

Previsão de preços de commodities agrícolas através do modelo de análise espectral singular multivariada

Carlos Alberto Orge Pinheiro
Universidade do Estado da Bahia (UNEB) e Faculdade de Tecnologia Senai Cimatec
Av. Silveira Martins, 2555 – Cabula
carlos.orge@terra.com.br

Valter de Senna
Faculdade de Tecnologia Senai Cimatec
Orlando Gomes, 1845 – Piatã
vsenna@terra.com.br

RESUMO

Esta pesquisa teve como proposta realizar previsões dos preços de um grupo de commodities através do modelo espectral de análise multivariada e compará-las com aquelas obtidas por modelos clássicos de previsão. A escolha das commodities etanol, boi gordo, milho, café e soja deu-se por conta do destaque na pauta de exportações no ano de 2013. O modelo espectral multivariado demonstrou-se adequado, quando comparado com os demais, ao permitir melhores desempenhos preditivos. Os resultados obtidos no período fora da amostra, mediante o uso das medidas de erro e do teste estatístico, confirmam isso. A pesquisa pode ajudar os profissionais do mercado na formulação e na aplicação de políticas direcionadas ao setor agrícola por conta da relevância da previsão dos preços como instrumento de planejamento bem como na análise do comportamento do mercado agrícola ao especificar a tendência dos preços.

Palavras-Chave: Análise espectral; previsão; commodities.

ABSTRACT

This study aimed to forecast the prices of a group of agricultural commodities through the multivariate spectral analysis model and compare them with those obtained by classical forecasting models. The choice of commodities such as ethanol, cattle, corn, coffee and soybeans was due to the emphasis in the exports in 2013. The multivariate spectral model has proved to be suitable, when compared with others, by enabling a better predictive performance. The results obtained in the out-of-sample period, through the use of measurement error and statistical test, confirm this. This research may help market professionals in formulating and implementing policies targeted to the agricultural sector due to the relevance of price forecasting as a planning instrument as well as in the analysis of the agricultural market behavior by specifying the price trend.

Keywords: Spectrum analysis; forecast; commodities.

1. Introdução

Nos últimos anos a agricultura no Brasil se desenvolveu e modernizou, destacando-se como atividade de alto valor econômico e social com tendência de crescimento acentuado (CAMPOS, 2007). O desenvolvimento dos mercados interno e externo contribuiu para o caráter dinâmico da agricultura, com a incorporação de tecnologias produtivas. Para Oranje (2003) o melhor desempenho da agricultura pode ser obtido através da competitividade dos preços, explicada, pela opção dos importadores em definirem entre os diferentes exportadores de produtos, aqueles que apresentam os menores preços.

A análise de preços, como elemento do mecanismo de troca, reveste-se de singular importância para os participantes do mercado agrícola, sejam eles compradores, vendedores ou especuladores. Então, as decisões tomadas pelos produtores, mesmo antes da realização da colheita, presumem o conhecimento do comportamento dos preços (RIBEIRO *et al.*, 2010). Além delas, as decisões referentes ao conjunto de produção e à adoção de políticas de financiamento como alternativa de assegurar preços, são baseadas em expectativas dos preços.

O principal problema sobre a previsão dos preços dos produtos agrícolas reside na sazonalidade que ocorre em razão dos fatores climáticos, mercadológicos e conjunturais que promovem incerteza de renda para os agricultores, armazenadores, exportadores bem como processadores desses produtos (MARTINS e MARTINELLI, 2010). Assim, a previsão de preços passa a atrair o interesse dos profissionais do mercado e dos acadêmicos uma vez que através dela é possível reduzir a incerteza no processo de decisão para aqueles que negociam no mercado de produtos agrícolas. Schwager (1995) explica que através da previsão é possível auxiliar os envolvidos nos mercados de commodities (mercadoria padronizada de baixo valor agregado) agrícolas tanto como os *hedgers*, que lidam com transações físicas de produtos, buscando proteção de futuras oscilações de preços. Logo, os participantes do mercado de commodities agrícolas necessitam de informações sobre os preços que são fundamentais para a rentabilidade de suas atividades. Pesquisas (Bressan, 2004; Lima *et al.*, 2010; Miranda Coronel e Pereira, 2013; Tibulo e Carli, 2014) sobre previsão de preços de commodities agrícolas focam no comportamento dos preços a partir da utilização de dados da commodity em análise. Portanto, utilizam modelos univariados de previsão que estão estruturados para realização de previsões de preços com base nos dados da série temporal em estudo. No trabalho realizado por Aguiar e Borestein (2012) os autores argumentam sobre a importância da utilização de outras informações que influenciam os preços da commodity butadieno, derivada do petróleo, com o intuito de monitorar as flutuações dos seus preços. A partir desta constatação o objetivo da pesquisa é o de realizar a previsão dos preços das commodities agrícolas através do modelo de Análise Espectral Singular Multivariada (AESM) por conta do modelo capturar as estruturas das séries temporais que representam o comportamento abrangente levando em consideração os efeitos do conjunto e avaliar seu desempenho preditivo. O artigo foi organizado da seguinte forma. Na próxima seção é apresentada a fundamentação teórica dos modelos aplicados para séries temporais com sazonalidade. Na seção 3 são descritos a metodologia além dos modelos de previsão utilizados na pesquisa. Na seção 4 os dados e os resultados dos testes de normalidade, normalidade multivariada, linearidade, estacionaridade e resultados empíricos são apresentados. As conclusões e sugestões para próximas pesquisas são expostos na seção 5.

