Recebido: 24/05/11
Aprovado: 04/07/12

A study about the performance of time series models for the analysis of agricultural prices

Dra. Sandra Cristina de Oliveira (UNESP - SP/Brazil) - sandra@tupa.unesp.br,
• Av. Domingos da Costa Lopes, 780, Jardim Unesp, 17602-496, Tupã-SP, fone: (55) 14-3404- 4200
Leonardo Matheus Marcondes Pereira (UNESP - SP/Brazil) - leonardommp@gmail.com
Juliana Tiemi Sakaguchi Hanashiro (UNESP - SP/Brazil) - jhanashiro@bol.com.br
Patricia Carvalho do Val (UNESP – SP/Brazil) - patty.doval@gmail.com

ABSTRACT *Agribusiness is one of the most relevant achievements in Brazilian and world economies. Price is an important variable within this sector since economic and market conditions vary over time. Efficient modeling methods are needed to describe better the trends and the characteristics of that variable. The performance of time series techniques of fixed and open models for analysis of agricultural prices will be provided to demonstrate the feasibility of techniques for the generation of results in the wake of economic decisions. Historical time series of monthly average price of cash crops, such as groundnut, sugarcane, banana and orange, received by Brazilian producers, were analyzed for the period between August 1994 and December 2009 and updated by the General Price Index - Internal Availability. Current study shows that exponential smoothing models and Box & Jenkins modeling are viable alternatives for the adjustment of the above price series, each one with its specific characteristics.*

Keywords *Statistical Models; Agricultural Marketing; Resource Efficiency; Management; Decision Taking.*

OLIVEIRA, S. C.; PEREIRA, L. M. M.; HANASHIRO, J. T. S.; VAL, P. C. A study about the performance of time series models for the analysis of agricultural prices. **GEPROS. Gestão da Produção, Operações e Sistemas**, Ano 7, nº 3, jul-set/2012, p. 11-27.

1. INTRODUCTION

Brazil's economy has full favorable conditions to increase world food production since agribusiness is one of the most important activities, with approximately 25% share of the national economy (CEPEA, 2010). Further, agribusiness engages almost 40% of the labor force and 36% of Brazilian exports (MENDES; PADILHA JR., 2007).

The four main agricultural commodities, namely, groundnut, sugarcane, banana and orange, will be discussed in current research owing to the relevant role they have in food supply and in the generation of labor and income for the population, specifically for that of the state of São Paulo.

The state of São Paulo, Brazil, is the greatest producer of groundnut in Brazil, with approximately 80% of the national production. Alta Mogiana and Alta Paulista are the most prominent regions with significant production amounts (MARTINS, 2006).

High productivity has been the result of technology that brought about creeping-type cultivars and planting-harvest mechanization. The above-mentioned technologies are mainly used in the state of São Paulo and in the savannah region where groundnut is alternately cultivated with sugarcane without the excessive use of fertilizers.

The sugarcane-alcohol sector in Brazil provides real advantages when compared to those furnished by sugar and alcohol-producing world competitors, such as the United States, China and the European Community. The main advantage, however, is that Brazil has the lowest sugarcane production costs in the world, due to climate and soil properties. The Brazilian sugar-alcohol agro-industry contracts almost a million laborers, half of which work on the fields and the other half in the sugar-alcohol plants. The state of São Paulo produces 60% of the sugarcane in Brazil, with 400,000 directly engaged employees (MDIC, 2010).

Brazil is also highly important in fruit crop and the country is the third biggest producer of fresh fruits worldwide. This is mainly highlighted by increasingly high quality Brazilian fruits. Banana and orange crops reached almost 12% of the total value in agribusiness produce of the state of São Paulo, with the largest (80%) volume of orange production. Further, Brazil has almost 60% world production in orange juice and more than 80% of the world market (IEA, 2010).

The Brazilian Ministry of Agriculture, Livestock and Supply (MAPA) currently coordinates the Integrated Fruit Production Program (PIF) that comprises a list of twenty-one products within the fruit crop sector, including banana and orange. These activities aim at replacing conventional practices by a process that would decrease production costs, improve quality, reduce environment degradation and better the national and foreign consumption of fruit produced in Brazil (SANTINI; OLIVEIRA; PIGATTO, 2010).

In the wake of the above characteristics and within the economic scenario of groundnut, sugarcane, banana and orange crops, studies related to the inherent variables of these crops and of others involving agribusiness are highly relevant.

Since price is an important and decisive variable for the rural producer and for agribusiness, understanding the possibilities of its analysis is of paramount important. It also contributes in planning and decision-taking for the products' commercialization.

Current research shows several technical statistics of time series analysis (fixed and open models) to verify their feasibility in forecasting mean monthly prices of groundnut, sugarcane, banana and orange received by the Brazilian producer as from a historical price series between August 1994 and December 2009 and updated by General Price Index - Internal Availability (IGP-DI).

Evaluations and discussions are available on what type of techniques provided the best estimates of average monthly price of the above-mentioned crops by methods that informed the exactness of these forecasting techniques.

2. LITERATURE REVIEW

Currently, statistical techniques of time series analysis have been widely disseminated in the literature, and there is a great variety of circumstances and researches in which they can be used, especially in studies involving economic decisions, such as: national and international economies behavior, stock market, price forecast, supply and demand for agricultural products, among others.

Barbosa (2005) used the simple average, simple moving average and exponential smoothing models, seeking the one that best suited to the alcohol demand behavior, observing a historical series of data from January 1980 to December 1995. The most appropriate model was the simple moving average, as it presented lower mean deviation (difference between the actual value and the predicted value) and therefore greater precision.

