

MONTHLY CLOSING PRICE FORECASTING OF SOYBEAN GRAIN IN PARANÁ USING SARIMA MODELING WITH INTERVENTION

Rodrigo Clemente Thom de Souza^{1,*}; Osvaldo Guedes Filho¹; Marco Aurélio Reis dos Santos¹ and Leandro dos Santos Coelho²

¹Federal University of Paraná, Jandaia do Sul, Paraná (Brazil). ²Catholic Pontifical University of Paraná, Curitiba, Paraná (Brazil).

* Corresponding author: thom@ufpr.br

ABSTRACT: This article presents a procedure to find a suitable model for forecasting financial time series of agribusiness commodities considering the negotiation and circumstantial factors that influence them, especially seasonality and abrupt economic changes. The used data were the series of monthly closing prices of soybean in the state of Paraná between the months of February 1998 and January 2004. This series and its transformations were tested empirically in terms of ACF analysis (*Auto-Correlation Function*) until stationarized changes were found. For this, the application of seasonal adjustment was necessary due to the characteristic of the series and interventions (dummy variables) due to structural breaks occurred during the analyzed period. The seasonal adjustment was done through the X-12-ARIMA method, which implements the univariate ARIMA model (*AutoRegressive Integrated Moving Average*) and the multiplicative seasonal (SARIMA). The combined use of SARIMA model and of interventions got, in terms of the statistical analysis adopted, modeling the negotiation factors known to influence their oscillations and realize the forecast with 1, 2 and 3 steps ahead.

Key words: Time series forecasting, SARIMA modeling, X-12-ARIMA seasonal adjustment, interventions.

PREVISÃO DO PREÇO DE FECHAMENTO MENSAL DA SOJA EM GRÃO NO PARANÁ UTILIZANDO MODELAGEM SARIMA COM INTERVENÇÕES

RESUMO: Este artigo apresenta um procedimento para se encontrar um modelo adequado para previsão de séries temporais financeiras de commodities do agronegócio considerando-se os fatores negociais e conjunturais que as influenciam, em especial sazonalidade e mudanças econômicas abruptas. Os dados utilizados foram da série de preços de fechamento mensal da soja em grão no Estado do Paraná entre os meses de fevereiro de 1998 e janeiro de 2004. Esta série e suas transformações foram testadas empiricamente em termos da análise ACF (*Auto-Correlation Function*) até que fossem encontradas transformações estacionarizadas. Para tal, foi necessária a aplicação de ajustamento sazonal devido à característica da série e de intervenções (variáveis dummy) devido às quebras estruturais ocorridas no período analisado. O ajustamento sazonal foi feito através do método X-12-ARIMA que implementa o modelo univariado ARIMA (*AutoRegressive Integrated Moving Average*) sazonal (SARIMA) multiplicativo. O uso combinado do modelo SARIMA e das intervenções conseguiu, em termos da análise estatística adotada, modelar os fatores negociais conhecidos por influenciar suas oscilações e realizar a previsão com 1, 2 e 3 passos à frente.

Palavras-chave: Previsão de séries temporais, Modelagem SARIMA, Ajuste sazonal X-12-ARIMA, Intervenções.

Received: Aug. 16, 2016 - Accepted: Dec. 06, 2016

1. INTRODUCTION

According to IBGE (2015), Brazil is currently found among the world's largest grain producers, being the second largest soybean producer, just behind the United States of

America (USA). This justifies in part, the great amount of recent scientific research involving this subject in Brazil (Guedes Filho et al., 2010; Vieira et al., 2010; Souza et al., 2013a; Souza et al., 2013b; Silva et al., 2014). In the agricultural sector, especially in soybean production, besides concentrate efforts on culture management activities (sowing,

irrigation, pest and diseases control, harvest), the producer need to be alert to the market to get a good sale price of the product. In this context, the time series forecasting is an important way for establishing future prices based on past information.

Aiming to improve the performance of the forecast, a model must consider several influencing factors to the price of the product and indicate the best time to complete the sale. In the case of soybean, the main factors are the international price, the exchange, the seasonality, among other factors, diseases as the Asian soybean rust, most recently pests as the caterpillar, *Helicoverpa*, unfavorable climatic and meteorological conditions, and quotation of substitute products.

As the international price (supply/demand), soybeans are traded on the Chicago Stock Exchange/USA, where the oscillations are influenced by "isolated" situations as the increase of over 1000% of demand from China between 1999 and 2006, and the super crop of soybeans in the USA of 61.5 billion dollars in 2004 (USDA, 2004).

In relation to seasonality, it is known that in Brazil, the soybean sowing usually occurs between October/November, but eventually, depending on weather conditions, some producers can anticipate the sowing. On the other hand, if there is a delay of rains, the producers are obliged to delay the sowing until December. The harvest is between 100 and 130 days after sowing, depending on the cultivation: early, semi-early and late.

Finally, in relation to the exchange, for being an exporting product, the soybean price is influenced by the dollar quotation. Thereby national and international financial crisis as the devaluation of the Real (1999), the Argentine crisis (2001) and the election crisis

(2002) caused impacts on soybean quotation. Below, a brief description of each of these three crises:

a) Devaluation of the Real: With the crises in Mexico (1995), Asia (1997) and Russia (1998), raised the degree of mistrust of foreign investors with regard to so-called emerging countries, becoming harder, expensive and insecure the maintenance of the supported Brazil's economic policy on the anchors of the overvalued exchange rate and of the high interest rates (DIEESE, 2015). Facing the growing risk of a devaluation of the real and its impact on the public debt, in 1999 was started an outflow movement of national and international speculative capital. To contain this capital outflow, on January 13, 1999, the Brazilian government announced the expansion of the Real Floating banking. The exchange, which until then had been keeping artificially valued, enters in a rapid devaluation process. On 03 March of this same year, when it reached its highest point, the dollar was quoted at R\$ 2.16 (on January 12, 1999, was quoted at R\$ 1.21). With some measures like raising interest rates, a loan with the International Monetary Fund (IMF) and the supply of government securities indexed to the exchange rate, the government has succeeded in stabilizing the exchange rate at R\$ 1.65 - R\$ 1.70 from the second half of April 1999.

b) Argentina crises: Owing one of the most fertile soils in the world, Argentina is distinguished by high grain production. Being soybeans, together with wheat and maize, one of its main export commodities. A financial crisis in this country severely affected by international soybean prices. On

December 19, 2001, the Minister of Economy of Argentina, Domingo Cavallo, leaves his position after Congress take the special powers that were granted months before. Some of the results of this crisis, which in less than 12 days made the Argentines knew five presidents, were: the most official moratorium on foreign debt in history (\$ 132 billion) and the release of "Argentine", a new bonus that would act as third currency in the country market.

c) Electoral crisis: In the second semester of 2002, the election eminence of the candidate Luis Inácio Lula da Silva, considered by the market as unpredictable in the conduct of economic policies if elected, caused a great escape of speculative capital in Brazil. This foreign currency departure of the country has made that the Real were depreciated over the months. The situation only began to reverse after the elections, when the economic policy of Lula, winner of these elections, was considered appropriate by the market.

On the other hand, researchers of econometrics discipline have been analyzing time series of agribusiness commodities prices as being the composition of several factors, which relates to the trend, the cycle, the seasonal effects and to an irregular component (Fávero et al., 2003).

The seasonality, which is one of the objects of this study, is interpreted as a regular movement of a series in a year, representing the systematic movements caused by non-economic phenomena (Thomas and Wallis, 1971) as, for example, climate changes. The irregular component relates to the unforeseen movements, randomly generated within of a series, such as strikes, non-

seasonal climatic conditions, etc. (Campos, 1991).

According to Durbin (1983), the trend-cycle component refers to the residual part of a series, when from the same were extracted the irregular and seasonal components. In addition, as these components contribute in a very strong manner for the total variance of the series, they need to be "removed" in order to enable the policy analysis of the effects caused by the economic variables (Campos, 1991).

