Analysis of the price of the apple using a SARIMA model

Jaime Antonio Ruiz Hernández[§]
Gerónimo Barrios Puente
Alma Alicia Gómez Gómez

Division of Economic and Administrative Sciences-Autonomous University Chapingo. Mexico-Texcoco Highway km 38.5, Chapingo, State of Mexico, Mexico. CP. 56230. Tel. 01 (595) 9521674. (gbarriospuente55@gmail.com; almaaliciamx@yahoo.com).

[§]Corresponding author: jaime_r9@hotmail.com.

Abstract

The monitoring and prediction of prices is an important tool in decision making in productive activities. Based on the monthly average price data of the red delicious apple at a national level, from January 1998 to July 2017, published by the national system of information and market integration, with the Gretl® software. The objective of the research was to generate information on the price behavior of the apple by means of a time series analysis, generating a prediction for the following 12 months with a Sarima model ARIMA $(2, 1, 0) \times$ ARIMA $(1, 1, 0)_{12}$. The results indicated that the relative error for the prediction is 2%, which guarantees relatively good forecasts, it was predicted that the price for July 2018 would be $31.99.

Keywords: ARIMA, agricultural prices, planning, prediction, time series, uncertainty.

Introduction

The production of food in the world satisfies a wide range of functions of vital importance for the welfare of humanity. Besides generating the world's food, they are also indispensable as a means of subsistence for billions of people as a generator of employment, income and for contributing to rural and general development (OECD-FAO, 2016).

Among them, fruit production stands out, which from 2005 to 2014 had an annual increase in exports of 8.8% and imports of 8.2%. Therefore, fruit trade has practically doubled in the last ten years, with the most important being bananas, grapes and apples (SIECA, 2016).

The fruits of temperate climate constitute one of the most important sources of income of this type of regions since they are more labor-intensive and, therefore, are a greater source of employment. These crops occupy the first place in cultivated area and the second place in the total value of total agricultural production in Mexico (Smith and Somerset, 2003).

On the other hand, the apple in Mexico is one of the five most consumed fruits nationwide with a consumption of 9 kg per year per inhabitant. Its consumption in the country is around 1 000 000 tonnes (t) of which imports 200 000 t. The main national producers are: Chihuahua, whose planted area is 43% of the national surface, Durango 18%, Puebla 14% and Coahuila 12%. Imports come mostly from the United States of America, occupying 94% of them and 5% for Chile (SAGARPA, 2015).

Therefore, the predictive models in the activity of the apple can benefit the producers, packers and buyers, since it can improve the efficiency in making both market decisions and production planning (Logan *et al.*, 2016).

In this sense, for the understanding of a market, it is important to consider the response of prices to changes in supply and demand, as evidenced by the work of Ahrens *et al*. (2014), as well as the importance of price behavior, since for business cycles, for example, it is an important factor in decision making. Also, an inertia in prices comes to cause shocks (unpredictable unexpected event that affects an economy, positively or negatively) permanent demand or supply (Noussair *et al*., 2015).

To acquire information, economic agents, conceptualize it and take it into account for future decisions, especially prices, it is also pointed out that small price variations can be taken by agents without further analysis, unlike larger variations, where the opposite happens and the decision has to be deliberated (Spears, 2014).

Markets and prices are the central scenario in economic theory, especially its operation and training; however, they are not easy to model, therefore, frequently, the price analysis is aimed at describing the dynamics of the market, although they can be described based on some static analysis or comparative statics (Flam and Godal, 2008).

In this regard, the components of the price that are typically considered in international economy are the production cost, mark ups (value that the economic agent adds to the cost of production) and transaction costs where it is indicated that the cost of production and mark ups they are the main source of price variation (Yilmazkuday, 2014), although one should be aware that economic agents are not always perfectly informed about prices (Chudik, 2012).

The information contained in the market prices is fundamental to the assumption that a market of price takers is being dealt with. Therefore, the efficiency with which market information is processed is important (Nelson, 1995). It becomes necessary, the construction of a useful model, which serves to carry out price predictions (Shah and Ghonasgi, 2016).

The monitoring of prices is important within the budget process, helping to establish objectives, in a regional and national framework, it is essential to establish a price tracking system. The objective of this work was to analyze the price of the apple in order to provide information to the main economic agents of this sector, with a prediction based on a Sarima model that serves as a guide for decision making.

It should be mentioned that this type of model is based on quality data that facilitate its estimation and evaluation in statistical packages (Jose and Sojan, 2013). For the purposes, price will be understood as the amount of money paid in a transaction (Fetter, 2016). But the concept is more than that, given that prices have a decisive influence on the market economy (Van Dalen and Thurik, 1998). For example, one can speak of price policy as one of the main factors for competing in the world market (Van Dalen and Thurik, 1998).