Freire (2005) investigated the exponential smoothing and multiple linear regression models to verify which of them would best exhibit accounting data to perform the sales revenue forecasting from a power company. It was considered a historical series of data from January 1998 to December 2004 and, in this study, both methods showed statistically acceptable degrees of accuracy and significance.

Carvalho (2005) used different models in order to make a prediction of short and medium term for the arrivals flow at the Belém (PA) bus station, among them, the autoregressive integrated moving average (ARIMA), classical decomposition and exponential smoothing models. The models that presented smaller deviations for the data series from January 1999 to December 2003 were: Winter's triple exponential smoothing and additive classical decomposition, respectively.

Medeiros *et al.* (2006) analyzed the evolution of the prices received by commodities of sugarcane, sugar and ethanol hydrated from February 1999 to June 2005. For such, the moving average, exponential smoothing, classical decomposition, Fourier series, ARIMA models and multiple linear regression methods were used. For the three products studied, the best model identified with the smallest error was Winter's triple exponential smoothing in which: the additive model for sugarcane and the multiplicative model for sugar and alcohol hydrate.

Duarte (2007), analyzing the market for soybeans, used exponential smoothing and ARIMA models to predict the price of the product, and thus use the best prediction to guide the daily negotiation for soybeans. The ARIMA models presented results closest to the actual values (daily prices from January 1998 to March 2007) and proved to be able to help in the negotiations of this commodity, followed by Winters' triple exponential smoothing methods.

Klidzio *et al.* (2008) investigated which model best explains the crude soybean oil weekly acidity index series dynamic from April 2004 to September 2008. ARIMA and Seasonal ARIMA (SARIMA) models were compared. It was concluded that the most appropriate model to represent the variable behavior was ARIMA model, since it provided the knowledge of future values for the acidity index more precisely.

In a more comprehensive manner, Makridakis, Wheelwright & Hyndman (1998) investigated the predictive ability of various methods of time series forecasting and reported that the ones of exponential smoothing are viable and facilitate the interpretation and data analysis, but they can generate correlated prediction errors compromising the long term prediction. However, Pindyck & Rubinfeld (2004) stated that simple forecasting methods can provide very satisfactory performance under certain conditions and that the adoption of a more complex forecasting method should be investigated.

Therefore, due to its wide applicability and relevance, the importance of a thorough study on the performance of time series statistical techniques of fixed and open models (in the generation of values within the sample period and in forecasting future values) is confirmed, specifically in the agricultural prices analysis context.

3. MATERIALS AND METHODS

3.1. Data collection

Information by the Getulio Vargas Foundation (FGV) and available at the database of the Institute for Applied Economical research (IPEA) has been employed to develop current research. The historical series of the General Price Index - Internal Availability (IGP-DI) in August 1994, and of monthly mean prices (in Brazilian Real) for groundnut (kg), sugarcane (ton), banana (kg) and orange (hundred), received by Brazilian producers between August 1994 (when the Real Monetary Plan was executed) and December 2009, were taken into consideration (IPEADATA, 2010).

The period between August 1994 and December 2008, comprising 174 observations, was taken into consideration for the adjustments of the proposed models, whereas the period between January and December 2009 was employed for forecast verification.

Due to constant devaluation of Brazilian money during years of high inflation, the correction of prices of agricultural and livestock products had to be corrected for a given period, or rather, real prices or money at current value had to be calculated to interpret the variations which occurred with the variable throughout a certain time period. As a rule, real prices are obtained by deflating the nominal prices (with inflation) by a general price index or by any other accumulated inflation index. Real prices (without inflation) make possible the analysis of variation of the variable over time that is not due to inflation.

Prices in current research have been deflated by IGP-DI, or rather, the weighted average of three indexes: Wholesale Price Index (IPA), Consumer Price Index (IPC) and the Civil Construction Costs National Index (INCC), respectively with weights 6, 3 and 1. Although IGP-DI is not the ideal inflation index for price behavior in agribusiness, it has conventionally been adopted for many years as a measure of Brazilian inflation and thus it has acquired special importance and wide acceptance in economical analyses (MENDES; PADILHA JR., 2007).

So that data collection would be standardized, nominal prices (NP) were duly updated for real prices (RP) by the formula below:

$$RP = ([IGP\_DI_{\text{basis month}}] \div [IGP\_DI_{\text{other months}}] \times NP \quad (1)$$

3.2. Data analysis

Techniques of time series analysis are based on the identification of standards (or components) in historical series so that forecasts on a given variability could be given. The main components (or movements) of a time series comprise trend, seasonality, cycle and randomness.

Fixed model techniques may be easily executed and employed since they present equations based on a priori evaluations of determined components in historical series. They also adjust themselves quickly to changes in agricultural and livestock price behavior. The above techniques do not require a very long series and are convenient in short and medium term forecasts. Simple average, simple moving average, simple exponential smoothing, Holt's double exponential smoothing and Winter's triple exponential smoothing, described in Table 1, are the main techniques for the fixed model time series.

Symbols in the expressions (a) to (e) in Table 1 comprise: P_{t+p}: forecast for period p (when p = 1, forecast is made for the next period); R_t: real price during period t; n: number of information of price series; M_t: moving average during period t; k: number of periods given in the moving average; α: damping coefficient that subtracts randomness from the series ($0 \leq \alpha \leq 1$); β: damping coefficient for trend estimates ($0 \leq \beta \leq 1$); γ: damping coefficient for seasonality estimates ($0 \leq \gamma \leq 1$); N_t: level component; T_t: trend component; S_t: seasonal component.

Table 1 – Time series techniques of fixed model.