The methods with this purpose approached in the present work are the SARIMA modeling with the application of the X-12-ARIMA method for "deseasonalisation" and the application of intervention functions (dummy variables) for the "smoothing" of structural breaks of the time series.

Kim et al. (2003) and Zhang et al. (2007) assessed the SARIMA models for forecasting of weather series of precipitation (rainfall) of South Korea and Australia, respectively. Pai et al. (2005) applied the SARIMA models to series of industrial production in Taiwan. As an example of X-12-ARIMA method application, can be specifically cited the work of Ribeiro and Dias (2004), which compares it with the Kalman filter for the forecast of Economic Activity Index of Maringá city. In addition, it is also cited the application of this method by Oliveira (1998), when he analyzed the seasonal pattern of industrial production, Brazilian exports, and imports. Most recently, Ferraz and Sáfyadi (2007) analyzed the related series to the minimum essential ration cost in the metropolitan areas of Belo Horizonte, São Paulo, Porto Alegre and Rio de Janeiro.

About using the dummy variables in time series, Hassler (1999) explains in detail this procedure in a practical case of Germany's

demand forecast. Alana (2003) also uses interventions to smoothing abrupt changes in the economy by modeling the rate of inflation in Brazil. Few studies until the moment, simultaneously involve the consideration of seasonality and abrupt economic disruptions. Noland et al. (2007) jointly applied the analysis of interventions and the use of the SARIMA models in an urban traffic problem of London. Alonso and Arcila (2013) analyzed the seasonal behavior for improvement of forecasting of the sugar price in this context. Finally, Asche and Oglend (2016), study the relation between the factors in question for future pricing of Norwegian salmon.

A time series is any set of observations ordered in time (Morettin and Toloj, 1987). The time series forecasting consists of establishing the future values for the series and is based either on the current information or in the past. The forecast horizon is the length of time, counted from a predetermined origin (forecast origin).

The time series forecasting is a relevant area of study, considering the numerous examples of time series in the literature, such as Souza (2008), Souza et al. (2009), Cui et al. (2015a, 2015b), Liu et al. (2015) and Prema et al. (2015). In order to be established future values for the series under study, it is necessary that, in some way, it can capture and formulate a mathematical model able to represent the behavior and the characteristics of time series that is wanted to be predict.

The present study differs from other previous studies of literature that made studies a greater or lower similarity over future pricing of commodities, such as Hassler (1999), Kim et al. (2003), Alana (2003), Ribeiro & Dias (2004), Pai et al. (2005), Zhang et al. (2007), Ferraz & Sáfyadi (2007), Noland et al.

(2007), Alonso & Arcila (2013) and Asche & Oglend (2016), because simultaneously focuses on the analysis of seasonal effects and abrupt economic changes in a specific clipping to Paraná, one of the most relevant federal states of the country in terms of soybean production and location where this research was sheltered. It is supposed that, possibly, if the same analysis were applied to other states or even soybean producer countries, the results would be similar for the same period, once the studied active is about commodity and the main factors considered related to macroeconomic issues.

The aim of this work was the modeling and forecasting of time series of monthly closing prices of soybeans in the state of Paraná between the months of February 1998 and January 2004 considering some of the main negotiating factors that influence their fluctuations, especially seasonality and national and international financial crises. It is justified the choice of the period, justly, to forecast the reaction of the future prices of the above-mentioned international financial crisis commodity

2. MATERIAL AND METHODS

One of the most adequate forecasting models to perform such task are the univariate models. In these models, the future values of a series are explained only by past values of this same series. The ARIMA models (AutoRegressive Integrated Moving Average) are framed within that classification.

An ARIMA model is intended to represent and predict the behavior of a time series through autoregressive processes, integrated and with moving averages (Morettin, 2006).

The autoregressive process AR (p) is defined by:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t \quad (1)$$

wherein p is the number of observations or order of the AR model, ϕ It is the weighting parameter, δ is intercept related to the average of y_t and u_t is the random disturbance.

The process of moving averages MA (q) is generated by the weighted average of random disturbances and is expressed by:

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (2)$$

wherein q is the number of past periods or order of the MA model, $\mu = E(y_t)$, u_t are the uncorrelated random errors with zero average and constant variance (white noise) and θ can assume positive and negative values.

The d number differences required to make the stationary series is called integration order. The combination of all these terms allows the ARIMA models to be used (p, d, q) given by:

$$\Delta^d y_t = \delta + \phi_1 \Delta^d y_{t-1} + \phi_2 \Delta^d y_{t-2} + \dots + \phi_p \Delta^d y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (3)$$

where $\Delta^d = y_t - y_{t-1}$

2.1 SEASONAL SERIES

The ARIMA models explore the self-correlation between the values of the series at successive moments, but when the data are observed in periods lower than one year, the series may also present autocorrelation for a seasonal station s . The models, which

contemplate the series that present seasonal autocorrelation, are known as SARIMA (Seasonal AutoRegressive Integrated Moving Average).

The SARIMA contain a non-seasonal part, with parameters (p, d, q) and other seasonal with parameters (P, D, Q) _{s} . The model multiplicative seasonal ARIMA (SARIMA) of order (p, d, q) _{s} - (P, D, Q) _{s} is defined by:

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\Theta(B^s)\varepsilon_t \quad (4)$$

wherein: $\phi(B)$ is the autoregressive stationary operator (non-seasonal) of order p ; $\Phi(B^s)$ is the autoregressive stationary operator of order P and seasonal station s ; $\theta(B)$ is the operator of invertible moving average (non-seasonal) of order q ; $\Theta(B^s)$ is the seasonal invertible of moving average operator of order Q and seasonal station s ; Δ^d is the part of non-seasonal integration order d ; Δ_s^D is the part of seasonal integration order D and seasonal station s .

In practice, is applied a signal extraction procedure for the seasonal adjustment in the series modeled through the SARIMA model. Thus, the seasonal component is estimated and removed from the original observations. One of the most applied approaches to seasonal adjustment is implemented by X-12-ARIMA program developed by U.S. Bureau of the Census (Findley et al., 1996).

Before the rise of X-12-ARIMA, the U.S. Bureau of the Census developed the X-11 method (Shiskin et al., 1967). X-11 offered a greater number of options of choice, such as the previously adjustment of the original data due to the calendar variations, the treatment of extreme values and, mainly, the choice of the filters for estimating the series of components (Fávero et al., 2003).

With the X-11 evolution, the Statistics Canadá created the X-11-ARIMA, which consisted in adding at the extremities of the original series a certain number of expected data and, then, apply the 11 X method to the extended series. Due to its adherence to the forecasting using ARIMA models, this new seasonal adjustment method received the designation of X-11-ARIMA method (Fávero et al., 2003).

The X-12-ARIMA method is the direct evolution of the X-11-ARIMA. The main characteristics of this method are regarding the good degree of accuracy, simplicity, a small number of parameters, stability, and low cost. In short, the X-12-ARIMA is the application of a two-sided sequence of filters in the original observations of the time series, where the final and initial values of the sample are supplemented, respectively, from the future forecast and initial observations for the adequacy of the appropriate seasonal ARIMA model (SARIMA) (Marini & Moauro, 2006). Figure 1 shows the sequence of filters X-12-ARIMA method, through which the prediction models suffer from the seasonal adjustment.

2.2 INTERVENTION IN STRUCTURAL BREAKS

According to Hotta (1988) *apud* Fávero et al. (2003), The adoption of a seasonal adjustment technique on a large scale must be designed in careful manner in development countries, such as Brazil, due to these be often subjected to major structural and conjunctural changes, which causes, according Dagum (1978), great irregularities, affecting the use and the proper adjustment results.