2. Fundamentação teórica dos modelos de previsão

Na literatura de séries temporais é possível distinguir duas estratégias clássicas de modelagem. A primeira é considerada simples e refere-se aos modelos de suavização exponencial enquanto que a segunda caracteriza a metodologia de Box-Jenkins. Os modelos de suavização exponencial, também definidos como modelos de alisamento ou amortecimento, são técnicas desenvolvidas para um fim específico e que não requerem fundamentação probabilística. Utilizam a ideia de distribuição de pesos no decorrer do período, com o objetivo de considerar ponderações variantes no tempo. Dentre os modelos de suavização exponencial o algoritmo sazonal Holt-Winters (HW) é indicado para séries temporais com padrão de comportamento mais complexo, que além de apresentarem sazonalidade também consideram tendência e ruído (MORETTIN e TOLOI, 2006).

Levando em conta que, através da suavização, um processo de médias móveis é pouco intuitivo para representar o comportamento de uma particular série temporal e, considerando que a aplicação dos modelos autorregressivos (definido por valores anteriores e adição do ruído) é comum em diferentes áreas do conhecimento, pode-se utilizar os termos autorregressivos e de médias móveis simultaneamente com o objetivo de melhoria. Tal combinação caracteriza o modelo definido pela literatura como *Autoregressive Moving Average Model* (ARMA).

Outra possibilidade é tornar a série temporal estacionária através de um processo de diferenciação, ou seja, tomar diferenças sucessivas da série temporal original. Com isso, é dada a formação do modelo *Autoregressive Integrated Moving Average* (ARIMA). Esse modelo fundamenta-se na construção de métodos ajustados em suas propriedades probabilísticas.

Em algumas situações as séries temporais podem apresentar flutuações periódicas a exemplo dos fenômenos meteorológicos que, quando avaliados trimestralmente costumam apresentar correlações mais elevadas quando são usadas defasagens múltiplas de quatro, conforme as estações do ano (ESQUIVEL, 2012). Quando o modelo ARIMA leva em consideração essas flutuações ele passa a ser conhecido como *Seasonal Autoregressive Integrated Moving Average* (SARIMA).

Se para esses modelos clássicos o sinal (tendência e periodicidade) da série temporal é estudado em termos de unidades de tempo para os modelos espectrais a extração das informações sobre a série temporal é realizada em termos de unidades de frequência. A base dos modelos espectrais reside no fato de que qualquer função no tempo pode ser definida pela superposição de ondas senoidais de diferentes frequências. Na literatura os modelos espectrais quando comparados com os modelos clássicos, decompõem a série temporal em várias componentes com características de periodicidade mais simples apresentando vantagens na eliminação do ruído da série original, de acordo com Marques e Antunes (2009) como, para Vityazev *et al.*, (2010), investigam a estrutura das séries temporais com mais detalhes quando comparados com as Transformadas de Fourier e de Wavelet.

3. Abordagem metodológica

Com o objetivo de avaliar realizar a previsão dos preços das commodities agrícolas e verificar a contribuição do modelo AESM a pesquisa irá, inicialmente, identificar as características das séries temporais, através dos testes estatísticos para a normalidade, conforme Anderson-Darling (AD) e Shapiro-Wilk (SW) além de Doornik-Hansen-Omnibus (DHO) para a normalidade multivariada. Também serão utilizados os testes de McLeod e Li (1983) e Tsay (1986) para linearidade e de Dickey-Fuller Aumentado (DFA) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS) para a estacionariedade das séries. Em seguida o modelo AESM descrito na literatura e, a proposta de agrupamento ao modelo (Pinheiro e Senna, 2014) além dos modelos clássicos HW e SARIMA serão aplicados às séries temporais. O aspecto importante da pesquisa é poder confrontar as previsões, fora do período da amostra e, avaliar o desempenho preditivo através das medidas Erro Quadrático Médio (EQM) e Erro Quadrático Acumulado (EQA) e do teste estatístico preditivo proposto por Diebold e Mariano (1995), definido nesta pesquisa como teste DM. A seguir são descritos os modelos de previsão utilizados na pesquisa além das medidas de erro e do teste DM.

3.1 Modelo espectral de análise multivariada

Em relação ao modelo AESM suas primeiras pesquisas foram inicialmente feitas com dados atmosféricos. As séries temporais estavam associadas ao clima e representadas por localidades ou regiões num mapa (KEPPENNE e GHIL, 1993, PLAUT e VAUTARD, 1994). O modelo AESM é definido em dois estágios: decomposição e reconstrução. O estágio de decomposição é dado por dois passos: incorporação e decomposição em valor singular. A incorporação pode ser considerada como um mapeamento que transfere um conjunto M de séries temporais unidimensionais para uma matriz multidimensional.

No passo incorporação são definidas as matrizes trajetórias $X^{(i)}$, para $i = 1, \dots, M$, em cada série temporal $Y_t^{(i)}$, considerando que a matriz trajetória é uma sequência dos vetores defasados. O

resultado da incorporação, conforme descrito por Hassani e Mahmoudvand (2013), é a formação de um bloco de matrizes trajetórias X_V , conforme:

$$X_V = \begin{bmatrix} X^{(1)} \\ \vdots \\ X^{(M)} \end{bmatrix} \quad (01)$$

No passo decomposição em valor singular é realizada a decomposição da matriz $X_V X_V^T$ obtendo uma soma de matrizes elementares. Assim, denota-se por $\lambda_{V_1}, \dots, \lambda_{V_{M \times L}}$ os autovalores de $X_V X_V^T$ em ordem decrescente de magnitude ($\lambda_{V_1} \geq \dots \geq \lambda_{V_{M \times L}} \geq 0$) e por $U_{V_1}, \dots, U_{V_{M \times L}}$ os autovetores ortogonais. A matriz $X_V X_V^T$, é dada conforme:

$$X_V X_V^T = \begin{bmatrix} X^{(1)} X^{(1)T} & X^{(1)} X^{(2)T} & \dots & X^{(1)} X^{(M)T} \\ X^{(2)} X^{(1)T} & X^{(2)} X^{(2)T} & \dots & X^{(2)} X^{(M)T} \\ \vdots & \vdots & \ddots & \vdots \\ X^{(M)} X^{(1)T} & X^{(M)} X^{(2)T} & \dots & X^{(M)} X^{(M)T} \end{bmatrix} \quad (02)$$