(a) simple average (SA)	$P_{t+1} = \sum_{t=1}^{n} R_t \Big/ n$
(b) simple moving average (SMA)	$P_{t+1} = (R_t + R_{t-1} + R_{t-2} + \ldots + R_{t-k+1})/k$
(c) simple exponential smoothing (SES)	$P_{t+1} = \alpha R_t + (1-\alpha)P_t$
(d) Holt's double exponential smoothing (DES)	$N_t = \alpha R_t + (1-\alpha)(N_{t-1} + T_{t-1})$ $T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1}$ $P_{t+p} = N_t + pT_t$
(e) Winter's triple exponential smoothing (TES)	$S_t = \gamma(R_t/N_t) + (1-\gamma)S_{t-c}$ $T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1}$ $N_t = \alpha(R_t/S_{t-c}) + (1-\alpha)(N_{t-1} + T_{t-1})$ $P_{t+p} = (N_t + pT_t)S_{t-c+p}$

Source: Box, Jenkins & Heinsel (1994), modified.

The exactness of the simple moving average technique is associated to coefficient k and the precision of exponential smoothing techniques is conditioned to the determination of damping coefficients α, β and γ. Consequently, values of these coefficients should be chosen to minimize forecast errors and allow a precision that would be satisfying (MORETTIN; TOLOI, 2004). Minitab software, which provides the best results for these coefficients, based on the monitoring of forecast errors, selected damping coefficients α and/or β and/or γ (ranging from 0 to 1) in current research.

Open model techniques analyze historical price series to identify or investigate which components are present so that a model that would project these components by forecasting the variable's future values could be established. Classical decomposition and Box & Jenkins parameter modeling are the main techniques of the time series of the existing open model.

Classical decomposition technique comprises the decomposition of the historical price series into its non-observable time components with the highest preciseness possible: trend (T), seasonality (S), cycle (C) and randomness (I). After the analysis of each component, forecasting (F) is calculated by recomposing the series, according to the formula below (SPIEGEL, 1994):

(f) Addition Classical Decomposition (ACD)

$$F = T + S + C + I \tag{2}$$

(g) Multiplication Classical Decomposition (MCD)

$$F = T \times S \times C \times I \tag{3}$$

Box & Jenkins modeling or autoregressive moving average models of orders p and q, denoted by ARMA(p,q), start from the concept that time series involved in the analysis are produced by a stationary process. They adjust the data of such series as a linear combination of past values and/or errors by employing the autoregressive moving average processes and a residual correlation structure. According to Box, Jenkins & Heinsel (1994), ARMA(p,q) models are given by:

$$R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + \ldots + \phi_p R_{t-p} + a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q R_{t-q} \tag{4}$$

or, in an equivalent way, $\phi(B)R_t = \theta(B)a_t$, where R_t is the real price of the crop during the period t, $t = 1, 2,\ldots, n$, $\phi(B) = (1 - B\phi_1 - B^2\phi_2 - \ldots B^p\phi_p)$ and $\theta(B) = (1 - B\theta_1 - B^2\theta_2 - \ldots B^q\theta_q)$.

Terms $\phi(B)$ and $\theta(B)$ are polynomial functions of B (delay operator, such that $BR_t = R_{t-1}$ and $Ba_t = a_{t-1}$) and which represent price the autoregressive component and the moving average component respectively. These processes are supposedly stationary and invertible, or rather, the roots of $\phi(B)$ and $\theta(B)$ are outside the unitary radius circle, and that $\{a_t, t \in 0\ Z\}$ is a white noise (normally distributed randomized variable, with zero as average and constant variance σ^2), where a_t is still a non co-related variable, or rather, $Cov(a_t,a_s) = 0$ for $t \neq s$ (BOX; JENKINS; HEINSEL, 1994).

When series are not stationary, one or two differentiations transform them into being so. Differentiated series are called order d integrated processes and denoted by ARIMA(p,d,q), where d represents the number of differentiations undertaken until the series becomes stationary.

As a rule, the strategy for the construction of the model consists in an iterative cycle in which the choice for model structure is based on its own data, with three stages: identification, estimate, diagnosis and forecasting. More details may be obtained from Box, Jenkins & Heinsel (1994), Morettin & Toloi (2004) and Morettin (2008).

Although autoregressive parameters of ARMA(p,q) models are linear with simple methods for estimates, the parameters of moving averages are not linear and, as a consequence, the estimate process becomes more complex. The maximum likelihood method, an extremely current estimation technique in statistical inference, should be opted for.

If the model is correctly identified, the likelihood principle foregrounds that all information that data provide on the parameters are contained within the likelihood function. The method selects the estimators (of the model's parameters) that maximize the probability for obtaining the sample under analysis (SOUZA; CAMARGO, 1996).

The likelihood function (L) of an ARMA(p,q) model is given by:

$$L\,(R_1, R_2,..., R_n \mid \varphi, \theta, \tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}\sum_{t=1}^{n}(R_t - \phi_1 R_{t-1} - ... - \phi_p R_{t-p} + \theta_1 a_{t-1} + ... + \theta_q a_{t-q})^2\right\} \qquad (5)$$

where R_t is the real price of the crop during period t ($t = 1, 2,..., n$), $\tau = \frac{1}{\sigma^2}$, $\tau > 0$, $\varphi = (\phi_1, \phi_2,..., \phi_p)$ and $\theta = (\theta_1, \theta_2,..., \theta_q)$ are the parameters of an ARMA(p,q) model to be estimated (SÁFADI; ANDRADE, 2007). Estimates of parameters should be obtained by some sort of numerical procedure. Estimates in current research have been obtained by procedure available by Minitab.

