1997 a 2010, num total de 3.144 observações. Essa série de preços, advindos dos ativos do mercado *spot* (à vista) da BM&FBovespa, foi objeto de comparação entre dois modelos quantitativos: o modelo ARIMA (Autorregressivo Integrado de Médias Móveis), por ser um modelo tradicional e comumente utilizado, e a Rede Neural Artificial (RNA), devido ao crescente interesse de pesquisadores de diversas áreas em compreendê-lo, e também por agregar critérios menos restritivos que outros modelos quantitativos. Como critério de comparação, visando analisar os ajustes adequados dos modelos à amostra apresentada, foi utilizado o coeficiente de determinação R^2 , além do teste F , para confirmar a significância dos modelos estimados.

Ao analisar os resultados obtidos, verificou-se no modelo ARIMA que o retorno adquirido sobre o preço da soja é impactado em 0,282 pelo retorno obtido no dia de negociação anterior, devido ao fato de que todas as informações a respeito da variação na diferença dos preços estejam incorporadas no preço da mesma, e, dessa forma, a variação do preço apenas é impactada pela informação do dia anterior. Isso confirma também a presença certo grau de eficiência na série de preços, dificultando, assim, a previsão, dando aos modelos ajustes muito baixos (7,9% do modelo ARIMA e 8,3% do modelo de RNA), mesmo com a significativa estimativa dos coeficientes dos parâmetros dos modelos propostos.

Não obstante, observou-se uma leve superioridade na qualidade do ajuste da RNA avaliada na comparação dos modelos, por captar dados não-lineares de forma menos limitativa que o tradicional modelo ARIMA, e assim sendo, este apresentou maior capacidade em configurar ajustes possíveis para a previsão da série temporal.

Além disso, o neurônio linear (N_1) contribuiu mais, na ordem de 0,082, se comparado com o neurônio quadrático (N_2), que obteve impacto de 0,017, no que tange ao ajuste do modelo de Redes Neurais Artificiais à amostra estudada.

Dentre as diversas métricas apresentadas na tabela 6, deve-se destacar que em termos de ME, MPE, MAPE, o modelo RNA apresenta uma pequena vantagem; já para o MSE e para o MAE não há uma clara distinção entre os modelos. Porém, a métrica chamada U de Theil, que avalia o desempenho das estimativas em relação a uma previsão ingênua a vantagem, fica com o modelo ARIMA(2,1,0), porém esse indicador deveria estar abaixo de 1 para indicar que os modelos têm desempenho melhor que uma previsão ingênua. Deve-se destacar, também, que todos os indicadores são muito próximos entre si, e que, de certa forma, estão classificando as variações de preços da soja como uma série imprevisível, pelo menos em períodos fora da amostra.

Porém, a parcial superioridade do modelo RNA não é confirmada nas estimativas fora da amostra, onde ambos os modelos apresentam desempenho sofrível se comparados com uma previsão ingênua. Portanto, ao seguir essa ideia de combinação e comparação de modelos para previsão de dados históricos, sugestionam-se possíveis comparações entre outros modelos quantitativos, ou mesmo comparações entre outros ativos de mesma natureza que a soja, já que o estudo comparativo entre modelos de previsão é ainda incipiente em *commodities* como o do objeto de estudo proposto. Não obstante, também é sugerida a aplicação de possíveis estudos que possam dar explicação a ajustes tão baixos dos modelos e que possibilite um melhor conhecimento do comportamento de ativos como o da *commodity* estudada.

LITERATURA CITADA

ABITANTE, K. G. Co-integração entre os mercados spot e futuro: evidências dos mercados de boi gordo e soja. **Rev. Econ. Sociol. Rural**, Brasília, v. 46, n. 1, p. 75-96, jan./mar. 2008.

BECERRA, K. **Mercado de capitais: Bolsa de Mercadorias e Futuros do Brasil (BM&F)**. 2007. 16p. Monografia (Pós-graduação em Contabilidade para Gestão de Negócios) - Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2007. Disponível em: <http://artigo cientifico.tebas.kinghost.net/uploads/artc_1199900017_64.pdf>. Acesso em: 02 abr. 2010.