Materials and methods

For the analysis, the price series of the red delicious apple (*Malus domestic*) was used on a monthly basis, published by the national information and market integration system (SNIIM) on its website. Data were taken from January 1998 to July 2017 with a total of 235 observations which were analyzed with the GRETL® software.

Taking into account that a time series is a succession of observations of a variable measured at regular intervals of time, its analysis objective being to know the behavior of said variable; through time, to be able to make predictions about it, clarifying uncertainty for economic agents (Parra, 2011).

To complete the objective, a univariate analysis of time series was applied, that is, they use the same information contained in the previous values of the series. In this regard, it should be borne in mind that a time series consists of trend, cyclical fluctuation, seasonal variation and irregular movements or error; however, in a specific series, the four components do not have to be given (Parra, 2011).

Within the previous foundations for its quantification, it is worth noting that in 1970, Box and Jenkins developed a methodological body destined to identify, estimate and diagnose dynamic models of time series in which the time variable plays a fundamental role (Arce, 2010). Concepts that have become later in the integrated autoregressive models of moving averages (ARIMA).

In this context, a model is defined as autoregressive if the endogenous variable of a period t is explained by observations of itself corresponding to previous periods, adding, as in structural models, an error term. In the case of stationary processes with normal distribution, the statistical theory of stochastic processes says that, under certain preconditions, all Y_t can be expressed as a linear combination of its past values plus an error term (Arce, 2010).

One of the most frequent approaches to manage seasonality in the time series has been to calculate it using a seasonal decomposition procedure, since the analysis of its predictions shows a higher precision, indicating that an estimate is more efficient combining the seasonal component than with a single individual pattern. However, there seems to be no consensus on which model is preferable to use for this type of analysis and prediction (De Gooijer and Hyndman, 2006).

For its quantification, it is assumed that the series can be represented by $Y_t = 1, \dots ,T$ being Y_t a monthly random variable that follows a SARIMA model $(p,d,q)(P,D,Q)_{12}$, in such a way that we proceed to identify and estimate the model, which generally is:

$$\Phi\left(L^{12}\right)\phi(L)\Delta^D\Delta^d Y_t = \mu + \Theta\left(L^{12}\right)\theta(L)\varepsilon_t$$

Where: L= is the delay operator; μ= is the mean of the differentiated series, $\phi(L)$, $\theta(L)$, $\Phi\left(L^{12}\right)$ and $\Theta(L^{12})$ are polynomials of order p, q, P and Q respectively, with the short-term dynamics collected by the first two and the seasonal dynamics by the last two, $\phi(L)$ and $\Phi\left(L^{12}\right)$ forming the autoregressive part of the model, while $\theta(L)$ and $\Theta(L^{12})$ make up the moving average. $\Delta^d = (1\text{-}L)^d$ and $\Delta^D = (1\text{-}L^{12})^D$, where d= is the order of integration and D= is the order of seasonal integration. Finally, we assume that ε_t= is a white noise with Gaussian distribution i. i. d. $N(0, \sigma_\varepsilon^2)$.

Results and discussion

In the time series of the price of the apple from January 1998 to July 2017, a variance was noticed that is increasing; that is, the oscillation within a year is getting bigger and bigger. To minimize this variance, the series was transformed into logarithms (Figure 1).

In Figure 2, after having converted the series to logarithms, a reduction in variance was analyzed.