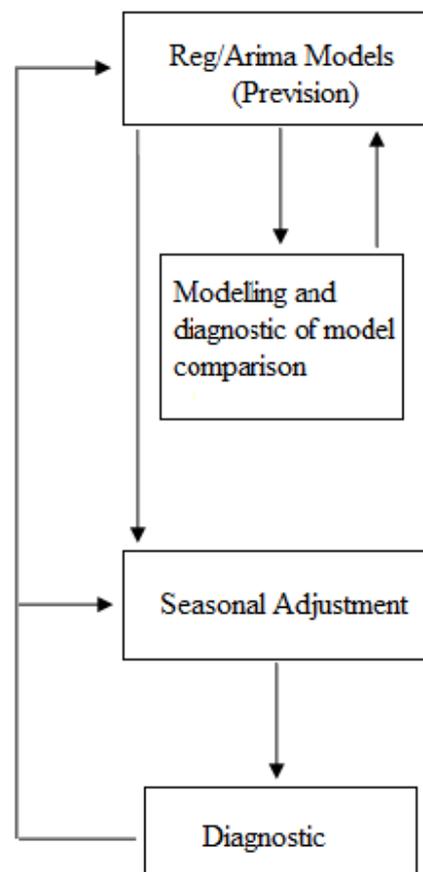


Figure 1. Structure of the X-12-ARIMA method. Source: Adapted from Findley et al. (1996) - Free Translation.

One way to consider these structural changes (or breaks), is through the application of interventions in the model. Be the algebraic expression below the representation of our model:

$$\hat{Y} = \alpha + \beta_1 T + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 \quad (5)$$

wherein β_1 represents the trend component coefficient and β_2 , β_3 and β_4 represent the coefficients of the intervention (devaluation of the real, the Argentine crisis and electoral crisis, respectively).

The variables Q_2 , Q_3 , and Q_4 are related to dummy variables. A dummy variable is an artificial variable, being 1 for the cases in

which occurs a given situation and 0 for the other cases. For this example, if a given observation is related to the devaluation of the real, then $Q_2 = 1$, $Q_3 = 0$ and $Q_4 = 0$.

2.3 DESCRIPTION OF THE USED DATA

The soybean price indicator of the Center for Advanced Studies in Applied Economics / College of Agriculture "Luiz de Queiroz" (CEPEA / ESALQ), previously called ESALQ / BM & F, is a weighted arithmetic average of the observed prices in Paraná for soybeans in bulk, exportation type, according to standard Concex: up to 14% of moisture, up to 2% of impurities, and maximum 8% for faulty grains (up to 5% of burnt) and 30% of broken grains (CEPEA, 2015).

The soybean indicator is released daily by CEPEA since August 1997 and is used by agents and institutions that deal with the physical market as one of the main price references.

Despite the indicator only consider Paraná's prices, an important state in the production, industrialization, and exportation of soybeans, the CEPEA staff also consulted agents in the states of Rio Grande do Sul, Mato Grosso, Mato Grosso do Sul, Goiás, Minas Gerais and Sao Paulo (CEPEA, 2015).

For obtaining the indicator, the state of Paraná was divided into five regions: Port of Paranaguá, Ponta Grossa, Southwest, West, and North. In each region is consulted daily a representative sample of buyers, sellers, brokers and dealers that provide prices (in Reais), free of taxes, which will comprise the indicator for this day. The cash payment discount calculations are made from the Rural Promissory Note (RPN) collected daily with several banks (CEPEA, 2015).

The indicator will be determined by the general arithmetic average of the simple arithmetic regional average weighted by the relative participation of the regions in the overall static crushing capacity of Paraná state, in the previous year. For example, for 1997, the regional weightings for soybean price indicator were: Port of Paranaguá (19.13%), Ponta Grossa (25.49%), North (34.72%), West (9.73%) and southwest (10.93%). The percentage of participation of each region is determined by its installed capacity of crushing (CEPEA, 2015).

Such participations will be updated, since the moving average of the last two years for the static capacity of processing of Paraná state, with data provided by the Brazilian Association of Vegetable Oil Industries (BAVEOI), present statistically significant changes. In the case of having no prices in a given region, this location will not be considered and its participation will be distributed proportionally among the others for indicator calculation (CEPEA, 2015).

Although the original series be daily, for this work the data were tabulated on a monthly basis, considering only the closing price of the last day of each month in the period of February 1998 to January 2004.

The implementation of ARIMA models is based on an iterative process, in which the choice of the model is made based on its own data (Morretin and Toloj, 1987). According to Box and Jenkins (1976), in the establishment of an ARIMA model for a time series there are three stages to consider:

1. Identification: consists on discovering through the analysis of the autocorrelation (ACF) and partial autocorrelation (PACF) function, which of the several versions of the ARIMA models (seasonal or not) best

describes the series behavior or its transformations and/or differences;

2. Estimation: having a provisional model for the time series, the next step is to estimate the autoregressive component parameters (AR) function, the moving average (MA) and the variance of residues (ϵ_t);

3. Verification or Diagnosis consists in verifying whether it represent, or not, adequately, the data (MORETTIN, 2006).

According to Werner & Ribeiro (2003), a commonly used procedure is to identify not only a single model but also some models that will then, be estimated and verified. When you get a satisfactory model, function, it passes to the last step of the methodology,

which constitutes the main goal is running forecasts.

The simulation tools used in this work were PCGIVE 10.0 and X12arima 2.09 modules do software GiveWin 2.20 (OxMetrics).

3. RESULTS AND DISCUSSION

Figure 2 presents the monthly series price of soybeans in Brazilian reais without changes. The whole series presents increasing trend with reasonable oscillations in its first half (between 1998 and 2001). In the second half (between 2001 and 2004) presents three expressive peaks (approximately 30 reais at the end of January 2001, and 50 reais at the end of 2002 and at the end of 2003) and two valleys (approximately 20 reais earlier in 2002function, and 35 reais in mid-2003).

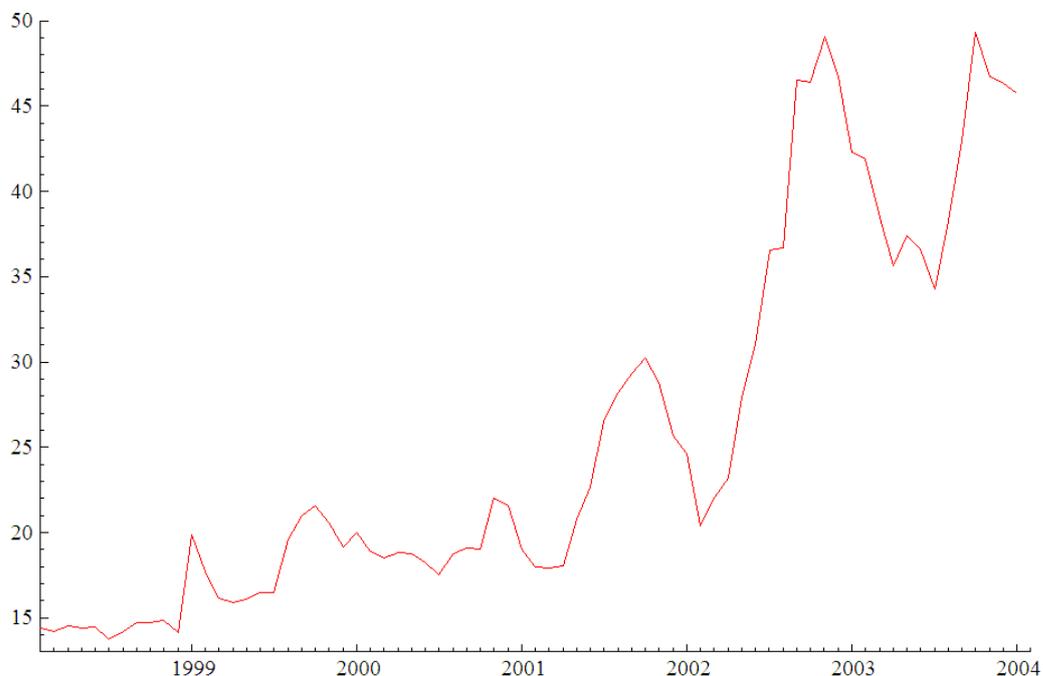


Figure 2. Monthly series of soybean prices.