Methodological procedures for the analysis of time series by Box & Jenkins models are given below:

Identification

1) Verification of estimated samplings' autocorrelation functions (ACF) and partial autocorrelation functions (PACF) of the crop's real prices series to analyze visually the stationarity condition and to have indications of the model and the order to be adjusted; testing of the hypothesis of stationarity by test KPSS (proposed by Kwiatkowski *et al.* 1992). It may be concluded that the series is not stationary for the statistics of test > critical value (for risk α). Since differentiations in the series are not required, the models ARMA(p,q) or ARIMA(p,0,q) should be adjusted to the data; otherwise, required differentiations have to be worked out and ARIMA(p,d,q) models adjusted;
2) According to the above, several competitive models with different p and q orders are estimated and the model with the lowest values for Akaike Information Criterion – AIC (SCHWARZ, 1978) should be chosen, namely,

$$\text{AIC} = -2l + 2k \qquad (6)$$

in which k is the total number of estimated parameters; N is the number of observations for each series of real prices under analysis and l is the mean of natural logarithm of the likelihood function defined by equation (5).

Estimates

1) When the provisional model for the price series is identified, the parameters of the model should be estimated by the highest likelihood method;
2) Verification of each parameter's significance by test t. In the case of p-value $\leq \alpha$, it may be concluded that (for risk α) the associated parameter may be taken as significantly different from zero[1].

Diagnosis

1) When the final adjustment model to the real price series is identified, the residual autocorrelation should be verified by a graphic analysis of ACF and PACF;
2) Ljung-Box[2] test (proposed by Ljung & Box, 1978) should be performed for estimated residues. In the case of p-value $\leq \alpha$, it may be concluded (for risk α) that residues are not white noise, or rather, they are auto correlated (MORETTIN, 2008).

Forecasting

The performance of fixed and open models in the adjustment of mean monthly price of groundnut, sugarcane, banana and orange has been verified by the exactness (or adequacy of adjustment) given by the application of each previously mentioned technique, (a)-(h). The technique with the lowest error is chosen (the difference between real values and values adjusted to each model). Two statistically standard measures were taken into account to generate values within the sampling period: mean absolute deviation (MAD), i.e. MAD $= \sum_{t=1}^{n} |R_t - P_t| \Big/ n$ and mean squared error (MSE), i.e. MSE $= \sum_{t=1}^{n} (R_t - P_t)^2 \Big/ n$; and a relative measure: mean absolute percentage error (MAPE), i.e. MAPE $= \sum_{t=1}^{n} \frac{|R_t - P_t|}{R_t} \Big/ n \times 100$; where R_t is the real price during period t; P_t is the adjusted rate for period t; n is the amount of information of the price series (MORETTIN; TOLOI, 2004).

The variable's future values (forecasts of the crops' mean monthly price) were obtained for twelve months (from January to December 2009) from tested models with the best performance. They were then compared to real values by a graphic analysis and mean absolute percentage error. A relative measure during the forecasting stage is recommended since it provides a more simple interpretation of results in percentage (MORETTIN, 2008; GONÇALVES, 2003).

It should be emphasized that the Minitab (mentioned above), Gretl and Microsoft Excel were employed for tests, analyses, estimates and modeling of data.

---

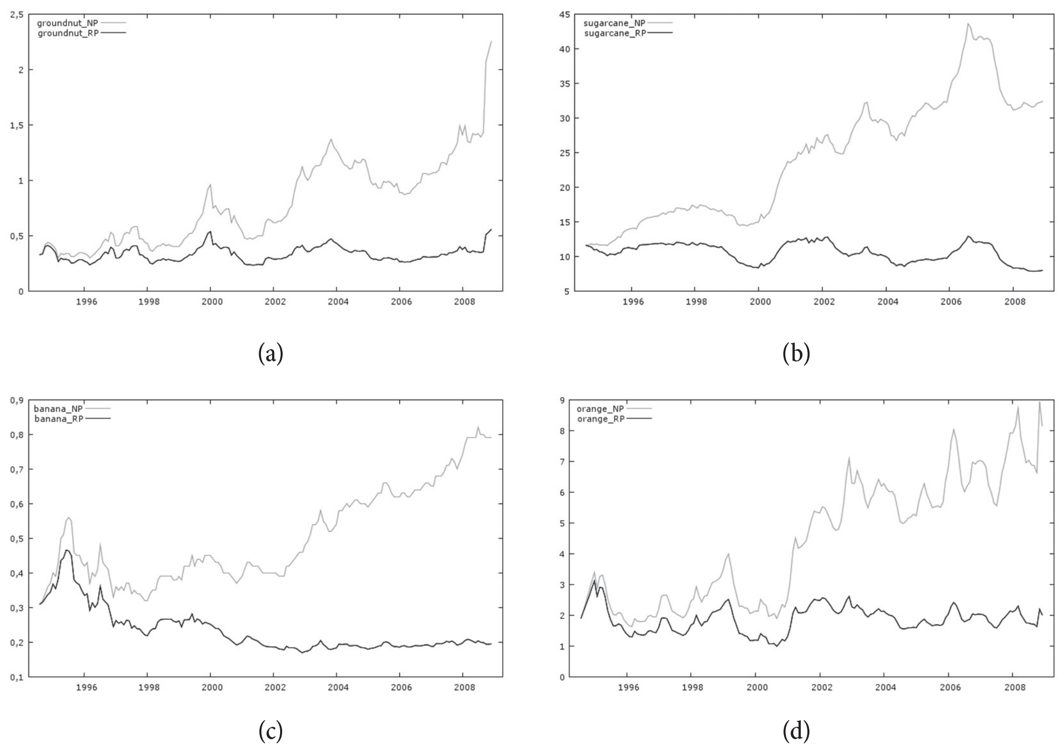
1 If the parameter is not significant, it lies in the model. However, the model will be less parsimonious.

2 Resídues should not be co-related so that it would be guaranteed that the model proposed for modeling the real prices is able to retrieve the autocorrelation from the series.