BOX, G. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3. ed. New Jersey: Prentice Hall, 1994. 598p.

BRESSAN, A. A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. *RAE-eletrônica*, São Paulo, v. 3, n. 1, art. 9, jan./jun. 2004. Disponível em: <<http://www.scielo.br/pdf/raeel/v3n1/v3n1a04.pdf>>. Acesso em: abr. 2010.

CAO, Q.; LEGGIO, K. B.; SCHNIEDERJANS, M. J. A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. **Computers & Operations Research**, v. 32, issue 10, p. 2499-2512, oct. 2005.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA - CEPEA. **Indicador soja CEPEA/ESALQ**, 2010. Disponível em: <<http://www.cepea.esalq.usp.br/soja/>>. Acesso em: 05 maio 2010.

CHATFIELD, C. Model uncertainty and forecast accuracy. **Journal of Forecasting**, v. 15, issue 7, p. 495-508, dec. 1996.

CHURCH, K. B.; CURRAM, S. P. Forecasting consumers expenditure: a comparison between econometric and neural network models. **Journal of Forecasting**, v. 12, issue 2, p. 255-267, june 1996.

CO, H. C.; BOOSARAWONGSE, R. Forecasting Thailand's rice export: statistical techniques vs. artificial neural networks. **Computers & Industrial Engineering**, v. 53, issue 4, p. 610-627, nov. 2007.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Indicadores da agropecuária**. Brasília: CONAB, 2009. Disponível em: <<http://www.conab.gov.br/conabweb/IA-mar09.pdf>>. Acesso em: 29 abr. 2010.

CORRÊA, W. 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.

FOOD AND AGRICULTURE ORGANIZATION - FAO. **Major food and agricultural commodities and producers**. Disponível em: <<http://www.fao.org/es/ess/top/commodity.html?lang=en&item=236&year=2005>>. Acesso em: 29 abr. 2010.

FERNANDES, L. G. L.; PORTUGAL, M. S.; NAVAU, P. O. A. Previsão de séries de tempo: redes neurais artificiais e modelos estruturais. In: ENCONTRO BRASILEIRO DE ECONOMETRIA, 17., Salvador, 1995. **Anais...** Salvador: Soc. Bras. de Econometria, 1995.

FITZPATRICK, D. W. **Neural net primer: a brief introduction to the use of neural networks suitable for futures forecast**. Disponível em: <<http://www.jurikres.com/down/nnprimer.txt>>. Acesso em: 22 mai. 2010.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176p.

GUJARATI, D. N. **Econometria básica**. Rio de Janeiro: Elsevier, 2006. 840p.

INTER-AMERICAN INSTITUTE FOR COOPERATION ON AGRICULTURE - IICA. **Agroenergy and biofuels atlas of the Americas**, 2007. Disponível em: <<http://www.iica.int/Esp/organizacion/LTGC/agroenergia/Documentos%20Agroenergia%20y%20Biocombustibles/Agroenergia%20and%20Biofuels%20Atlas%20of%20the%20Americas.pdf>>. Acesso em: 29 abr. 2010.

KOHZADI, N. et al. A comparison of artificial neural network and time series models for forecasting commodity prices. **Neurocomputing**, v. 10, issue 2, p.169-181, 1996.

KUMAR, U. A. Comparison of neural networks and regression analysis: a new insight. **Expert Systems with Applications**, v. 29, issue 2, p. 424-430, 2005.

LATORRE, M. R. D. O; CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. **Revista Brasileira de Epidemiologia**, São Paulo, v. 4, n. 3, p.145-152, 2001.

MARGARIDO, M. A.; SOUZA, E. L. L. Formação de preços da soja no Brasil. **Agricultura em São Paulo**, São Paulo, v. 45, n. 2, p. 52-61, 1998.

MELO FILHO, G. A. de; RICHETTI, A. **Perfil socioeconômico e tecnológico dos produtores de soja e milho de Mato Grosso do Sul**. Dourados: Embrapa-CPAO, 1998. 57 p. (Documentos, 15).