The next step is the attainment of the ACF analysis in the series. The aim of this analysis is to identify if the series is stationary. Can be observed that the confidence interval (horizontal black line in Figure 3) is always overtaken in ACF test, indicating the presence of autocorrelation in series, and consequently, non-stationarity.

The next step is the accomplishment of series transformations aiming to stationary it. The accomplished transformations were the

first and the second differences (not recommended more than two differences, because in this case, the data become spurious), logarithm and first logarithm difference. In Figure 4 ACF analyses are presented, this time for the transformed series.

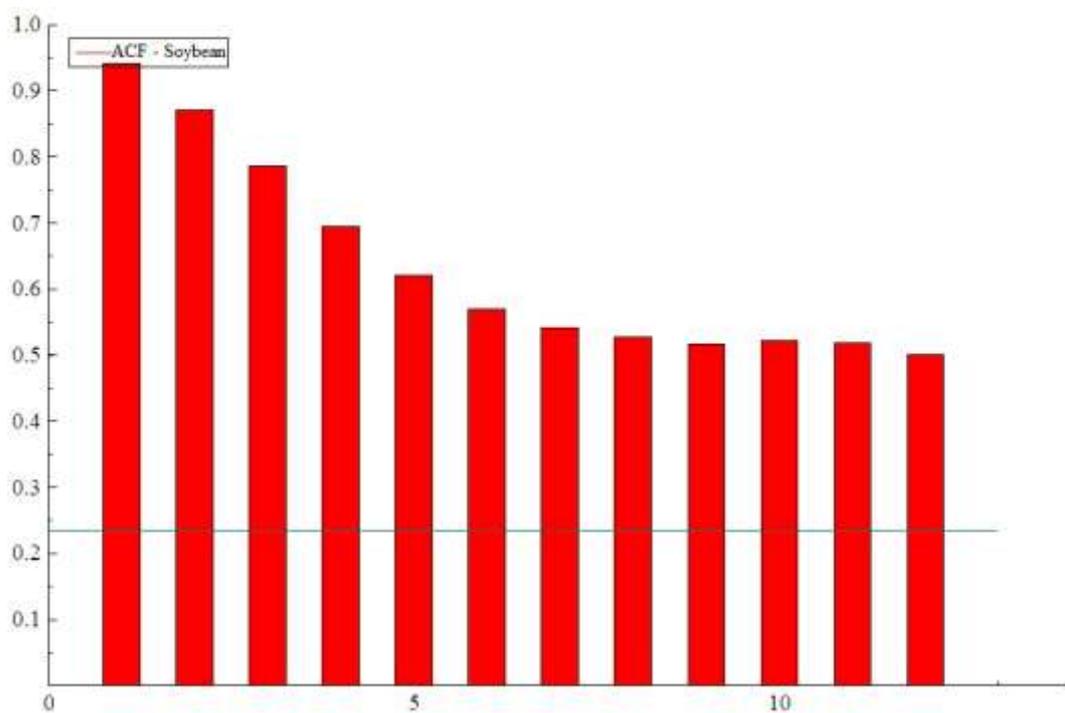


Figure 3. ACF test of the soybean prices monthly series.

As the autocorrelation series presented, it was not necessary to analyze the PACF, which would indicate for how many lags the autocorrelation is significant, in other words, the degree of the autoregressive component (AR).

The next step is the accomplishment of series transformations aiming to stationary it. The accomplished transformations were the first and the second differences (not recommended more than two differences,

because in this case, the data become spurious), logarithm and first logarithm difference. In Figure 4 ACF analyses are presented, this time for the transformed series.

As once again all series continued presenting autocorrelation, as shown in Figure 5, interventions were included (dummy variables) in their structural breaks aiming to stationary it (Figure 6). The considered structural breaks were the devaluation of the real (with $\beta = -3$ in the period from

January/99 to February/99), the Argentine crisis (with $\beta = +3$ in the period from December/2001 to April/2002) and the

electoral crisis (with $\beta = -6$ in the period from September/2002 to February/2003).

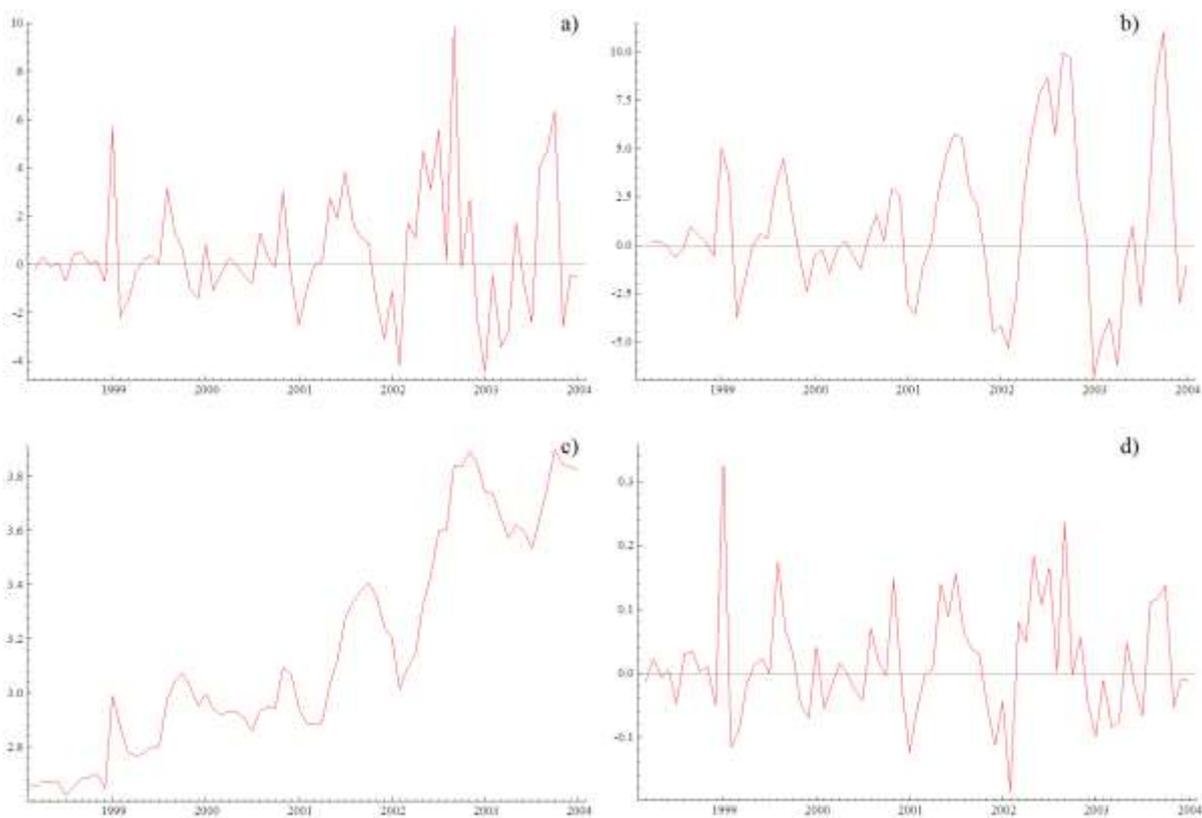


Figure 4. Monthly series of transformed soybean prices: a) first difference; b) second difference; c) logarithmic; d) logarithmic (first difference).

