Agrociencia

INTERVENTION AND FORECAST MODELS FOR THE PRICE PAID TO PRODUCER OF BEE (*Apis mellifera* L.) HONEY IN MEXICO

Samuel **Luis-Rojas**[1], Roberto Carlos **García-Sánchez**[1]*, Roberto **García-Mata**[1], Oscar Antonio **Arana-Coronado**[1], Benito **Ramírez-Valverde**[2]

[1] Colegio de Postgraduados Campus Montecillo, Posgrado en Socioeconomía, Estadística e Informática-Economía, Carretera México-Texcoco km 36.5, Montecillo, Texcoco, Estado de México, México, C. P. 56230.

[2] Colegio de Postgraduados Campus Puebla, Posgrado en Estrategias para el Desarrollo Agrícola Regional, Km 125.5 carretera federal México-Puebla (Boulevard Forjadores de Puebla), Puebla, México, C. P. 72760.

* Corresponding author: rcgarcia@colpos.mx

Citation: Luis-Rojas S, García-Sánchez RC, García-Mata R, Arana-Coronado OA, Ramírez-Valverde B. 2022. Intervention and forecast models for the price paid to producers of bee (*Apis mellifera* L.) honey in México.
Agrociencia. https://doi.org/ 10.47163/agrociencia.v56i3.2807

Editor in Chief:
Dr. Fernando C. Gómez Merino

Received: August 10, 2021.
Approved: May 16, 2022.
Published in Agrociencia: June 07, 2022.

ABSTRACT

Bee (*Apis mellifera* L) honey is one of the oldest foods that humans have used. Since ancient times, it has been used as a healthy product due to its sweetening and healing properties. In 2020, Mexico produced 54 121 tons (Mg), which ranked the country as the tenth largest producer in the world. The hypothesis was that current honey prices can be explained by previous prices and that they influence the increase in the population of hives and the production of honey in Mexico. To test this hypothesis, the objective of this research was to develop a forecast model for the annual average prices of honey in Mexico (AAPH). The data comprised the 1966 to 2019 prices and the Box-Jenkins methodology of Autoregressive Integrated Moving Average (ARIMA), with and without intervention, was used. The parameters of the models were estimated with the maximum likelihood method of the SAS® software, while the structural change was calculated with the corresponding library (*strucchange*) of the R software. A model based on the AAPH series was adapted for the 1966−2019 period and validated with data from 2018 and 2019. The series presents five periods of trend structural changes of AAPH: 1966−1985; 1986−1995; 1996−2003; 2004−2008; and 2009−2019. The best estimated model without intervention was ARIMA (1, 1, 1) and the best model with intervention was ARIMA (1, 1, 0), which indicates that the prices of previous years can explain the AAPH. The predictions had a mean absolute percentage error (MAPE) of 8.16 % for the model without intervention and 4.02 % for the model with intervention. Both estimated models suggested that the AAPH have an upward trend in the medium term. The ARIMA model with intervention provided a more accurate estimation of the AAPH and information to plan and make decisions for the next five years.

Keywords: ARIMA models, intervention models, beekeeping, livestock planning, predictions.

INTRODUCTION

Beekeeping in Mexico, as a generator of foreign currency, ranks among the top three activities in the livestock sector. The economic income of this activity mainly benefits small producers (Dolores *et al*., 2017). The main product of this activity is honey, which is one of the oldest foods that humans have used to nourish. Additionally, it has been used since ancient times as a beneficial health product, due to its well-known healing properties (Ramos and Pacheco, 2016).

The domestic production of bee (*Apis mellifera* L) honey decreased from 55 687 Mg in 2010 to 54 121 Mg in 2020; an average annual growth rate (AAGR) of −0.31 % was recorded (SIAP, 2019). The decrease is associated with climate change, deforestation, and the use of herbicides and insecticides in agricultural plots. Large drought periods, erratic rainfall, and the lack of organization among producers contribute greatly to the downward trend (Magaña *et al*., 2016).

In 2020, the world production of natural honey was 1 724 195 Mg. Mexico (54 121 Mg), China (447 007 Mg), Turkey (109 330 Mg), and Canada (80 345 Mg) accounted for 3.60, 25.9, 6.3, and 4.6 % respectively of the total production (FAO, 2020). Mexico was the tenth world producer.