A decomposição é representada por:

$$X_V = H_{V_1} + \dots + H_{V_D} \quad (03)$$

onde $H_{V_i} = \sqrt{\lambda_{V_i}} U_{V_i} V_{V_i}^T$ representa o bloco de matriz elementar, $V_{V_i} = X_V^T U_{V_i} / \sqrt{\lambda_{V_i}}$, o conjunto $\sqrt{\lambda_{V_i}}, U_{V_i}, V_{V_i}$ como autotripla e V_D o posto do bloco da matriz que corresponde ao número de autovalores não nulos.

No estágio de reconstrução, o passo agrupamento para o modelo AESM corresponde em dividir os blocos de matrizes elementares H_{V_1}, \dots, H_{V_D} em grupos disjuntos somando-os dentro de cada grupo. Assim, o desdobramento corresponde a representação:

$$X_V = H_{I_1} + \dots + H_{I_M} \quad (04)$$

onde H_{I_1}, \dots, H_{I_M} são definidos como bloco de matrizes resultantes (HASSANI e MAHMOUDVAND, 2013).

A ferramenta que auxilia na separação das componentes é o gráfico w -correlação cumulativa. Sua metodologia considera na definição da w -correlação $C(f)$ valores cumulativos, conforme explicam Patterson *et al.*, (2011). Com isso, a w -correlação $C(1)$ é definida com o primeiro conjunto autotripla como integrante da subsérie sinal $Y_{N_i}^{(s)}$ e os demais conjuntos de autotripas para formação da subsérie ruído $Y_{N_i}^{(n)}$. A w -correlação $C(2)$ é definida com o primeiro e segundo conjuntos de autotripas como integrantes da subsérie sinal $Y_{N_i}^{(s)}$ e os demais conjuntos para formação da subsérie ruído $Y_{N_i}^{(n)}$ e, assim sucessivamente. Essas w -correlações cumulativas são plotadas num gráfico, conforme Figura 1, adaptada de Patterson *et al.*, (2011). Logo, a existência da estrutura da série temporal está indicada por mínimos locais e máximos. Um padrão típico é um declínio para as w -correlações cumulativas e isso correspondente a uma separação das componentes sinal e ruído. Com base nisso, percebe-se na Figura 1 que a subsérie sinal $Y_{N_i}^{(s)}$ será dada pelos conjuntos 1-5 e, a subsérie ruído $Y_{N_i}^{(n)}$ pelos conjuntos 6-12, uma vez que $C(6)$ indica uma alteração deste declínio.

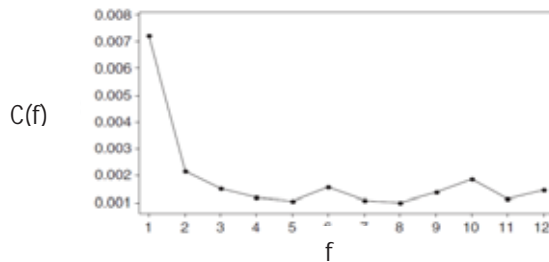


Figura 1 – Correlações cumulativas para as primeiras autotriplas

A proposta de agrupamento, conforme Pinheiro e Senna (2014), considera que a autocorrelação \hat{r}_q da subsérie ruído $Y_{N_i}^{(n)}$ é dada por:

$$\hat{r}_q = \frac{\sum_{t=q+1}^{N-1} \tilde{Y}_{N_i}^{(n)} \tilde{Y}_{N_i-t}^{(n)}}{\sum_{t=1}^{N-1} (\tilde{Y}_{N_i}^{(n)})^2} \quad (05)$$

com q representado o número de defasagens. Com base numa escolha binária das matrizes $\tilde{E}_1^{(i)}, \dots, \tilde{E}_d^{(i)}$, são formadas as subséries sinal e ruído reconstruídas pela média diagonal e dadas por $\tilde{Y}_{N_i}^{(s)} = (\tilde{y}_1^{(i)}, \dots, \tilde{y}_j^{(i)}, \dots, \tilde{y}_{N_i}^{(i)})$ e $\tilde{Y}_{N_i}^{(n)} = (\tilde{y}_1^{(i)}, \dots, \tilde{y}_j^{(i)}, \dots, \tilde{y}_{N_i}^{(i)})$. O processo segue conforme:

$$\begin{aligned} & \text{Min } \hat{r}_q \\ & \text{sujeito} \\ & \tilde{Y}_{N_i}^{(s)} = S_1 \tilde{E}_1^{(i)} + \dots + S_d \tilde{E}_d^{(i)} \\ & \tilde{Y}_{N_i}^{(n)} = N_1 \tilde{E}_1^{(i)} + \dots + N_d \tilde{E}_d^{(i)} \\ & \text{com} \\ & \text{se } S_1 = 0 \text{ então } N_1 = 1 \\ & \text{se } S_1 = 1 \text{ então } N_1 = 0 \\ & \dots \\ & \text{se } S_d = 0 \text{ então } N_d = 1 \\ & \text{se } S_d = 1 \text{ então } N_d = 0 \\ & e \\ & \mu_{\tilde{Y}_{N_i}^{(n)}} = 0 \end{aligned} \quad (06)$$

em que $\tilde{Y}_{N_i}^{(s)}$ representa a subsérie reconstruída sinal, $\tilde{Y}_{N_i}^{(n)}$ a subsérie reconstruída ruído e $S_1, N_1, \dots, S_d, N_d$ as variáveis de decisão binárias condicionais. Assim, determina-se a separabilidade entre sinal e ruído para o modelo, a partir de agora, definido como AESMP. As exigências $\text{Min } \hat{r}_q$ e $\mu_{\tilde{Y}_{N_i}^{(n)}} = 0$ em (06) são necessárias para que a subsérie ruído $\tilde{Y}_{N_i}^{(n)}$ seja um ruído branco.