According to the methodology of Box and Jenkins (1970), the series of logarithms must be transformed into first differences, since we intend to work with a stationary series in variance and trend. For the series to be stationary in trend, then the first difference is applied.
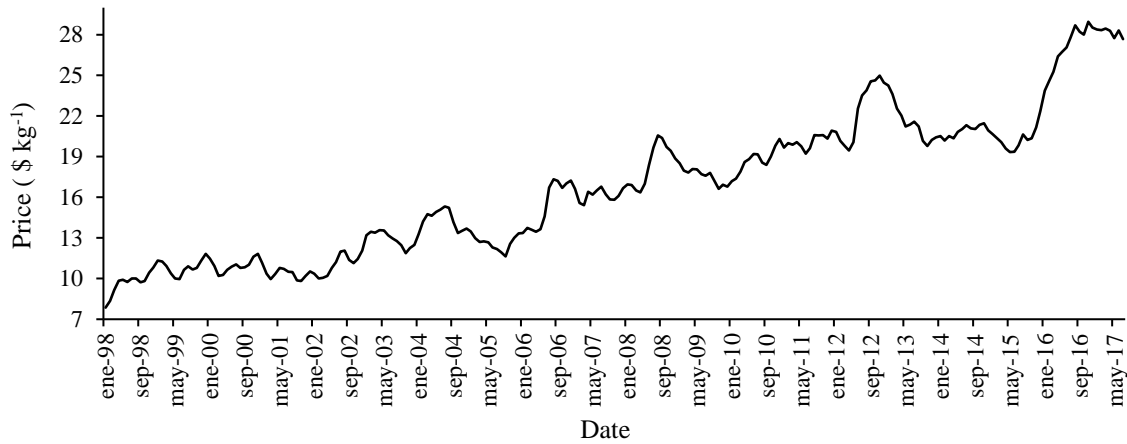
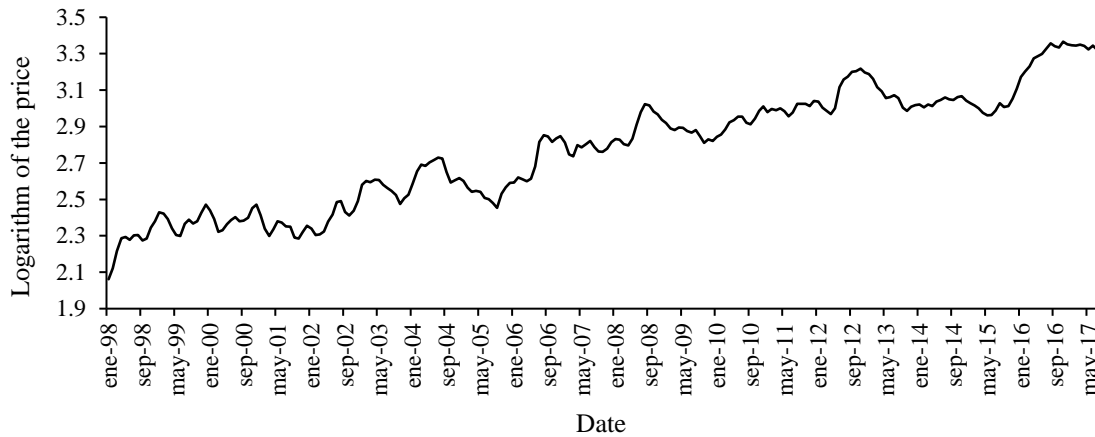**Figure 1. Price of the apple.** Elaboration with SNIIM data.

Figure 2. Logarithm of the price of the apple. Elaboration with SNIIM data.

In Figure 3, in first differences, it is perceived that the series oscillates around 0, no longer has a tendency and its variance is more or less constant. But, a seasonal component is detected, therefore, a transformation is made at 12 months or seasonal difference.
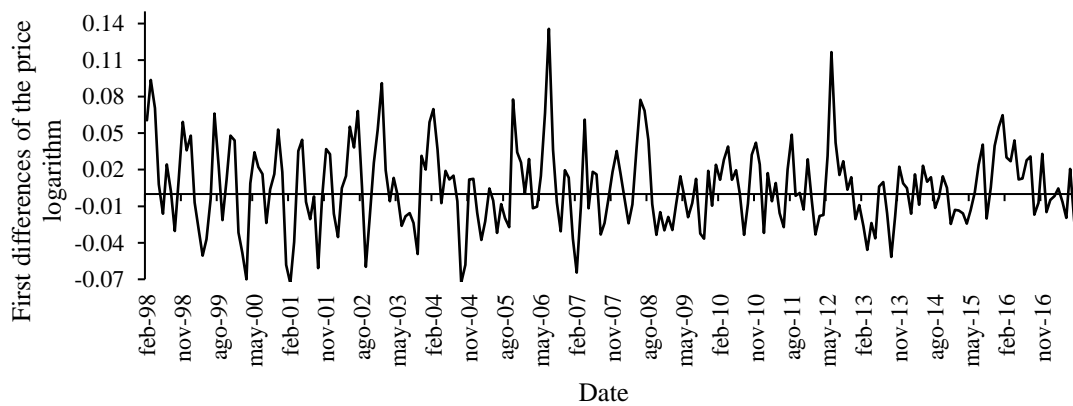

Figure 3. First differences of the logarithm of the price of the apple. Elaboration with SNIIM data.

In the Figure 4 shows the first regular and seasonal difference, from this transformation is where we work. The correlogram is obtained to identify the type of pattern it contains.
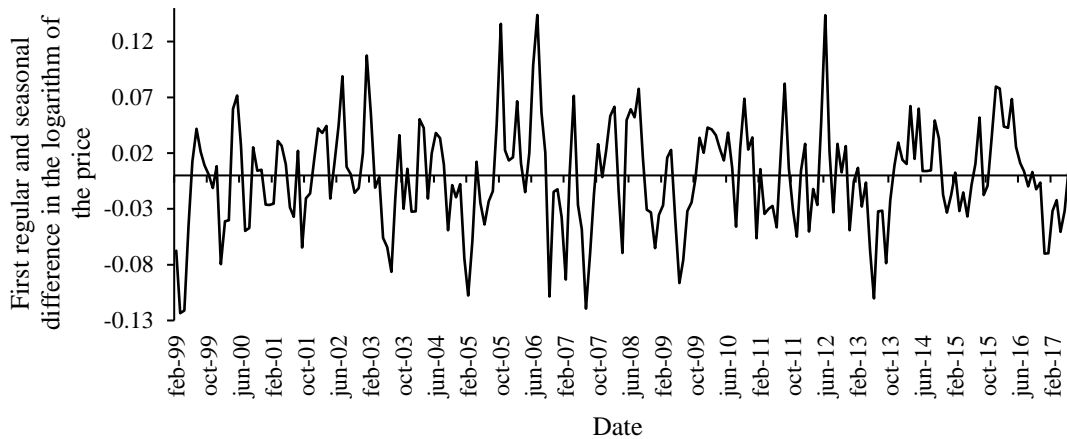