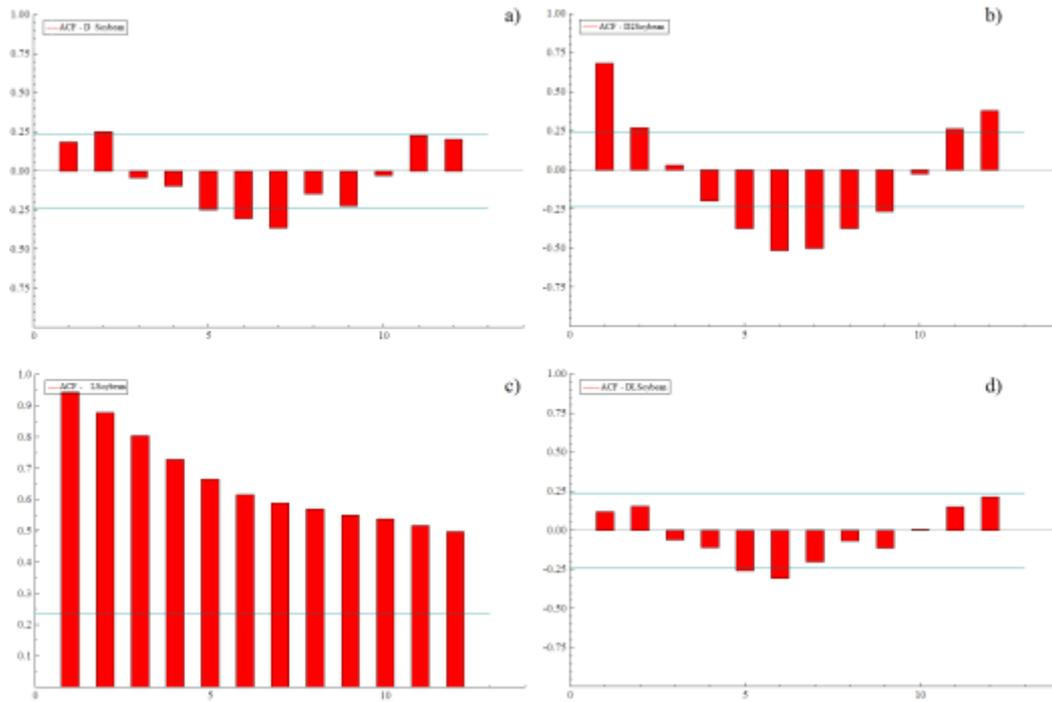


Figure 5. ACF test of monthly series of transformed soybean prices: a) first difference; b) second difference; c) logarithmic; d) logarithmic (first difference).

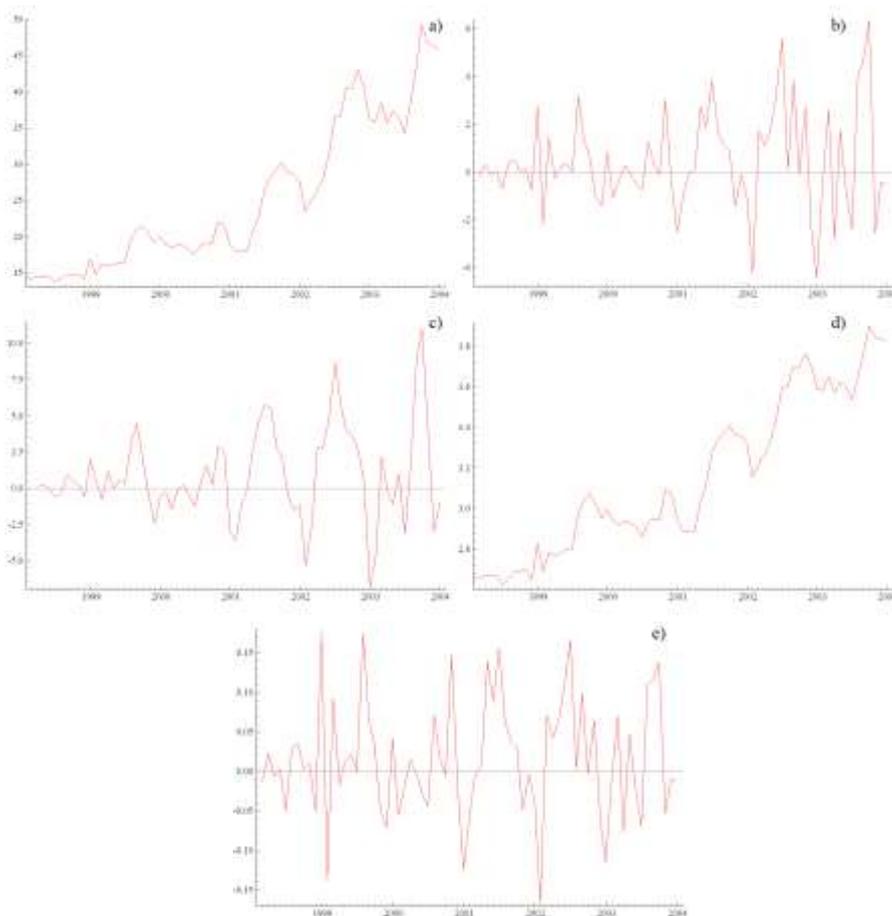


Figure 6. Monthly series of intervention soybean prices: a) no transformation; b) first difference; c) second difference; d) logarithmic; e) logarithmic (first difference).

Figure 7 shows that the series still present autocorrelation (the first logarithmic difference with interventions almost has not provided), the next step is "deseasonalize" the series through SARIMA modeling (Doornik, 2001). For convenience, the type of seasonal adjustment adopted in this article was the multiplicative.

Finally, the ACF test presented in Figure 8 indicated the absence of the autocorrelation in the series. The next step becomes to evaluate the quality of SARIMA model in

order to accept or reject a particular type of seasonal adjustment.

According to Hotta & Cazorla (1990), the most pervasive criterion for deciding the type of seasonal adjustment is the use of M statistics and quality control statistical Q presented in Table 1.