3.1.1 Previsão do modelo espectral multivariado

A previsão obtida a partir de um conjunto M de séries temporais é dada para h passos à frente (HASSANI e MAHMOUDVAND, 2013):

$$\begin{bmatrix} \hat{y}_{j_1}^{(1)}, \dots, \hat{y}_{j_M}^{(M)} \end{bmatrix}^T = \begin{cases} \begin{bmatrix} \tilde{y}_{j_1}^{(1)}, \dots, \tilde{y}_{j_M}^{(M)} \end{bmatrix} & j_i = 1, \dots, N_i \\ (I_{M \times M} - WW^T)^{-1} W U^{\nabla M^T} Z_h & j_i = N_i + 1, \dots, N_i + h \end{cases} \quad (07)$$

com $U_j^{(i)\nabla}$ representando os primeiros $L_i - 1$ componentes do vetor $U_j^{(i)}$ e $\pi_j^{(i)}$ os últimos componentes do vetor $U_j^{(i)}$, com $(i = 1, \dots, M)$. A matriz $U_j^{\nabla(M)}$ é dada conforme:

$$U_j^{\nabla(M)} = \begin{bmatrix} U_j^{(1)\nabla} \\ \vdots \\ U_j^{(M)\nabla} \end{bmatrix} \quad (08)$$

e a matriz W representada por:

$$W = \begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} & \dots & \pi_r^{(1)} \\ \pi_1^{(2)} & \pi_2^{(2)} & \dots & \pi_r^{(2)} \\ \vdots & \vdots & \dots & \vdots \\ \pi_1^{(M)} & \pi_2^{(M)} & \dots & \pi_r^{(M)} \end{bmatrix} \quad (09)$$

Além disso, $Z_h = [Z_h^{(1)}, \dots, Z_h^{(M)}]^T$ e $Z_h^{(i)} = [\hat{y}_{Ni-Li+h+1}^{(i)}, \dots, \hat{y}_{Ni+h-1}^{(i)}]$.

3.2 Algoritmo sazonal HW

A incorporação das flutuações no algoritmo sazonal HW pode ser realizada através de dois enfoques distintos: sazonalidade multiplicativa e aditiva. Ao considerar a sazonalidade multiplicativa Morettin e Toloi (2006) explicam que série temporal pode ser definida por:

$$Y_t = N_t S_t + m_t + \varepsilon_t \quad (10)$$

com N_t o nível da série, S_t o fator sazonal, m_t o componente de tendência, ε_t o erro aleatório no instante t e $t = 1, \dots, N$. A forma de recorrência para o enfoque multiplicativo, nesta pesquisa, é dado por HW_m , com o fator de sazonalidade multiplicativo representado pelas equações que envolvem as três constantes de suavização, α , β e γ , conforme:

$$\begin{cases} N_t = \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)(N_{t-1} + m_{t-1}) \\ m_t = \beta(N_t - N_{t-1}) + (1-\beta)m_{t-1} \\ S_t = \gamma \frac{y_t}{N_t} + (1-\gamma)S_{t-s} \end{cases} \quad (11)$$

em que $0 < \alpha < 1$, $0 < \beta < 1$ e $0 < \gamma < 1$ são as condições das constantes de suavização do modelo e s representa o número de observações. As previsões para os valores futuros levam em conta os passos à frente h , de acordo com as equações:

$$\begin{cases} \hat{Y}_t(h) = (N_t + hm_t)S_{t+h-s}; & h = 1, 2, 3, \dots, s \\ \hat{Y}_t(h) = (N_t + hm_t)S_{t+h-2s}; & h = s+1, s+2, s+3, \dots, 2s \\ \vdots \end{cases} \quad (12)$$

Para o enfoque sazonalidade multiplicativa a correção dos erros e_t é dada:

$$\begin{cases} N_t = N_{t-1} + m_{t-1} + \alpha e_t / F_{t-s} \\ m_t = m_{t-1} + \alpha \beta e_t / F_{t-s} \\ S_t = S_{t-s} + \gamma(1-\alpha)e_t / N_t \end{cases} \quad (13)$$

O outro enfoque do método, dado nesta pesquisa por HW_a , é aplicado quando a série apresenta sazonalidade da forma aditiva. Para Morettin e Toloi (2006) ao tomar por base o fator sazonal aditivo a série temporal fica representada pela soma de todos os componentes conforme:

$$Y_t = N_t + m_t + S_t + \varepsilon_t \quad (14)$$

Na sazonalidade aditiva a forma de recorrência é dada pelas equações:

$$\begin{cases} N_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(N_{t-1} + m_{t-1}) \\ m_t = \beta(N_t - N_{t-1}) + (1-\beta)m_{t-1} \\ S_t = \gamma(Y_t - N_t) + (1-\gamma)S_{t-s} \end{cases} \quad (15)$$

com as mesmas condições das constantes de suavização do modelo para o enfoque multiplicativo bem como s representando o número de observações.