Figure 4. First regular and seasonal difference of the series in logarithms of the apple. Elaboration with SNIIM data.

The Dickey-Fuller unit root test was applied in Table 1, where it was perceived that, given the null hypothesis that the series has a unit root is rejected, it is concluded that it is a seasonal series.

Table 1. Dickey-Fuller unit root test.

Null Hypothesis: SDLP has a unit root			
Exogenous: constant			
Lag length: 12 (automatic - based on SIC, maxlag=12)			
		t-Statistic	Probability[*]
Augmented Dickey-Fuller test statistic		-6.0285915	3.26E-07
Test critical values:	1%	-3.4625736	
	5%	-2.8756082	
	10%	-2.5743463	

[*] = MacKinnon (1996) one-sided *p*-values. Source: results obtained in Eviews 9.

The correlogram of Figure 5 was analyzed, in order to determine the type of model that best suits the series in the simple and partial autocorrelation functions, it was observed that it can be an autoregressive model of order 2 in the regular part and in the seasonal part of an autoregressive of order 1, being able to be identified as a typical process of an economic series according to Arce (2011).
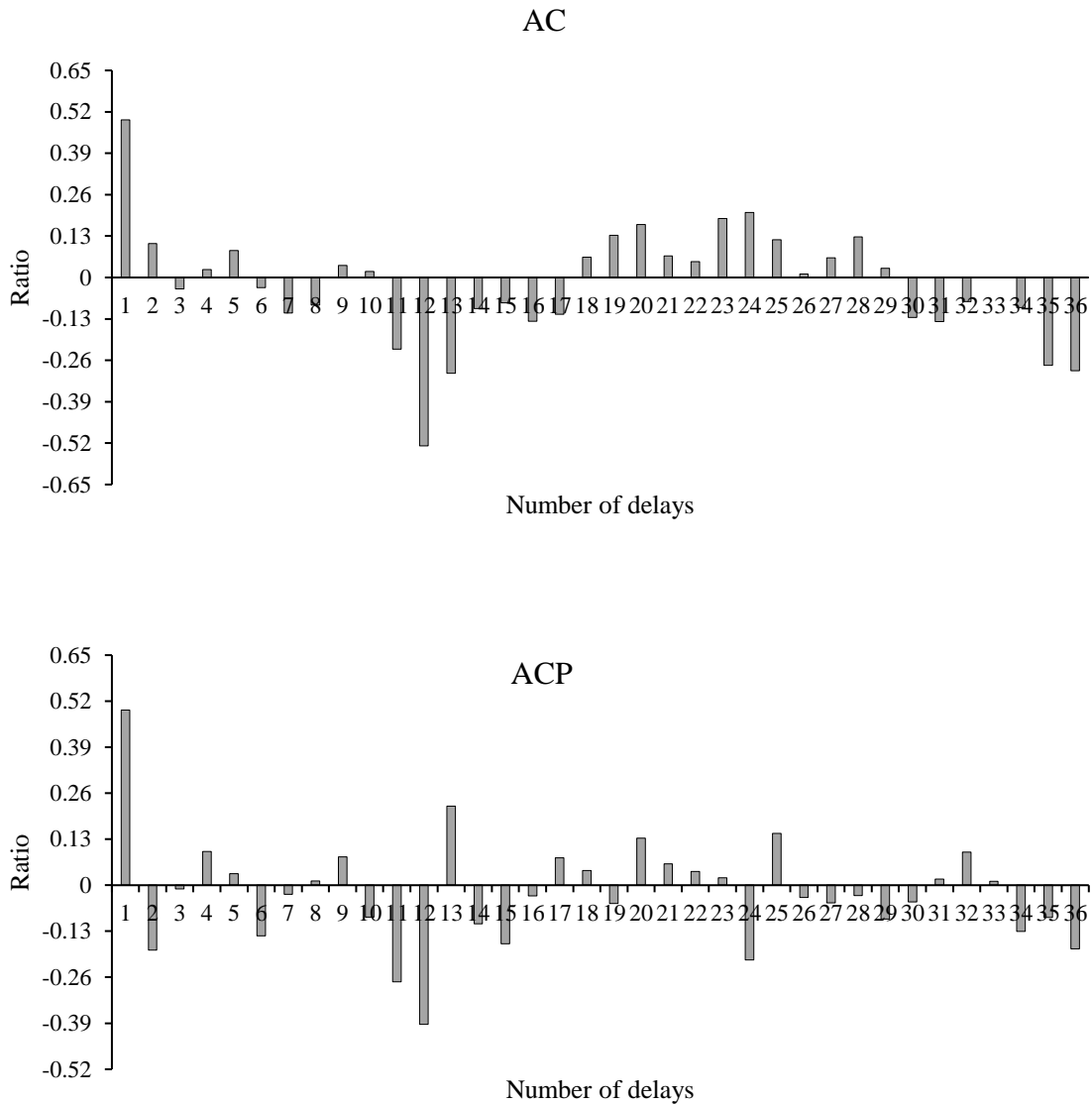
AC