5. RESULTS AND DISCUSSION

Figure 1 shows some characteristics of the price series and suggests the most adequate modeling type for their study and forecasting. Although a preliminary diagnosis of nominal data would identify an increasing price trend, an analysis of real data demonstrates only a slightly growing trend component. It actually reveals a significant weight of inflation during the period. Figure 1 also shows that the seasonal component is only slightly or practically not significant throughout the years. ACF and PACF samples in Figure 2 corroborate the above statement. In fact, no peaks in the lag multiples of 12 exist; otherwise they would have indicated seasonality series.

Figure 1 – Nominal (black) and real (dots) prices of groundnut (a), sugarcane (b), banana (c) and orange (d) – from August 1994 to December 2008.

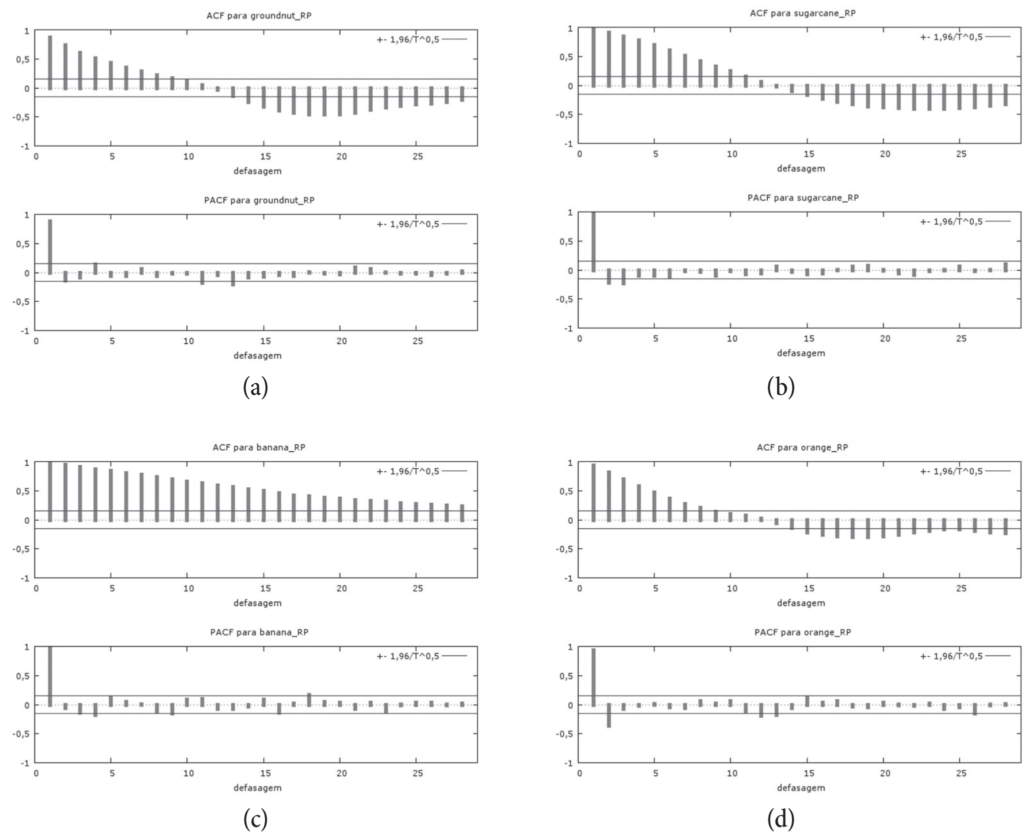


Source: IPEADATA, 2010.

The identification of stationarity of the series of real prices may be undertaken by analyzing samples ACF and PACF in Figure 2. ACF and PACF, typically associated with ARMA(p,q) models, may be observed for groundnut, sugar cane and orange, with stationarity evidences. However, price series of banana presents ACF with a slow decrease and indicates that the series is not stationary.

So that the presence or absence of stationarity may be statistically tested and a more robust criterion of the process which causes real price series may be obtained, KPSS hypotheses tests were undertaken. Nil hypothesis of stationarity for the price series of groundnut (test statistics = 0.114778), sugarcane (test statistics = 0.285236) and orange (test statistics = 0.0933421) was not discarded at 5% significance level (critical value = 0.464). However, the same hypothesis (test statistics = 0.695517) was discarded for the price series of banana. This confirmed the visual survey of samples ACF and PACF of Figure 2.

Figure 2 – ACF and PACF of real prices of groundnut (a), sugarcane (b), banana (c) and orange (d).



Source: Research data.

Table 2 shows the main descriptive measures from real price series of groundnut, sugarcane, banana and orange, with an overall view of the behavior of monthly mean prices during the period.

Real prices for groundnut, banana and orange had a moderate uniformity since variation coefficients (the ratio between mean price and standard deviation) were 19.4%, 27.2% and 22.2%, respectively. On the other hand, real prices of sugarcane presented variation, asymmetry and kurtosis coefficients significantly higher than the others'. This fact corroborated the non-stationarity condition of the series previously analyzed.

Table 2 – Descriptive measures of real prices.

Parameters	groundnut (R$/kg)	sugarcane (R$/ton)	banana (R$/kg)	orange (R$/hundred)
Mean	0.335	10.605	0.233	1.874
Median	0.326	10.709	0.203	1.876
Standard deviation	0.065	1.357	0.064	0.416
Maximum price	0.557	12.928	0.466	3.116
Minimum price	0.235	7.879	0.170	0.995
Asymmetry	0.863	-0.349	1.679	0.220
Kurtosis	0.755	-1.036	2.532	-0.229

Source: Research data.