Os valores futuros são obtidos através das equações:

$$\begin{cases} \hat{Y}_t(h) = (N_t + hm_t) + S_{t+h-s}; & h = 1, 2, 3, \dots, s \\ \hat{Y}_t(h) = (N_t + hm_t) + S_{t+h-2s}; & h = s+1, s+2, s+3, \dots, 2s \\ \vdots \end{cases} \quad (16)$$

O procedimento de correção dos erros para este tipo de sazonalidade passa a ser dado por:

$$\begin{cases} N_t = N_{t-1} + m_{t-1} + \alpha e_t \\ m_t = m_{t-1} + \alpha \beta e_t \\ S_t = S_{t-s} + \gamma(1-\alpha)e_t \end{cases} \quad (17)$$

3.3 Modelo SARIMA

Em algumas situações é importante considerar a sazonalidade estocástica para explicar o comportamento sazonal da série temporal. A recomendação é que uma das variações do modelo ARIMA seja utilizado. Trata-se do modelo SARIMA. Para Box e Jenkins (1976) o modelo geral representado por ARIMA $(p,d,q) \times (P,D,Q)$ pode ser definido:

$$\phi(B)\Phi_P(B^s)\left[\Delta^d \Delta_s^D - \mu\right]Y_t = \theta(B)\Theta_Q(B^s)\varepsilon_t \quad (18)$$

com $\phi(B)$ o operador autoregressivo, Φ_P o polinômio estacionário autoregressivo sazonal de ordem P , Δ o operador diferença, μ o valor esperado da série, $\theta(B)$ o operador de médias móveis, Θ_Q o polinômio invertível de médias móveis sazonal de ordem Q e ε_t um erro aleatório.

O polinômio estacionário autoregressivo sazonal de ordem P é dado por:

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \quad (19)$$

O polinômio invertível de médias móveis sazonal de ordem Q é dado por:

$$\Theta_Q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs} \quad (20)$$

com o operador diferença sazonal de ordem D representado por:

$$\Delta_s^D = (1 - B^s)^D Y_t \quad (21)$$

em que de modo geral, a primeira diferenciação sazonal $\Delta_s^D = (1 - B^s)^D Y_t = Y_t - Y_{t-s}$ consegue excluir a sazonalidade da série temporal (ESQUIVEL, 2012).

3.4 Avaliação do desempenho das previsões

Como as previsões podem apresentar erros, independente do modelo adotado, costuma-se avaliar o resultado das previsões ao comparar os valores obtidos com os valores da série temporal original e, determinar o seu desempenho através de alguma medida. Então, na pesquisa, as previsões serão confrontadas com a 12ª semana subsequente a semana final da amostra. Para isso, a avaliação de desempenho faz uso da medida EQM definida por:

$$EQM = \frac{1}{h} \sum_{N+1}^{N+h} (Y_j - \hat{Y}_j)^2 \quad (10)$$

com Y_j representando o valor da série original, \hat{Y}_j o valor da previsão e h a quantidade de observações previstas e reservadas para avaliação. A pesquisa também faz uso da metodologia proposta por Goyal e Welch (2003), dada pela diferença entre os erros de previsão ao quadrado acumulados do modelo de melhor desempenho e os erros de previsão ao quadrado acumulados do modelo de melhor desempenho subsequente, considerando o EQA dado por:

$$EQA = \sum_{N+1}^{N+h} (Y_j - \hat{Y}_j)^2 \quad (11)$$

Sempre que esta diferença for positiva, o modelo de melhor desempenho subsequente supera o de melhor desempenho. Agora, ao considerar duas previsões de uma série temporal Y_t , e definindo e_{it} e e_{jt} como os respectivos erros de previsão, uma análise das perdas associadas a cada uma dessas previsões pode ser feita através do teste DM, que faz uso de uma função perda para medir o erro de previsão, ou seja, a perda é calculada a partir de valores realizados e previstos da variável em questão. Com isso, o teste DM verifica se a perda diferencial não é significativa entre duas previsões realizadas.

4. Resultados e discussões

4.1 Testes estatísticos aplicados sobre as séries temporais

O motivo da escolha das commodities açúcar, algodão, milho, café e soja não está somente no crescimento do seu volume de exportações nos últimos 5 anos, conforme dados disponibilizados pelo Ministério da Agricultura, Pecuária e Abastecimento (2013) mas, por desempenharem papel significativo na pauta das exportações brasileiras. Assim, em 2013 do total de exportações 42% correspondem aos produtos agrícolas. Relatórios do MAPA indicam que as commodities açúcar, algodão, milho, café e soja responderam por 13,72%, 7,11%, 8,35%, 7,50% e 34,97%,

respectivamente, desse total. As séries temporais preços das commodities agrícolas são identificadas na pesquisa conforme: AÇUC (açúcar), ALGO (algodão), MILH (milho), CAFE (café) e SOJA (soja). Os preços foram obtidos no banco de dados da Escola Superior de Agricultura Luiz de Queiroz (ESALQ) e correspondem ao período de 13 de janeiro de 2012 a 20 de dezembro de 2013, cuja periodicidade é semanal totalizando 96 semanas.

Para testar se a amostra de dados tem origem de uma população com uma distribuição específica os testes AD e SW são usados. Na pesquisa os dois testes permitem uma visão abrangente dos resultados. A partir dos resultados dispostos na Tabela 1, para um nível de 1% de significância, a série temporal CAFE está distribuída normalmente, enquanto que as demais séries temporais são exceções.