ACP

Figure 5. Correlogram of the first regular and seasonal difference of the series in logarithms of the apple price. Results obtained in Gretl® and edited in Excel.

Then we select an ARIMA $(2, 1, 0)\times$ARIMA$(1, 1, 0)_{12}$ model over the series in logarithms of the price. The model is run in Gretl® to obtain the results of Table 2, which highlights that both the regular and seasonal part are significant. It should be noted that Gretl® calculates by the maximum likelihood method.

Table 2. Sarima model of the price in logarithms of the apple.

Model: ARIMA, using observations 1999:02-2017:07 (T= 222)
Estimated using the Kalman filter (exact MV)
Dependent variable: (1-L)(1-Ls) l_P
Typical deviations based on the Hessian

Coefficient standard deviation		z	value p	
phi_1 0.554395	0.0668646	8.291	1.12e-16 ***	
phi_2 -0.152999	0.0688784	-2.221	0.0263 **	
Phi_1 -0.540887	0.0567055	-9.539	1.45e-21 ***	

Mean of the dependent variable -0.001675		D.T. of the dependent variable 0.046708	
Mean innovations	-0.001225	D.T. innovations	0.033422
Log-likelihood	437.2424	Akaike criterion	-866.4848
Schwarz criterion	-852.8741	Hannan-Quinn criterion	-860.9896

	Real	imaginary	Frequency	module
AR				
Root 1	1.8118	-1.8037	2.5566	-0.1246
Root 2	1.8118	1.8037	2.5566	0.1246
AR (seasonal)				
Root 1	-1.8488	0.0000	1.8488	0.5000

Results obtained in Gretl®.

The result is presented in the following equation:

$$\left(1-0.56L+0.16L^2\right)\left(1+0.54L^{12}\right)LnY_t= \varepsilon_t$$

It is proceeded to verify that the contrast of normality of waste actually passes as it is distinguished in Figure 6 and in fact, it was found that the errors are distributed normally.
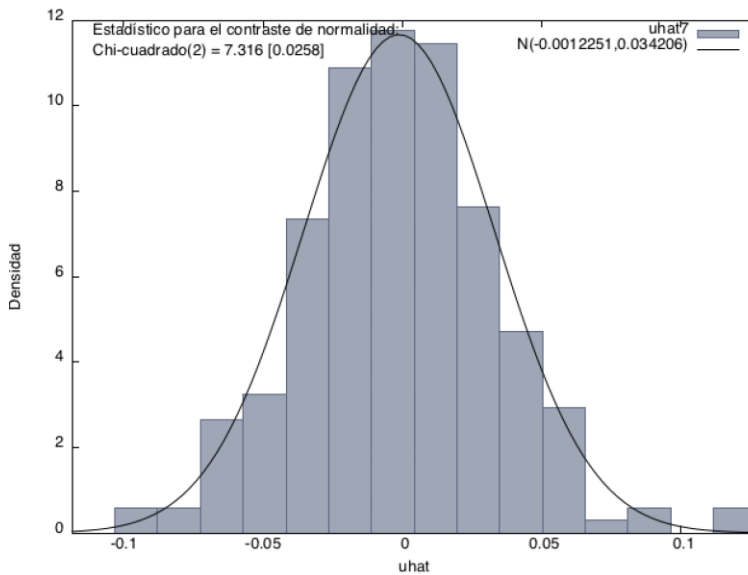

Figure 6. Contrast of normality of the errors of the Sarima model of the price in logarithms of the apple. Source: results obtained in Gretl®.