Different orders p and q were tested by Box & Jenkins modeling when adjusting each real price series (groundnut, sugarcane and orange). ARMA(p,q) or ARIMA (p,0,q) models, which minimize AIC procedures, and statistics of Ljung-Box tests for the first 24 co-relationships were obtained by ACF and PACF survey of the residues. Data showed that residues were white noise (for risk 5%). The need of a single series differentiation to make it stationary was required in the case of banana real prices. Nil hypothesis of stationarity for the first difference of banana price series (statistics of KPSS test = 0.0894865) was not discarded at 5% significance level (critical value = 0.464). ARIMA(p,1,q) model was later obtained and proved to be more adequate for the adjustment of these prices.

Although different orders of the simple moving average model were also tested, the order with the lowest forecasting errors was order 2 for all the real price series under analysis. However, the simple mean and moving average models failed to describe adequately the crops' real prices and frequently forecasts overestimated real values. It should be noted that these methods were adequate to evaluate the trend of one series and are highly sensitive to recent changes of behavior.

In the case of exponential smoothing models, damping coefficients were obtained by a good research with Minitab.

Table 3 summarizes the results from different analyses of time series of fixed and open models and the respective statistic measures for forecasting error (MAD, MAPE and MSE). The lowest errors, which suggest models for better adjustments for each crop, are highlighted.

Table 3 – Results for time series analyses.

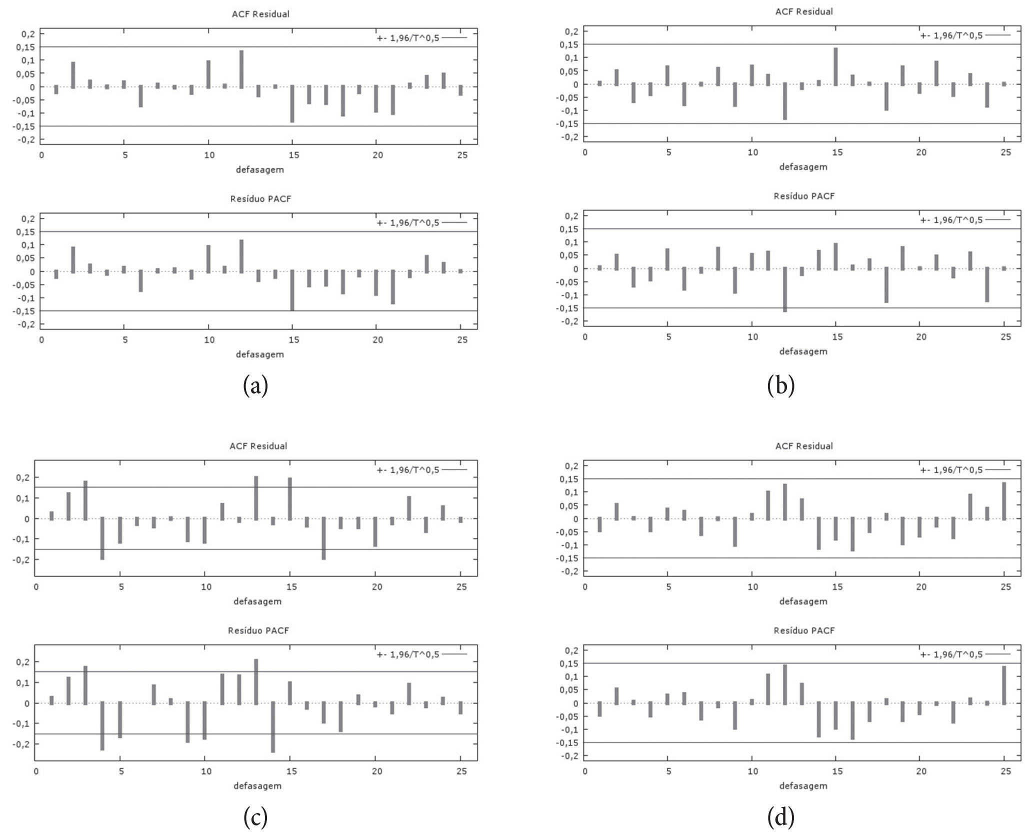
Models for time series	Forecasting errors		
Groundnut	MAD	MAPE	MSE
SA	0.05230	15.4010	0.00431
SMA	0.02183	6.30321	0.00101
SES	**0.01697**	**4.93952**	**0.00070**
DES – Holt	0.01749	5.10838	0.00074
TES – Winters	**0.01693**	**5.01096**	**0.00066**
ACD	0.04970	14.9555	0.00370
MCD	0.04960	14.9524	0.00370
ARIMA(2,0,0) $R_t = 0.0272 + 1.1137R_{t-1} - 0.1924R_{t-2} + a_t$ AIC = -754,2084 ; Ljung-Box test: p_value = 0.572	**0.01707**	**4.98360**	**0.00069**
Sugarcane	MAD	MAPE	MSE
SA	1.14153	11.6573	1.87416
SMA	0.26668	2.53124	0.12851
SES	0.20476	1.92214	0.07681
DES – Holt	**0.19953**	**1.87488**	**0.07352**
TES – Winters	**0.19896**	**1.86005**	**0.07285**
ACD	1.02028	9.94315	1.53252
MCD	1.02025	9.94291	1.53271
ARIMA(3,0,1) $R_t = 0.2432 + 1.7177R_{t-1} - 0.5907R_{t-2} - 0.1498R_{t-3} - 0.6071a_{t-1} + a_t$ AIC = 28.0914 ; Ljung-Box test: p_value = 0.466	**0.18096**	**1.6987**	**0.06436**