Tabela 1 – Teste de normalidade AD, SW e p -valor

	AÇUC	ALGO	MILH	CAFE	SOJA
Número de Observações	96	96	96	96	96
Shapiro-Wilk	0.96	0.94	0.91	0.92	0.95
p -valor	0.00	0.00	0.00	0.03	0.00
Anderson-Darling	1.05	2.02	2.95	0.83	1.87
p -valor	0.00	0.00	0.00	0.03	0.00

Fonte: Elaborado pelos autores

Em seguida, fez-se o uso do teste DHO que é um teste de normalidade multivariada aplicado entre pares formados pelas séries temporais. Os resultados apresentados na Tabela 2 indicam que existem fortes evidências de normalidade multivariada para um nível de 1% de significância entre os pares ALGO/CAFE e CAFE/SOJA. As exceções ocorrem para os demais pares. Na pesquisa seu uso justifica-se para conhecimento das características das séries temporais uma vez que o modelo espectral não pressupõe a hipótese de normalidade dos dados.

Tabela 2 – Teste de normalidade multivariada DHO e p -valor

	AÇUC	ALGO	MILH	CAFE	SOJA
AÇUC		19.06	34.27	22.02	19.49
p -valor		0.00	0.00	0.00	0.00
ALGO			27.67	8.35	16.38
p -valor			0.00	0.07	0.00
MILH				25.71	24.04
p -valor				0.00	0.00
CAFE					10.99
p -valor					0.03

Fonte: Elaborado pelos autores

Os testes de não linearidade de Tsay e de McLeod e Li também são aplicados nas séries temporais. Para o primeiro teste, antes de sua aplicação, os dados são filtrados por um modelo AR. A Tabela 3 apresenta os resultados para os dados da amostra com base em *lags* de 5 e 10 semanas. Para um nível de 1% de significância as séries temporais podem ser consideradas lineares.

Tabela 3 – p -valor para testes Tsay e McLeod do primeiro conjunto de séries reais

	Lags	AÇUC	ALGO	MILH	CAFE	SOJA
Número de Observações		96	96	96	96	96
Tsay	5					
p -valor		0.29	0.10	0.89	0.74	0.60

Tsay	10					
<i>p</i> -valor		0.77	0.30	0.58	0.10	0.57
McLeod	5					
<i>p</i> -valor		0.85	0.30	0.10	0.45	0.80
McLeod	10					
<i>p</i> -valor		0.90	0.25	0.30	0.80	0.90

Fonte: Elaborado pelos autores

A Tabela 4 apresenta os resultados dos testes DFA e KPSS. Para o primeiro a hipótese nula é que as séries temporais possuem raiz unitária e, portanto, não são estacionárias. Em relação ao segundo teste a hipótese nula é que as séries temporais não possuem raiz unitária e, portanto, são estacionárias. Desta forma, os testes confirmam que as séries temporais não são estacionárias. Em resumo, as séries temporais utilizadas na pesquisa não são estacionárias, são lineares e, não apresentam, de forma geral, normalidade multivariada.

Tabela 4 – Teste DFA e KPSS para série simulada e conjuntos de séries reais

Séries temporais	DFA	Valor Crítico 1%	KPSS	Valor Crítico 1%
AÇUC	-2,39	-4,05	0,26	0,21
ALGO	-2,26	-4,05	0,25	0,21
MILH	-2,39	-4,05	0,28	0,21
CAFÉ	-2,99	-4,05	0,28	0,21
SOJA	-2,28	-4,05	0,27	0,21

Fonte: Elaborado pelos autores

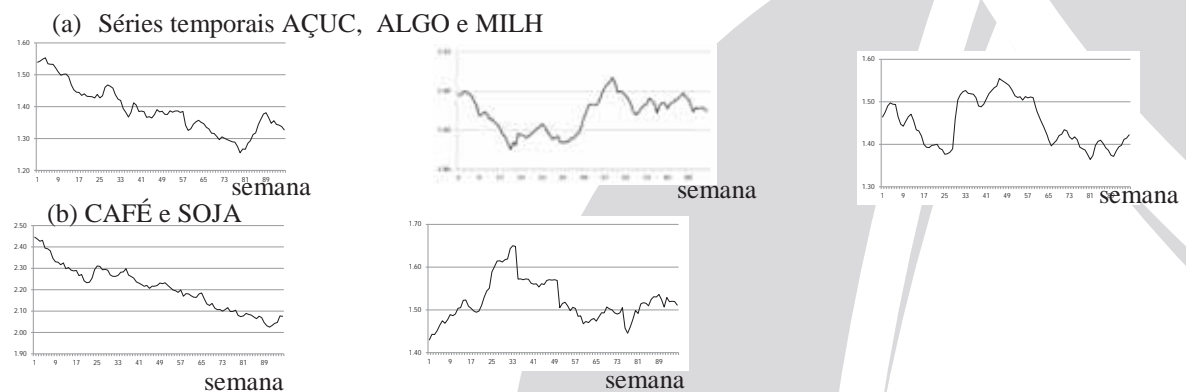


Figura 2 – Comportamento semanal dos preços das commodities destinadas à exportação

A Figura 2 apresenta o comportamento de cada série temporal no período de 13 de janeiro de 2012 a 20 de dezembro de 2013, representado pelas 96 semanas.

4.2 Aplicações empíricas

Nos modelos AESM e AESMP foram utilizados como comprimento da janela 16 semanas uma vez que este valor define o tamanho ótimo de janela. Observa-se na Tabela 5 os desempenhos das previsões, avaliados pela medida de erro EQM. As previsões obtidas pelo modelo AESMP quando comparadas com as previsões obtidas pelos demais modelos correspondem aos melhores desempenhos (MD na Tabela 5), por conta dos menores valores da medida de erro.