China is the first world producer of honey, as a result of the drastic increase in the population of hives, which are used for honey production. However, they also pollinate cotton, rapeseed, buckwheat, apples, citrus, sunflower, vetches, and other crops, whose production volumes significantly increased in the 2000–2011 period. Beekeeping in China has been developed to the point that the country is now the largest exporter in the world, as a result of the low prices it offers (Martínez and Pérez, 2013). However, consumers from importing countries report that Chinese honey lacks safety and traceability (Maté, 2012).

Honey production in Mexico depends on several factors, including floral characteristics, soil, and climate. The Coordinación General de Ganadería of the Secretaría de Agricultura y Desarrollo Rural (SADER) classifies beekeeping activity into five production regions: North, Pacific Coast, Gulf of Mexico, Altiplano, and Yucatan Peninsula (Martínez and Pérez, 2013). From highest to lowest, the honey production (2020) in these regions was divided as follows: Pacific Coast (39.10 %), Yucatan Peninsula (24.09 %), Altiplano (15.67 %), Gulf of Mexico (10.64 %), and North (10.50 %). There are still vast areas of the country where beekeeping can be promoted; however, Jalisco, Chiapas, Veracruz, and Oaxaca contributed 11.20, 10.04, 8.58, and 8.38 % of the domestic production, respectively (SIAP, 2020).

In 2018, the average volume of honey exports from Mexico was 55 674 Mg, ranking the country as the fourth largest exporter. Mexico is the main supplier of Germany, the largest importer in the world, which applies the highest quality standards (SIAP, 2019). The main destinations of Mexican exports were Germany, United Kingdom, USA, and Saudi Arabia, which together accounted for 90 % (FAO, 2020).

The variability of honey price in Mexico is the consequence of biological and climatic factors (Caro *et al*., 2012). Nevertheless, it mainly depends on the Chinese production; therefore, determining the behaviour of this variability and how it influences Mexican prices is fundamental. Autoregressive Integrated Moving Average (ARIMA) models are more appropriate for short-term predictions; they are designed to obtain information about processes that have a certain degree of homogeneity. That is to say, their analysis is based on a stationary series and at least 50 data are needed to achieve a reliable prediction (Box *et al.*, 2015).

Prior knowledge of the time series to be studied is important, since the presence of outliers can produce serious distortions in the results (Segura and Torres, 2014). It is also very likely that they cannot be explained by the ARIMA model and, therefore, violate the assumption of normality. Hence, outliers and structural changes influence the efficiency and goodness of fit of the best proposed ARIMA models.

Economic theory indicates that, in perfect competition, a higher price leads to an increase in supply, while a lower price induces a decrease (Varian, 2010). In the case of honey, the quantity supplied in recent years has not changed (perfectly inelastic supply). Meanwhile, the demand shift, based on the tastes and preferences of the consumer, causes the price to increase. The hypothesis is that honey prices can be explained by prior prices, which influence the increase in the population of hives and the production of honey in Mexico. Under this hypothesis, the objective of this research was to develop time series models from 1966 to 2019, with and without intervention, in order to forecast the average prices of honey (AAPH) in Mexico and to evaluate the functionality of the models.

MATERIALS AND METHODS

In order to determine the behaviour of the average prices paid to the producer of bee (*Apis mellifera* L.) honey in Mexico (AAPH) and to develop forecasts, an annual historical series of prices, expressed in Mexican pesos (MXN \$ kg^{-1}), was used, consulting the Sistema de Información Agroalimentaria y Pesquera (SIAP, 2020) and the Food and Agriculture Organization of the United Nations (FAO, 2020). The AAPH time series was divided into two parts: data from 1966 to 2019 were used to develop the time series models, with and without intervention; and price data from 2018 and 2019 were used to validate the models.

Assuming that $Y' = (Y_1, Y_2, ..., Y_n)$ is a time series, a pure ARIMA model is mathematically denoted as *(p, d, q)* and is expressed as follows:

$$W_t = \mu + \frac{\theta(B)}{\phi(B)}\alpha_t$$

where: t = indexes time; W_t = is the response series Y_t or a difference of the response series; μ = is the mean term; B = is the backshift operator, that is ($B^r Y_t = Y_{t-r}$); $\phi(B)$ is the autoregressive polynomial (AR) of order *"p"*, developed as follows: $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$; $\theta(B)$ = is the polynomial f moving averages (MA) of order *"q"*, where: $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... \theta_q B^q$; α_t = refers to random error terms (also called white noise), random variables independently distributed in an identical way, sampled from a distribution with preferably mean equal to zero and variance $\alpha_t \sim N(0, \delta^2)$ (Box *et al.*, 2015).