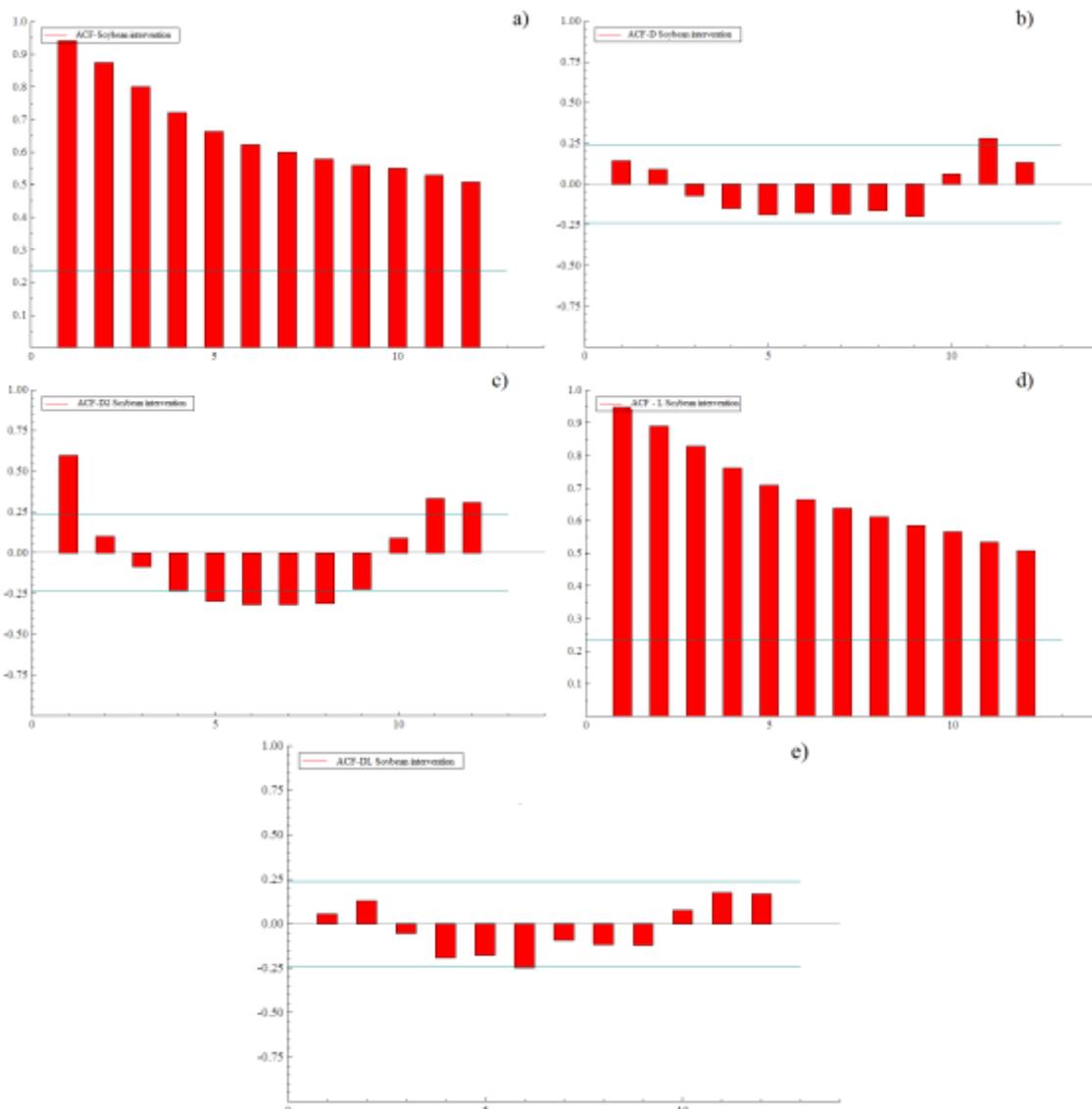


Figure 7. ACF test of monthly series of interventions soybean prices: a) no transformation; b) first difference; c) second difference; d) logarithmic; e) logarithmic (first difference).

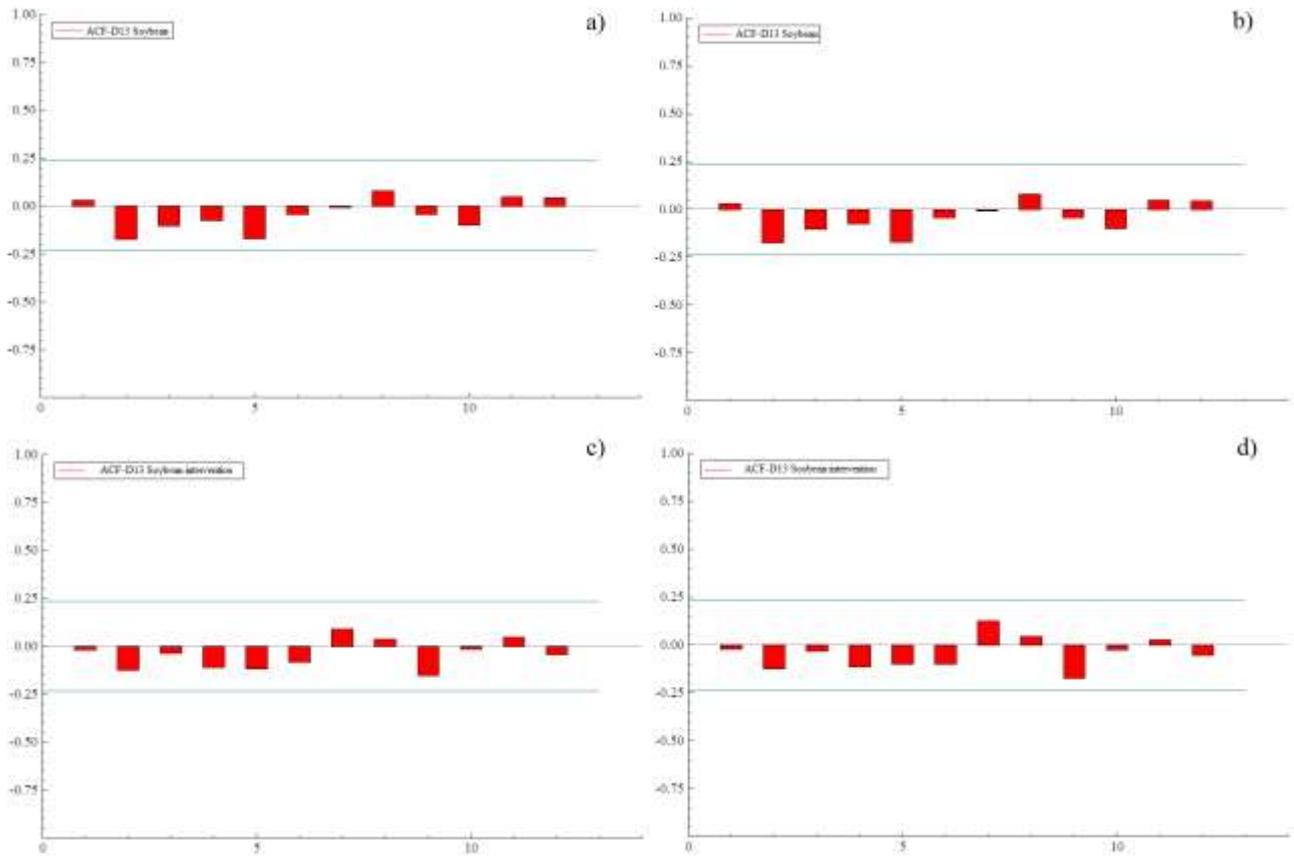


Figure 8. ACF test of monthly series of "deseasonalized" soybean prices: a) without intervention and without transformation; b) without intervention and with logarithmic transformation; c) with interventions and without transformations; d) with interventions and with logarithmic transformation).

Table 1. Quality assessment Statistics and monitoring of SARIMA models.

	SARIMA without intervention and without logarithmic transformation	SARIMA without interventions and with logarithmic transformation	SARIMA with intervention and without logarithmic transformation	SARIMA with intervention and with logarithmic transformation
M1	0.316	0.316	0.299	0.238
M2	0.309	0.309	0.324	0.239
M3	0.000	0.000	0.000	0.000
M4	0.513	0.513	0.146	0.146
M5	0.223	0.223	0.228	0.227
M6	0.273	0.273	0.182	0.137
M7	0.725	0.725	0.613	0.619
M8	0.753	0.753	0.653	0.656
M9	0.752	0.752	0.650	0.649
M10	0.815	0.815	0.701	0.701
M11	0.815	0.815	0.701	0.693
Acceptance level	0.48	0.48	0.38	0.36
Q (without M2)	0.50	0.50	0.38	0.37

Hotta & Cazorla (1990) propose the verifying of the dimensions of each of the M statistics, once that if any of them be greater

than 1 (the values may range between 0 and 3, but the acceptance region is generally between 0 and 1), it is necessary the use of

filters for seasonal adjustment. Although all SARIMA models were accepted by M and Q statistics (between 0 and 1), some values were relatively close to 1. So, it was decided to use the hybrid option of filters to seasonal adjustment (average 3 x 3 for the first alternatives, and an average of 3 x 5 for the final estimate). It is recommended that the choice of the best model involve the prior application of tests over the residue of each estimated model, such as autocorrelation and white noise. However, this study delimited the choice, only the identification of the best SARIMA model by the lower average of forecast error, as shown in Table 2. The choice by the evaluation criteria is based on the average forecast error results from the empirical decision. Moreover, this indicator is a very intuitive and of easy understanding criteria.

In Table 2, of the five not rejected models, four are with intervention and without logarithmic transformation, showing that the presence of interventions with dummy

variables provided greater consistency than when absent.

Finally, after choosing the best model, the series with interventions and with logarithmic transformation SARIMA (2 1 0) x (0 1 1)_s (line highlighted in blue in Table 2), the model can be used for forecasting. In this work were performed forecasts with 1, 2 and 3 steps ahead through the best model. Figure 9 shows the unadjusted, adjusted and the forecast series for the months of February, March and April 2004.