Tabela 5 – Desempenho de previsão pelo EQM

Série	L	h	EQM					MD
			AESMP	AESM	HWa	HWm	SARIMA	
AÇUC	16	12	1.3E-04	2.0E-03	1.2E-03	1.3E-03	1.3E-03	AESMP

ALGO	16	12	1.3E-04	4.1E-02	3.5E-02	3.3E-02	3.9E-02	AESMP
MILH	16	12	9.9E-04	2.3E-03	3.8E-03	4.0E-03	1.8E-03	AESMP
CAFE	16	12	1.5E-02	1.7E-02	1.4E-02	1.5E-02	7.9E-03	SARIMA
SOJA	16	12	4.0E-04	2.1E-03	1.6E-03	1.7E-03	4.9E-03	AESMP

Fonte: Elaborado pelos autores

Tabela 6 – Desempenho de previsão pela Diferença do EQA

Série	h	EQA		Diferença
		Modelos		
AÇUC	12	AESMP	HWa	-1.3E-02
		1.5E-03	1.4E-02	
ALGO	12	AESMP	HWm	-3.9E-00
		1.6E-03	3.9E-00	
MILH	12	AESMP	SARIMA	-1.0E-02
		1.0E-02	2.2E-02	
CAFE	12	SARIMA	HWa	-7.3E-02
		9.4E-02	1.7E-01	
SOJA	12	AESMP	HWa	-1.4E-02
		4.8E-03	1.9E-02	

Fonte: Elaborado pelos autores

Na Tabela 6, através da metodologia proposta por Goyal e Welch (2003), percebe-se que o modelo de melhor desempenho de previsão supera o de melhor desempenho subsequente uma vez que as diferenças apresentadas são negativas. Com o objetivo de avaliar se a diferença entre os EQMs do modelo de melhor desempenho com o modelo de melhor desempenho subsequente é estatisticamente significativa, aplicou-se o teste DM. Os resultados expostos na Tabela 7 indicam que, para os modelos comparados, a hipótese nula de que a diferença entre as medidas de erro é zero pode ser rejeitada ao nível estatístico 1% de significância para as séries temporais AÇUC, ALGO, MILH e SOJA. Em relação à série CAFE a hipótese nula de que a diferença entre as medidas de erro é zero não pode ser rejeitada ao mesmo nível de significância, não existindo superioridade do modelo de melhor desempenho de previsão.

Tabela 7 – Teste Diebold-Mariano e modelos comparados

Série	h	D-M	p-value	Modelos
AÇUC	12	5.07	0.00	AESMP, HWa
ALGO	12	11.24	0.00	AESMP, HWm
MILH	12	6.34	0.00	AESMP, SARIMA
CAFE	12	2.72	0.03	SARIMA, HWa
SOJA	12	4.75	0.00	AESMP, HWa

Fonte: Elaborado pelos autores

Então, pode-se concluir, com base no teste DM aplicado ao desempenho das previsões realizadas para o passo à frente h (12^a semanas), que o modelo AESMP agrega evidências favoráveis a sua aplicação aos preços das commodities agrícolas. Na Figura 3 são apresentados os comportamentos das séries temporais originais (linhas sólidas) e séries temporais previstas (linhas tracejadas) obtidos pelo modelo AESMP. Os gráficos foram elaborados no período de 27 de dezembro de 2012 a 14 de março de 2014, fora da amostra. Observa-se que o modelo AESMP,

no período, conseguiu detectar a tendência dos preços das commodities, com exceção para a série temporal CAFE.

(a) Série temporal AÇUC, ALGO, CAFÉ, SOJA E MILH

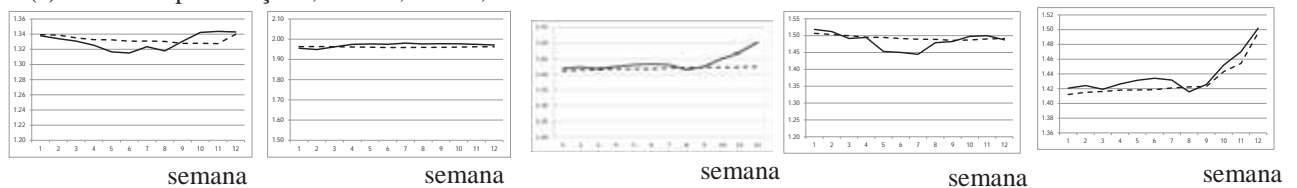


Figura 3 – Previsão para os preços das commodities destinadas à exportação

5. Conclusões e sugestões

A análise dos preços das commodities agrícolas, como elemento do mecanismo de troca, reveste-se de singular importância para os participantes do mercado, sejam eles compradores, vendedores ou especuladores. Assim, as decisões tomadas por estes participantes presumem o conhecimento do comportamento dos preços. Acontece que pesquisas realizadas sobre previsão de preços de commodities agrícolas são dadas pela análise empírica que observa o comportamento dos preços a partir da utilização de dados referentes apenas a commodity em estudo. Como a dinâmica das séries temporais de commodities agrícolas apresenta mudanças no tempo, por conta da sazonalidade, é preciso ter certeza de que o modelo de previsão não seja sensível a essas mudanças. A motivação pela utilização do modelo de análise multivariada dá-se por causa da sua capacidade de capturar estruturas que representem o comportamento mais abrangente e que levem em consideração os efeitos entre um conjunto de séries temporais. Os resultados dos testes estatísticos aplicados às séries temporais dos preços das commodities açúcar, algodão, café, milho e soja indicam que as séries não são estacionárias, são lineares e, não apresentam, de forma geral, normalidade multivariada. No contexto da pesquisa, os resultados empíricos demonstram que o modelo AESMP agregou evidências favoráveis a sua aplicação, uma vez que quando comparado com os modelos AESM, HWa, HWm e SARIMA, permitiu os melhores desempenhos das previsões. Os resultados obtidos no período fora da amostra, mediante o uso das medidas de erro EQM e EQA além do teste MD para o passo à frente h (12^a semana) confirmam isso. Do ponto de vista prático, os resultados obtidos podem ajudar os profissionais do mercado na formulação e na aplicação de políticas direcionadas ao setor agrícola por conta da relevância da previsão dos preços como instrumento de planejamento bem como na análise do comportamento do mercado agrícola ao especificar a tendência dos preços.