The contrast of the correlogram or correlogram of the residues was made in Figure 7, where it was verified that we are indeed in the presence of white noise. It is reaffirmed that it is before an ARIMA $(2, 1, 0) \times$ARIMA$(1, 1, 0)_{12}$ model.
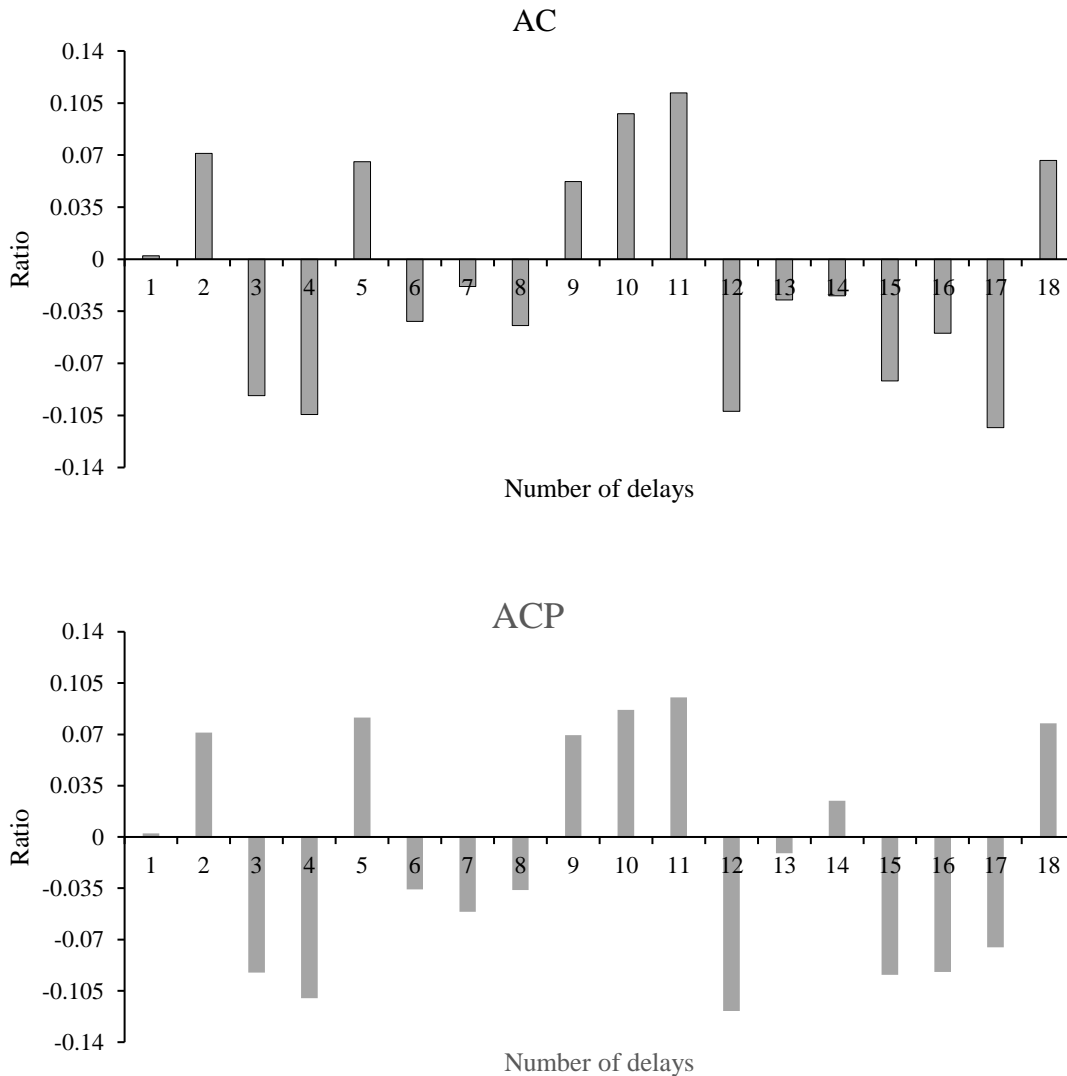

Figure 7. Correlogram of the residuals of the Sarima model of the price in logarithms of the apple. Results obtained in Gretl® and edited in Excel.

It is proceeded to make a forecast, which will be for one year before the last data; that is, 12 observations forward in time. As expected, as the value moves away in time, the confidence interval increases as it is perceived in Figure 8, where it only shows 100 observations.
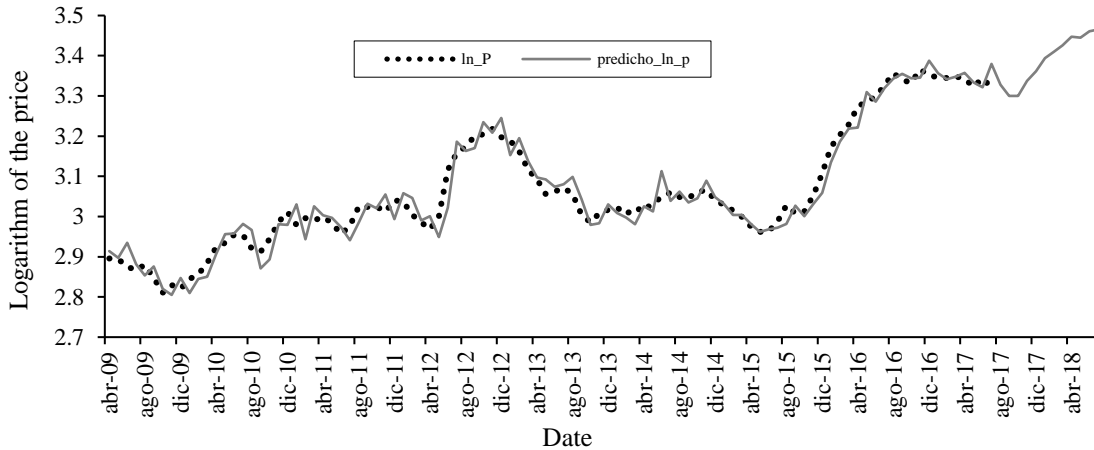
Figure 8. Prediction of the Sarima model of the logarithm price of the apple for one year. Results obtained in Gretl® and edited in Excel.