As presented in Table 3, the percentage difference ranged between 9.77% (one step ahead) and 22.80% (three steps ahead) and MSE (Mean Square Error) ranged from 16.89 (one step ahead) and 98.80 (three steps ahead). As the confidence interval with a probability of coverage over the original scale adopted was 0.95000, all forecasts were satisfactory.

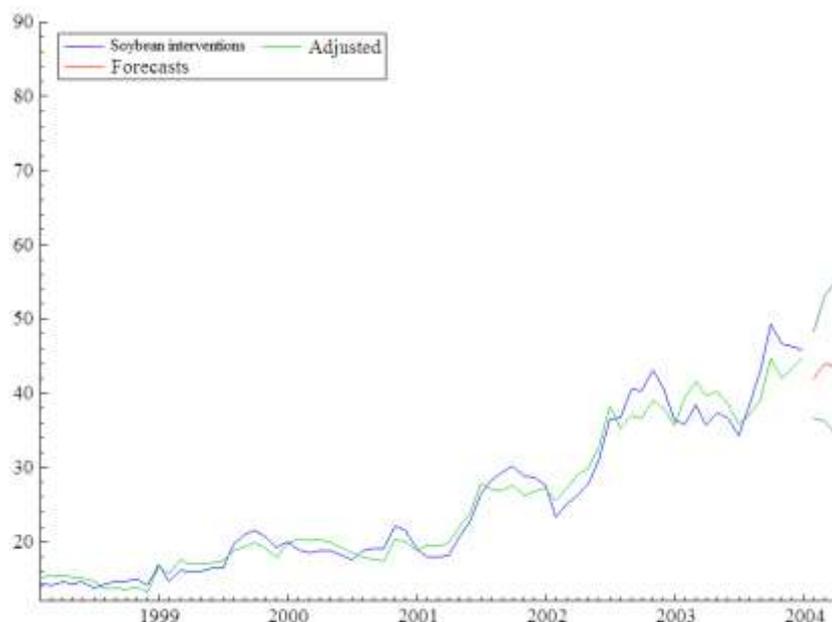


Figure 9. Forecast 1, 2 and 3 steps ahead.

Table 2. Comparison between ARIMA models. Best model: series with interventions and with logarithmic transformation SARIMA (2 1 0)×(0 1 1)_s.

Series	SARIMA	Absolut mean error in prevision into the sample (%)				Chi square probability (%)	Estimated non-seasonal parameters				Estimated seasonal parameters MA	Result	Mean error of prevision
		Last year	Penultimate year	Antepenultimate year	Last 3-years		AR		MA				
Without interventions and without logarithmic transformation	(0 1 1)(0 1 1)	17.33	24.84	23.68	21.95	1.84			0.014		0.606	Rejected	> 15%
	(0 1 2)(0 1 1)	16.84	25.55	27.85	23.41	18.99			0.069	-0.288	0.534	Rejected	> 15%
	(2 1 0)(0 1 1)	15.45	26.84	30.39	24.22	11.43	-0.022	0.257			0.558	Rejected	> 15%
	(0 2 2)(0 1 1)	9.26	52.95	78.98	47.06	1.63			1.008	-0.008	0.601	Rejected	> 15%
	(2 1 2)(0 1 1)	13.77	27.03	25.63	22.14	18.11	0.500	-0.607	0.575	-1.000	0.573	Rejected	> 15%
Without interventions and with logarithmic transformation	(0 1 1)(0 1 1)	23.71	24.72	24.27	24.23	58.97			-0.019		0.630	Rejected	> 15%
	(0 1 2)(0 1 1)	23.56	25.42	26.37	25.11	64.51			-0.015	-0.126	0.614	Rejected	> 15%
	(2 1 0)(0 1 1)	23.00	26.22	27.78	25.67	62.91	0.021	0.103			0.617	Rejected	> 15%
	(0 2 2)(0 1 1)	10.45	48.30	57.14	38.63	65.33			0.989	0.011	0.477	Rejected	> 15%
	(2 1 2)(0 1 1)	25.70	26.60	25.21	25.84	71.41	0.531	-0.678	0.586	-1.000	0.562	Rejected	> 15%
With interventions and without logarithmic transformation	(0 1 1)(0 1 1)	6.78	14.51	22.13	14.47	82.73			0.016		0.713	Accepted	
	(0 1 2)(0 1 1)	6.68	14.40	21.72	14.27	78.29			0.016	0.033	0.706	Accepted	
	(2 1 0)(0 1 1)	6.64	14.34	21.21	14.06	77.35	-0.016	-0.026			0.709	Accepted	
	(0 2 2)(0 1 1)	30.40	40.65	49.83	40.29	78.55			1.016	-0.016	0.735	Rejected	> 15%
	(2 1 2)(0 1 1)	6.60	14.35	23.65	14.86	73.56	-0.356	0.487	-0.382	0.618	0.658	Accepted	
With interventions and with logarithmic transformation	(0 1 1)(0 1 1)	6.70	15.32	23.38	15.13	91.57			0.045		0.889	Rejected	> 15%
	(0 1 2)(0 1 1)	6.74	15.49	23.60	15.28	87.90			0.046	-0.012	0.895	Rejected	> 15%
	(2 1 0)(0 1 1)	6.49	14.55	21.70	14.25	88.13	-0.046	0.003			0.892	Accepted	
	(0 2 2)(0 1 1)	299.79	459.30	309.20	356.10	91.72			1.037	-0.037	0.855	Rejected	> 15%
	(2 1 2)(0 1 1)	7.47	16.38	23.29	15.72	93.01	-0.175	0.569	-0.219	0.781	0.785	Rejected	> 15%

Table 3. Comparison between the forecast and the performed.

Month	Minimum (confidence Interval)	Forecast	Maximum (confidence Interval)	Performed	Absolute difference	Percentage difference	MSE
Feb. 2004	36.62	42.04	48.27	46.15	-4.11	9.77%	16.89
Mar. 2004	36.25	43.88	53.10	51.26	-7.38	16.81%	54.46
Apr. 2004	34.55	43.59	54.98	53.53	-9.94	22.80%	98.80

4. CONCLUSIONS

The combined use of SARIMA model and of interventions in structural breaks in the time series of soybeans prices in Paraná in the analyzed period could, somehow, treat the negotiating factors known by influence their oscillations, such as seasonality and national and international financial crisis.

The statistical analyses and the forecasts within the confidence interval also corroborate with this hypothesis. Although in absolute values the forecast has become worse at each step, this is easily explained, because the forecast errors of the previous steps are propagated at each new step.

For the future, it is suggested the reproduction of this work for other financial time series of agribusiness commodities, such as alcohol, wheat, maize, and cattle. Other interesting work would use different periods of their own monthly soybean series, but covering impacts due to other acute conjunctural factors, such as the North American super crop of 2004, or tendentious, as the increase in China's demand for soybeans in the last years.

Moreover, we intend to discuss the future of economic manner the achieved results. For example, a relatively small absolute difference (the difference between forecast and

observed value) at the soybean prices early in the first forecast period may lead the producer to make different decisions. These forecasts are an important part of planning for producers and businesses. Performing economic forecasting is essential to reduce the risks associated with the performance and economic viability in short, medium and long-term for soybean producers. Therefore, a greater accuracy of the forecast prices to those commercialized is able to intensify the production.

5. REFERENCES

- Alana, L.A.G. 2005. Unit and fractional roots in the presence of abrupt changes with an application to the brazilian inflation rate. *Empirical Economics*, v. 30, pp. 193-207.
- Alonso, J.C.; Arcila, A.M. 2013. Empleo del comportamiento estacional para mejorar el pronóstico de un commodity: El caso del mercado internacional del azúcar. *Estudios Gerenciales*, v. 29, n. 129, pp. 406-415.
- Asche, F.; Oglend, A. 2016. The relationship between input-factor and output prices in commodity industries: The case of Norwegian salmon aquaculture. *Journal of Commodity Markets*, v. 1, n. 1, pp. 35-47.
- Box, G.E.; Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- BUREAU OF THE CENSUS, X-12-ARIMA Reference Manual. v. 0.3, Washington, DC, Maio de 2007

<http://www.census.gov/srd/www/x12a/> [Acesso em 25 de outubro de 2015].