Para próximas pesquisas sugere-se a utilização de outras bases de dados, a inclusão de outras commodities agrícolas, a adoção de outros períodos de análise e o uso de outras variáveis que possam aumentar o poder explicativo do modelo.

6. Referências bibliográficas

- Aguiar, S. C. G. E. e Borestein, D. (2012). Redes bayesianas: uma ferramenta na previsão de preço de commodity. *Revista de Administração e Negócios da Amazônia*, 4, 237-253.
- Box, G. E. P. e JENKINS, G. M. (1976). *Time series analysis forecasting and control*. San Francisco: H. Day.
- Bressan, A. A. (2004). Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. *Revista de Administração Eletrônica*, 3, 1-20.
- Brockwell, P. J. e Davis R. A. (2002). *Introduction to Time Series and Forecasting*, 2nd edition. New York: Springer.
- Campos, K. C. (2007). Análise da volatilidade de preços de produtos agropecuários no Brasil. *Revista de Economia e Agronegócio*, 5, 303-328.
- Diebold, F. e Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253-265.

- Esquivel, R. M. (2012). *Análise espectral singular: modelagens de séries temporais através de estudos comparativos usando diferentes estratégias de previsão*. Dissertação de mestrado, Faculdade de Tecnologia Senai Cimatec, Salvador.
- Goyal, A. e Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49, 639–654.
- Hassani, H. e Mahmoudvand, R. (2013). Multivariate singular spectrum analysis: a general view and new vector forecasting approach. *International Journal of Energy and Statistics*, 1, 55-83.
- Keppenne, C. e Ghil, M. (1993). Adaptive filtering and prediction of noisy multivariate signals: An application to subannual variability in atmospheric angular momentum. *International Journal of Bifurcation and Chaos*, 3, 625-634.
- Lima, F. G., Kimura, H., Assaf Neto, A. e Perera, L. C. J. (2010). Previsão de preços de commodities com modelos ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias – novos resultados. *Revista de Administração*, 45, 188-202.
- Marques, J. e Antunes, S. (2009). A perigosidade natural da temperatura do ar em Portugal continental: a avaliação do risco na mortalidade. *V Encontro Nacional e I Congresso Internacional de Riscos*, 49-61.
- Martins, T. M. e Martinelli, D. P. (2010). Ciclos e previsão cíclica dos preços de commodities: um modelo indicador antecedente para commodity açúcar. *Revista de Administração, Contabilidade e Economia*, 2, 2-12.
- McLeod, A. I. e Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, 4, 169-176.
- Menezes, M. L., Cassiano, K. M., Souza, R. M., Teixeira Júnior, L. A., Pessanha, J. F. M. e Souza, R. C. (2014). Modelagem e previsão de demanda de energia com filtragem SSA. *Revista de Estatística UFOP*, 3, 170-187.
- Ministério da Agricultura, Pecuária e Abastecimento. (2013). Projeções do agronegócio 2013/14 a 2025/25. Assessoria de Gestão Estratégica. Brasília.
- Miranda, A. P.; Coronel, D. A.; Vieira, K. M. (2013). Previsão do mercado futuro do café arábica utilizando redes neurais e métodos econométricos. *Revista Estudos do CEPE*, 38, 66-98.
- Morettin, P. A. e Toloi, C. M. C. (2006). *Análise de séries temporais*. São Paulo: Blücher.
- Oliveira, V. A. e Aguiar, D. R. (2003). Determinantes do desempenho dos contratos futuros de commodities agropecuários no Brasil. Congresso Internacional de Economia e Gestão de Redes Agroalimentares. Ribeirão Preto: Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto.
- Oranje, M. (2003). *Competitividade das frutas brasileiras no comércio internacional*. Dissertação de Mestrado, Universidade Federal de Viçosa, Viçosa.
- Patterson, K., Hassani, H., Heravi, S. e Zhigljavsky, A. (2011). Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics*, v. 38, p. 2183-2211.
- Pinheiro, C. A. O. e Senna, Valter. (2014). Predição de índices acionários aos BRICS através de uma proposta de metodologia para a Análise Espectral Singular Multivariada. In: Simposio Brasileiro de Pesquisa Operacional, 2014, Salvador, Anais...
Salvador: XLVI Simpósio Brasileiro de Pesquisa Operacional, 2014.
- Plaut, G. e Vautard, R. (1994) Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. *Journal of the Atmospheric Sciences*, 51, 210-236.
- Ribeiro, C. O., Sosnoski, A. A. K. e Oliveira, S. M. (2010). Um modelo hierárquico para previsão de preços de commodities agrícolas. *Revista Produção On-line*, 10, 719-733.
- Schwager, J. D. (1995). *Fundamental analysis*. New York: John Wiley & Sons.
- Tibulo, C. e Carli, V. (2014). Previsão do preço do milho através de séries temporais. *Scientia Plena*, 10, 2-10.
- Tsay, R. (1986). Non-linearity tests for time series. *Biometrika*, 73, 461-466.
- Vityazev, V., Miller, N. e Prudnikova, E. J. (2010). Singular spectrum analysis in astrometry and geodynamics. *AIP Conference Proceedings*, 1283, 317-326.