Thus, the prediction of the price for the following months would be represented by Table 3, where the real price is also compared against its predicted and the relative error is calculated from July 2015 to July 2017, with the relative error being around 2%, which indicates that the prediction of the model can be relatively good. This prediction does not indicate an upward price trend until the end of the same in July 2018. Following the forecast generates a price of $31.99 for the month of July 2018, which is an increase of 4 pesos with respect to July 2017.

Table 3. Predicted price and relative error.

Date	Observed price	Predicted price	Relative error	Date	Observed price	Predicted price	Relative error
Jul.15	19.83	19.53	0.0152	feb-17	28.34	28.25	0.003
Ago.15	20.65	19.71	0.0454	mar-17	28.47	28.44	0.0011
Sep.15	20.24	20.63	0.0194	abr-17	28.29	28.72	0.0153
Oct.15	20.35	20.11	0.012	may-17	27.74	28.03	0.0104
Nov.15	21.17	20.72	0.0213	jun-17	28.32	27.71	0.0217
Dic.15	22.37	21.3	0.0479	jul-17	27.68	29.36	0.0609
Ene.16	23.87	22.98	0.0372	ago-17		27.87	
Feb.16	24.6	24.19	0.0166	sep-17		27.12	
Mar.16	25.27	25	0.0107	oct-17		27.11	
Abr.16	26.41	25.05	0.0515	nov-17		28.14	
May.16	26.73	27.37	0.0241	dic-17		28.82	
Jun.16	27.07	26.73	0.0125	ene-18		29.79	
Jul.16	27.83	27.65	0.0065	feb-18		30.25	
Ago.16	28.7	28.3	0.0141	mar-18		30.76	
Sep.16	28.22	28.65	0.0152	abr-18		31.41	
Oct.16	28.02	28.33	0.0112	may-18		31.33	
Nov.16	28.96	28.36	0.0206	jun-18		31.85	
Dic.16	28.53	29.59	0.0371	jul-18		31.99	
Ene.17	28.39	28.7	0.0109				

Elaboration with Gretl® results and edited in Excel.

From the results obtained, it means that the estimated model represents a good statistical adjustment of the time series of the price of the apple, given that a relative error of only 2% is handled when in Wei *et al*. (2010) a relative error of 5% is accepted. Regarding the identification of the model Caivano *et al*. (2016), indicates that with seasonal series a double or triple interaction are sufficient, to adjust it.

Unlike Yoonsuk and Wade (2016), the model was adjusted without the correction of outliers. However, as anticipated by Findley *et al*. (2016), with a difference in the seasonal part is enough to adjust the model. However, despite the fact that the use of this type of models has increased in recent years, the opinion persists that there is no consensus on which type of model is the one that best fits the economic data that present seasonality (Franses and Van Dijk, 2005).

Comparing with the ARIMA methodology, sugar production prediction works in Mexico have models with an accuracy in the forecast of 94%, such as that of Ruiz *et al*. (2011), which has an ARIMA structure (1, 2, 0).

Other production prediction models, that contemplate both ARMA structures, such as ARIMA, are those presented for the production of milk and the production of pork. In the works of Sánchez *et al*. (2013) and Barreras *et al*. (2013) in which it is accepted that they are useful to establish only short-term forecasts, suggesting the use of multivariate models for a longer term.

Martínez and Chalita (2011) use an ARIMA model (23, 0, 1) to make a 12-month forecast on the tomato price, where they conclude that the current and future prices of the vegetable can be explained by their past prices.

In any case, the presentation of a unique and universally accepted model is unreal and possibly unnecessary. The important thing is that the model is well specified, so that it is well understood and analyzed, coinciding with the structure of the series (Maravall, 1993).

Finally, as Franses and Van Dijk (2005) warn, simple seasonal models offer a better prediction in the short term, while more elaborate models can serve to predict in the long term.