Banana	MAD	MAPE	MSE
SA	0.06039	27.9832	0.00416
SMA	0.00833	3.20580	0.00020
SES	**0.00722**	**2.75974**	**0.00015**
DES – Holt	**0.00739**	**2.81513**	**0.00015**
TES – Winters	0.00750	2.93613	0.00015
ACD	0.02990	12.7546	0.00160
MCD	0.03000	12.7640	0.00150
ARIMA(1,1,1) $R_t = 0.5958R_{t-1}^{(1)} + a_t + 0.5122a_{t-1}$ AIC = -1012.9630 ; Ljung-Box test: p_value = 0.113	**0.00711**	**2.71810**	**0.00015**
Orange	MAD	MAPE	MSE
SA	0.35394	20.7333	0.17971
SMA	0.14227	7.56495	0.03546
SES	**0.10268**	**5.42635**	**0.02062**
DES – Holt	0.10349	5.46943	0.02087
TES – Winters	**0.09503**	**5.11471**	**0.01882**
ACD	0.31510	18.1869	0.15420
MCD	0.31530	18.1813	0.15450
ARIMA(2,0,0): $R_t = 0.1694 + 1.2792R_{t-1} - 0.3698R_{t-2}$ AIC = -187.2616 ; Ljung-Box test: p_value = 0.337	**0.09979**	**5.34360**	**0.01891**

Source: Research data.

Figure 3 shows ACF and PACF of residues obtained by adjustment of real price series through Box & Jenkins modeling, with small fluctuations around zero stability, corroborating statistics by Ljung-Box test and described in Table 3.

Figure 3 – ACF and PACF of residues of ARIMA(p,d,q) models adjusted to real price series of groundnut (a), sugarcane (b), banana (c) and orange (d).



Source: Research data.

According to the proposed selection criteria (MAD, MAPE and MSE), it has been verified that the main promising models for better adjustments for each crop were:

Groundnut: TES with coefficients $\alpha = 0.95$, $\beta = 0.01$ and $\gamma = 0.03$; SES with coefficient $\alpha = 0.985$; ARIMA(2,0,0) with $R_t = 0.0272 + 1.1137R_{t-1} - 0.1924R_{t-2} + a_t$, respectively.

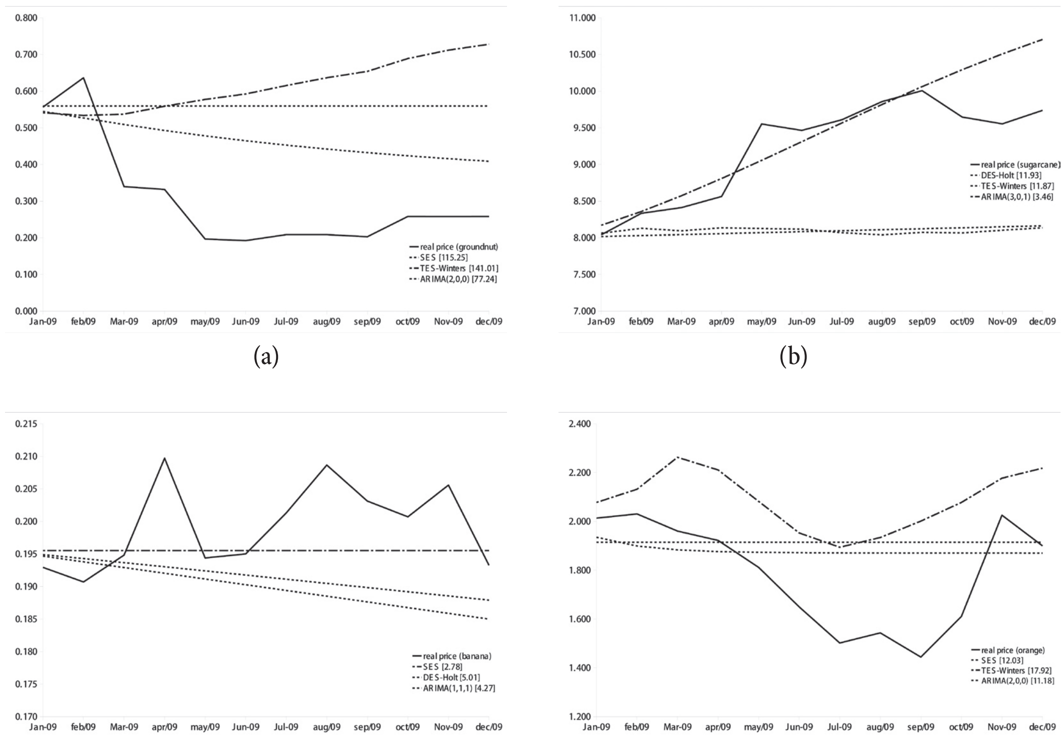
Sugarcane: ARIMA(3,0,1) with $R_t = 0.2432 + 1.7177R_{t-1} - 0.5907R_{t-2} - 0.1498R_{t-3} - 0.6071a_{t-1} + a_t$; TES with coefficients $\alpha = 0.95$, $\beta = 0.35$ and $\gamma = 0.01$; DES with coefficients $\alpha = 0.956$ and $\beta = 0.325$, respectively.

Banana: ARIMA(1,1,1) with $R_t = 0.5958R^{(1)}_{t-1} + a_t + 0.5122a_{t-1}$; SES with coefficient $\alpha = 0.95$; DES with coefficients $\alpha = 0.936$ and $\beta = 0.007$,respectively.

Orange: TES with coefficients $\alpha = 0.95$, $\beta = 0.37$ and $\gamma = 0.05$; ARIMA(2,0,0) with $R_t = 0.1694 + 1.2792R_{t-1} - 0.3698R_{t-2}$; SES with coefficient $\alpha = 0.995$, respectively.

Figure 4 shows real prices for the period between January and December 2009 compared to their respective forecasts. Models with the best adjustments among all those tested for each crop are employed.