Campos, J. 1991. A brief look on the literature on deseasonalization. *Revista de Econometria*, Rio de Janeiro, v.11, n. 2, pp. 217-236.

CEPEA, <http://www.cepea.esalq.uspp.br/> [Acesso em 10 de Outubro de 2015].

Cui, H.; Feng, J.; Guo, J.; Wang, T. 2015a. A novel single multiplicative neuron model trained by an improved glowworm swarm optimization algorithm for time series prediction, *Knowledge-Based Systems*, v. 88, pp. 195-209.

Cui, Y.; Shi, J.; Wan, Z. 2015b. Complex Rotation Quantum Dynamic Neural Networks (CRQDNN) using Complex Quantum Neuron (CQN): Applications to time series prediction, *Neural Networks*, v. 71, pp. 11-26.

Dagum, E.B. 1978. Comments on a survey and comparative analysis of various methods of seasonal adjustment by John Kuiper. Conference on Seasonal Analysis of Economic Time Series, US Department of Commerce, Bureau of Census.

Doornik, J.A. 2011. Ox Workshopp. IX ESTE - Escola de Séries Temporais e Econometria – Belo Horizonte.

Durbin, J. 1983. Theory and practice in time series analysis. Conference on Applied Time Series Analysis of Economic Data, US Department of Commerce, Bureau of Census.

Fávero, L.L.; Oliveira, M.A.; Angelo, C.F. 2003. Aplicação de métodos de ajustamento sazonal em séries temporais. VI Semead - Seminários em Administração FEA-USP, São Paulo.

Ferraz, M.I.F.; Sáfyadi, T. 2007. Sazonalidade do custo da ração essencial mínima nas regiões metropolitanas de Belo Horizonte, São Paulo, Porto Alegre e Rio de Janeiro, antes e após o Plano Real. *Informações Econômicas*, SP, v. 37, n. 8.

Findley, D.F.; Monsell, B.C.; Bell, W.R.; Otto, M.C.; Chen, B.C. 1996. New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program. U.S. Bureau of the Census.

Guedes Filho, O.; Vieira, S.R.; Chiba, M.K.; Grego, C.R. 2010. Geostatistical analysis of crop yield maps in a long term no tillage system. *Bragantia*, v.69, p 9-18.

Hassler, U. 1999. When Should cointegrating regressions be detrended? The case of a German money demand function. *Empirical Economics*, v. 24, pp. 155-172.

Hotta, L.K.; Cazorla, I.M.;1990. X-11 seasonal adjustment program options. *Revista de Econometria*, Rio de Janeiro, v. 10, n. 1, pp. 161-179.

IBGE, <http://www.ibge.gov.br/>, [Acesso em 10 de Outubro de 2015].

Kim, H.S.; Kang, D.S.; Kim, J.H. 2003. The BDS Statistic and Residual Test. *Stochastic Environmental Research and Risk Assessment*, v. 17, pp. 104–115.

Liu, Y.K.; Xie, F.; Xie, C.L.; Peng, M.J.; Wu, G.H.; Xia, H. 2015. Prediction of time series of NPP operating parameters using dynamic model based on BP neural network. *Annals of Nuclear Energy*, v. 85, pp. 566-575.

Marini, M.; Moauro, F. 2006. Seasonal Adjustment Procedures Using a Related Series: an Application on the Industrial Production Index. Conference on Seasonality, Seasonal Adjustment and Their Implications for Short-Term Analysis and Forecasting.

Morettin, A. 2006. *Econometria Financeira: Um Curso em Séries Temporais Financeiras*. 17ª SINAPE.ABE. 24 a 28/07/2006.

Morettin, A.; Toloi, C.M.C. 1987. *Previsão de Séries Temporais*. 2ª ed. São Paulo: Atual Editora.

Noland, R.B.; Quddus, M.A.; Ochieng, W.Y. 2007. The effect of the London congestion charge on road casualties: an intervention analysis. *Transportation*, Springer Netherlands.

Oliveira, A.X. 1998. Padrão sazonal da produção industrial, exportações e importações: uma aplicação do X-12-ARIMA. EAESP/FGV/NPP – Núcleo de Pesquisas e Publicações, Relatório de Pesquisa.

- Pai, F.; Lin, C.S. 2005. Using support vector machines to forecast the production values of the machinery industry in Taiwan. *The International Journal of Advanced Manufacturing Technology*, Springer-Verlag, v. 27, pp. 205–210.
- Prema, V.; Rao, K.U. 2015. Development of statistical time series models for solar power prediction, *Renewable Energy*, v. 83, pp. 100-109.
- Ribeiro, V.S.; Dias, J. 2004. Índice de atividade econômica: os modelos de filtro de Kalman e Box-Jenkins comparados. *Anais do VII Encontro de Economia da Região Sul – ANPEC SUL*, n. 7, Maringá.
- Shiskin, J.; Young, A.H.; Musgrave, J.C. 1967. The X-11 variant of the census method II seasonal adjustment program. Bureau of Census, Technical paper 15, US Department of Commerce.
- Silva, M.O.; Guedes Filho, O.; Krause, E.F.; Ramos, G.L.; Polizel, A.C. 2014. Densidade do solo e porosidade total em sistemas de manejo e de semeadura da cultura da soja. VI Congresso sobre Uso e Manejo do Solo, Recife-PE.
- Souza, R.C.T.; Coelho, L.S.; Mariani, V.C.; Steiner, M.T.A. 2009. Coffee Price Forecasting Through RBF Neural Network Trained by Kalman Filter Optimized by Genetic Algorithms. *Proceedings of the 1st International Congress of Mathematics, Engineering and Society - ICMES 2009*. Curitiba-PR: PUCPR.
- Souza, R.C.T. 2008. Previsão de Séries Temporais Utilizando Rede Neural Treinada por Filtro de Kalman e Evolução Diferencial, *Dissertação (Mestrado), Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba-PR*.
- Souza, R.C.T., Steiner, M.T.A., Coelho, L.S. 2013a. Proposta de Wrapper de Seleção de Atributos Combinando Algoritmos Culturais e de Estimação de Distribuição e o Classificador Naïve-Bayes. XX Simpósio de Engenharia de Produção, Bauru-SP.
- Souza, R.C.T.; Steiner, M.T.A.; Coelho, L.S. 2013b. Transformação Geométrica para Melhoria da Acurácia de Classificação de Dados Nominais no Processo KDD. XX Simpósio de Engenharia de Produção, Bauru-SP.
- Thomas, J.J.; Wallis, K.F. 1971. Seasonal variation in regression analysis. *Journal of the Royal Statistical Society*, v. 134, pp. 57-72.
- USDA, U.S. 2015. Department of Agriculture, 26 de Maio de 2004 <http://www.usda.gov/> [Acesso em 20 de outubro de 2015].
- Vieira, S.R.; Guedes Filho, O.; Chiba, M.K.; Mellis, E.V.; Dechen, S.C.F.; Maria, I.C. 2010. Variabilidade espacial dos teores foliares de nutrientes e da produtividade da soja em dois anos de cultivo em um latossolo vermelho. *Revista Brasileira de Ciência do Solo*, v. 34, pp. 1503-1514.
- Werner, L.; Ribeiro, J.L.D. 2003. Previsão de demanda: uma aplicação Box-Jenkins na área de assistência técnica de computadores pessoais. *Gestão & Produção, São Carlos*, v. 10, n. 1, pp. 47-67.
- Zhang, Y.; Hiller, J. 2007. Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. *The International Journal of Biometeorology*.