Conclusions

Through the analysis of the time series of the red delicious apple price, from January 1998 to February 2017, a Sarima ARIMA $(2, 1, 0) \times ARIMA(1, 1, 0)_{12}$ model was established, in which since the errors are distributed in a normal way; that is, it is in the presence of white noise. The adjustment of the results of the model was obtained in a convincing and practical way in the Gretl® software. In addition, apple prices were forecast for the months of August 2017 to July 2018, where it should be noted the trend to rise them, with a relative error around 2%, which is close to ideal. However, one must be aware of the limitation of the prediction, since they are predicted values and the economic dynamics ergo that of prices will always be complex.

Cited literature

Ahrens, S.; Pirschel, I. and Snower, D. J. 2014. Theory of price adjustment under loss aversion. J. Econ. Behavior Organization. 134:78-95.

Arce, R. 2011. Modelos arima. U.D.I. Econometría e Informática. 31 p.

Barreras, A. S.; Sánchez, E. L.; Pérez, C. L. y Figueroa, F. S. 2013. Uso de un modelo univariado de series de tiempo para la predicción del comportamiento de la producción de carne de cerdo en Baja California, México. Rev. Científ. 23(5):403-409.

Box, G. and Jenkins, G. 1970. Time series analysis: forecasting and control. Holden-Day, San Francisco. 575 p.

Caivano, M.; Harvey, A. and Luati, A. 2016. Robust time series models with trend and seasonal components. SERIEs 7:99-120.

Chudik, A. 2012. A simple model of price dispersion. Econ. Letters. 117(1):344-347.

Van, D. J. and Thurik, R. 1998. A model of pricing behavior: an econometric case study. J. Econ. Behavior Org. 36(2):177-195.

Fetter, F. A. 2016. The Definition of price published by: american economic association stable. http://www.jstor.org/stable/1828191. The Am. Econ. Review. Am. Econ. Assoc. 2(4):83-813.

Findley, D. F.; Lytras, D. P. and Marvall, A. 2016. Illuminating ARIMA model-based seasonal adjustment with three fundamental seasonal models. SERIEs. 7:11-52.

Flam, S. D. and Godal, O. 2008. Market clearing and price formation. J. Econ. Dynamics Control. 32(3):956-977.

Franses, P. H. and Van, D. D. 2005. The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production. Inter. J. Forecasting. 21(1):87-102.

de Gooijer, J. G. and Hyndman, R. J. 2006. 25 years of time series forecasting. Inter. J. Forecasting. 22(3):443-473.

Jose, J. and Sojan, L. P. 2013. Application of ARIMA(1,1,0) model for predicting time delay of search engine crawlers. Inf. Econ. 17(4):26-38.

Logan, T. M.; Mcleod, S. and Guikema, S. 2016. Predictive models in horticulture: a case study with Royal Gala apples. Sci. Hortic. 209:201-213.

Maravall, A. 1993. Stochastic linear trends. Time. 56:5-37.

Martínez, G. M. and Chalita, L. E. T. 2011. Aplicación de la metodología Box-Jenkins para pronóstico de precios en jitomate. Rev. Mex. Cienc. Agríc. 2(4):573-577.

Nelson, J. P. 1995. Market structure and incomplete information: Price formation in a real-world repeated English auction. J. Econ. Behavior Org. 27(3):421-437.

Noussair, C. N.; Pfajfar, D. and Zsiros, J. 2015. Pricing decisions in an experimental dynamic stochastic general equilibrium economy. J. Econ. Behavior Org. 109:188-202.

OECD-FAO. 2016. OCDE-FAO perspectivas agrícolas 2016-2025. Publishing, París. DOI: http://dx.doi.org/10.1787/agr_outlook-2016-es.

Parra, F. R. 2011. Econometría aplicada II. Creative Commons reconocimiento-nocomercial-compartir igual 4.0 Internacional License.

SAGARPA. 2015. SIAP. Available at: http://www.siap.gob.mx/cierre-de-la-produccion-agricola-por-cultivo/.

Shah, C. and Ghonasgi, N. 2016. Determinants and forecast of price level in India: a VAR Framework. The Indian Econometric Society. J. Quant. Econ. 14:57-86.

SIECA. 2016. Análisis de la competitividad regional del mercado de frutas. http://www.sieca.int/portaldata/documentos/d1ccb65b-6edd-4fc4-92e5-609efcf1e227.pdf.

Smith, L. G. and Somerset, S. M. 2003. Fruits of temperate climates commercial and dietary importance. *In*: Encyclopedia of Food Sciences and Nutrition. 2753-2761 pp.

Spears, D. 2014. Decision costs and price sensitivity: field experimental evidence from India. J. Econ. Behavior Org. 97:169-184.

Yilmazkuday, H. 2014. Price dispersion across US districts of entry. Econ. Letters. 123(3):361-365.