Figure 4 – Real prices (black) versus forecast prices (dots) by MAPE (between brackets) of groundnut (a), sugarcane (b), banana (c) and orange (d) between January and December 2009.



(c) (d)

Source: Research data.

According to results in Table 3, the exponential smoothing models were practically very similar to those by ARIMA models in the verification stage of adjustment. In fact, they were more precise for groundnut and orange. However, in the forecasting stage, somewhat further on, Box & Jenkins models were more efficient. In fact, they produced less mean absolute percentage errors for all crops, with the exception of banana (albeit with a slight difference), as Figure 4 demonstrates.

According to Morettin (2008) and Oliveira & Andrade (2012), forecasts with many steps forward and using the same adjustment model are not recommended. It is more convenient to update information data (as these occur) in the model and re-calculate the forecasts. As data are updated, the adjustment model should be revised. Smoothing models would be indicated for the crops' analysis and forecasting some stages forward. However, information updating should be constantly undertaken.

Pindyck & Rubinfeld (2004) state that the smoothing process is highly useful when it is necessary to remove, or at least to reduce, short-term fluctuations of a time series. In other words, exponential smoothing methods are indicated when data interpretation and analysis should be made easier. However, when the researcher's interest lies on forecasting results, these methods have disadvantages since they are not based on probabilistic models and the frequent forecasting errors are co-related with non-valid forecasts for long periods of time.

On the other hand, Makridakis, Wheelwright & Hyndman (1998) investigated the forecasting potential of several methods commonly employed in forecasting of time series. The above researchers found that increase in complexity and in statistical sophistication of forecasting methods of time series does not necessarily imply a better significance in forecasting preciseness. They report that simple forecasting methods may forward very satisfactory performance under certain conditions and normally achieve total comprehension in their suppositions and limitations and in the interpretation of results. Prior to the adoption of a more complex forecasting method, the advantages that it may produce with regard to costs in its application should be evaluated.

4. CONCLUSION

A product must be economically competitive to be feasible. Within this context, one of the most relevant issues is price behavior which oscillates mainly during the harvest and between-harvest periods and directly affects the planning process and decision-taking in commercialization.

The application of another forecasting method and the verification of the forecast's consistency may and should be taken into account in price analysis. In fact, it gives more reliability and reduces the possibility of wide deviations.

It should also be enhanced that decisions are actually taken from forecasts. Small errors may bring great hazards and, consequently, every effort should be spent to obtain a better model.

5. ACKNOWLEDGMENTS

The authors would like to thank FAPESP for funding the scientific initiation scholarship (proc. 2008/53369-3 and proc. 2009/50746-3) and FUNDUNESP for funding (proc. 00206/09).

REFERENCES

BOX, G. E.; JENKINS, G. M.; REINSEL, G. **Time Series Analysis: Forecasting and Control**. New York: Prentice Hall, 1994.

CEPEA, Centro de Estudos Avançados em Economia Aplicada. Available at < http://www.cepea.esalq.usp.br/ >. Acess on 20th June 2010.

GONÇALVES, G. **Prognosticação quantitativa utilizando técnicas de alisamento baseadas em séries temporais**. Undergraduate Monographin Production Engineering, Departamento de Engenharia de Produção, Administração e Economia, Universidade Federal de Ouro Preto – UFOP, 2003.

IEA, Instituto de Economia Agrícola. Available at <http://www.iea.sp.gov.br/>. Acess on 1st February 2010.

IPEADATA, Instituto de Pesquisa Econômica Aplicada. Available at <http://www.ipeadata.gov.br/>. Acess on 12th March 2010.

KWIATKOWSKI, D.; PHILLIPS, P.; SCHMIDT, P.; SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, v. 54, p. 159-178, 1992.

LJUNG, G.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, p. 297-303, 1978.

MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R. **Forecasting:** Methods and Applications. New York: John Wiley & Sons, 1998.

MARTINS, R. Cultivares de amendoim: Um estudo sobre as contribuições da pesquisa pública paulista. **Informações Econômicas**, v. 36, p. 37-49, 2006.

MENDES, J. T. G.; PADILHA JR., J. B. **Agronegócio:** uma abordagem econômica. São Paulo: Pearson Education do Brasil, 2007.

MDIC, Ministério do Desenvolvimento, Indústria e Comércio Exterior. Available at <http://www.mdic.gov.br/>. Acess on 16th April 2010.

MORETTIN, P. A. **Econometria Financeira**. São Paulo: Edgard Blücher Ltda., 2008.

MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. São Paulo: Edgard Blücher Ltda., 2004.

OLIVEIRA, S. C.; ANDRADE, M. G. Comparison between complete Bayesian method and empirical Bayesian method for ARCH models using Brazilian financial time series. **Pesquisa Operacional** (Impresso), v. 32, p. 293-313, 2012.

PINDYCK, R.; RUBINFELD, D. **Econometria:** Modelos e Previsões. Rio de Janeiro: Elsevier, 2004.

SÁFADI, T.; ANDRADE, M. G. **Abordagem Bayesiana de Modelos de Séries Temporais**. 12ª Escola de Séries Temporais e Econometria, ABE/SBE, Gramado-RS, 2007.

SANTINI, G. A.; OLIVEIRA, S. C.; PIGATTO, G. Análise da relação das variáveis preço e produção da mandioca tipo indústria no estado de São Paulo, 1996 a 2008. **Informações Econômicas**, v. 40, p. 41-52, 2010.

SCHWARZ, G.. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, p. 461-464, 1978.

SOUZA, R. C.; CAMARGO, M. E. **Análise e Previsão de Séries Temporais:** Os Modelos ARIMA. Ejuí: SEDIGRAF, Santa Maria-RS, 1996.

SPIEGEL, M. R. **Estatística**. São Paulo: Pearson Education do Brasil, 1994.