MISSÃO, M. R. Soja: origem, classificação, utilização e uma visão abrangente do mercado. **Maringá Management: Revista de Ciências Empresariais**, Maringá, v. 3, n. 1, p. 7-15, jan./jun. 2006.

PINDYCK, R. S.; RUBENFIELD, D. L. **Econometric models and economic forecasts**. 3. ed. New York: McGrawHill, 1991. 565p.

PINO, R. et al. Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks. **Engineering Applications of Artificial Intelligence**, v. 21, issue 1, p.53-62, feb. 2008.

SILVA, W. V.; SANTO, E. L.; SILVA, L. S. C. Co-integração entre os preços da soja cotados nos mercados brasileiro e norte-americano: uma análise empírica, **Caderno de Pesquisas em Administração**, São Paulo, v. 10, n. 3, p. 69-78, jul/set 2003. Disponível em: <<http://www.ead.fea.usp.br/cad-pesq/arquivos/v10n3art7.pdf>>. Acesso em: 13 abr. 2010.

SOBREIRO, V. A.; ARAÚJO, P. H. S. L.; NAGANO, M. S. Precificação do etanol utilizando técnicas de redes neurais artificiais. **Revista de Administração**, São Paulo, v. 44, n. 1, p. 46-58, jan./fev./mar. 2009.

SOUZA, R. C. Modelos estruturais para previsão de séries temporais: abordagens clássica e bayesiana. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 17., Rio de Janeiro, 1989. **Anais...** Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1989.

TORRES FILHO, E. T. Petróleo: concorrência, regulação e estratégia. **Economia Política Internacional**, Campinas, n. 2, jul./set. 2004.

TSAY, R. S. **Analysis of financial time series**. 2. ed. New Jersey: John Wiley & Sons, 2005. 640p.

TSENG, F.; TZENG, G. A fuzzy seasonal ARIMA model for forecasting. **Fuzzy Sets and Systems**, v. 126, issue 3, p. 367-376, mar. 2002.

TURBAN, E. **Decision support and expert systems: management support systems**. 3. ed. New York: MacMillan, 1993. 833p.

VIEIRA, V. A. As tipologias, variações e características da pesquisa de marketing. **Revista da FAE**, Curitiba, v. 5, n. 1, p. 61-70, jan./abr. 2002.

WANG, Y.; ELHAG, T. M. S. A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. **Expert Systems with Applications**, v. 32, issue 2, p. 336-348, feb. 2007.

**PREVISÃO DO PREÇO DA SOJA: uma comparação
entre os modelos ARIMA e redes neurais artificiais**

RESUMO: O artigo compara o modelo ARIMA e as Redes Neurais Artificiais (RNAs) no que tange ao ajuste e previsão das variações de preço da soja. A amostra compreende os anos de 1997 a 2010, constituindo 3.144 cotações, que foram divididas para ajuste e previsão. Os resultados alcançados indicam que os retornos obtidos dependem do dia de negociação anterior, dando indícios de possível eficiência deste mercado. Não obstante, verificou-se ligeira superioridade do modelo de RNA no ajuste. Tal vantagem pode ser explicada pelo fato de as RNAs captarem relações não lineares. Entretanto, no que tange à previsão, a superioridade da RNA não se mantém.

Palavras-chave: soja, séries temporais, previsão.

**SOYBEAN PRICE FORECAST: a comparison between
the ARIMA and the artificial neural network model**

ABSTRACT: This article aims at comparing the ARIMA and the Artificial Neural Network (ANN) models in relation to the fit and forecasting performance of soybean prices. The sample covers the years 1997 to 2010 - 3144 price quotes in total - which have been split for fit and prediction. The results show that the returns obtained depend on the previous trading day, indicating a possible efficiency in this market. Nevertheless, a slight superiority of the ANN model is observed in the setting. This advantage can be explained by the fact that ANNs capture nonlinear relationships. However, with respect to prediction, the superiority of ANN is not maintained.

Key-words: soybean, time series, forecasting.

Recebido em 30/06/2010. Liberado para publicação em 19/8/2010.