The inclusion of the effects of exogenous variables (intervention variables) in the general ARIMA model is achieved through the following transfer function:

$$W_t = \mu + \sum_i \frac{\omega_i(B)}{\delta_i(B)} B^{ki} X_{i,t} + \frac{\theta(B)}{\phi(B)} \alpha_t$$

where: $X_{i,t}$ is the *i*-th input time series or a difference of the series of *i*-th input at time t; k_i is the pure time lag for the effect of the *i*-th input series; w_i (B) is the numerator polynomial of the transfer function for the *i*-th input series; and δ_i (B) is the denominator polynomial of the transfer function for the *i*-th input series. In the intervention analysis, some of the $X_{i,t}$ variables are assumed to be binary variables that play the same role as the dummy variables in the regressions; therefore, the $X_{i,t}$ series are known as intervention indicators or outliers (Ferruz *et al.* 2011).

If the intervention is recurrent in some type of event at a certain moment in time, it can manifest itself in a later time, and temporarily or permanently affect the series under study.

For the analysis and treatment, PROC ARIMA of SAS® software, version 9.4, was used (SAS Institute Inc., 2014). The ARIMA model of the AAPH series for the 1966−2019 period was estimated using the methodology proposed by Box *et al.* (2015), which consists of the construction and adjustment of the forecast model. Meanwhile, the R program, version 3.6.2 (R Core Team, 2019) was used to determine the structural change of level, with the corresponding library (*strucchange*) developed by Zeileis *et al.* (2019).

The choice of the best model with and without intervention was parsimoniously suggested by Rodríguez *et al.* (2017), through the Akaike Information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) that compare the goodness of fit of the different models. Both criteria are based on the use of the sum of squared errors and seek to minimize it, based on various combinations of p and q. Lower AIC and SBC values indicate a better fit to the model. The aim of the graphical analysis was to identify the best model in which, in addition to reducing the mean squared error, the residuals were randomly distributed around zero without showing any pattern or trend. This is an indicator that the data are random and have a normal distribution (Moffat and Akpan, 2019).

RESULTS AND DISCUSSION

The AAPH series data fluctuated over time, following an upward trend. The results of the structural change analysis showed that the trend of the series presents structural changes and that these had an impact on the evolution over time of the data generation process. A trend break of the AAPH occurred in five periods. The first (1966−1985) was characterized by the existence of tariffs on food imports. During this period, prices presented a 19.27 % AAGR. Tariffs protected domestic production from international competition and they were the basis of the food supply for the population (CEDRSSA, 2018).

From 1986 to 1995, Mexican beekeeping suffered a major setback due to the entry of the African bee (*Apis mellifera scutellata*) through the states of Chiapas and Quintana Roo. In 1986, the Africanization process began to affect honey production in the states of Yucatan and Campeche. African bees are characterized by their defensive behaviour, their tendency to take flight, and their high capacity to build swarms or hives. Their beekeeping requires a more technical management and their exploitation demands a greater investment (Cervantes *et al*., 2018).

In 1988, hurricane Gilberto caused a considerable loss of hives and wild swarms in the Yucatan Peninsula, reducing honey production. This generated a constant rise in prices (AAGR: 17.66 %) during the 1986–1995 period.

During the 1996-2003 period, the Africanization of hives resulted in a decrease in production in Mexico, South America, and the southern United States. Meanwhile, the appearance in the Gulf of Mexico (and subsequent dispersion) of the varroa mite (*Varroa jacobsoni Oudemans*), which parasitizes *Apis mellifera* bees (Medina *et al*., 2014), generated a fall in honey production and, consequently, an increase in honey prices to a 6.02 % AAGR.

From 2004 to 2008, there was a decrease in honey production, as a result of hurricanes Wilma and Dean, which mainly affected southeastern Mexico and the Yucatan Peninsula (the most important production regions), causing a partial or total loss of hives. Other problems were the lack of water in other production regions, bee health, reduction of wild areas due to urbanization, and the use of pesticides and agrochemicals that affect bees (Martínez and Pérez, 2013). Consequently, prices fell to a −0.33 % AAGR.

In the 2009–2019 period, prices increased to a 4.31 % AAGR, largely as a result of the awareness of society about the preservation of bees and pollinating species. The most lucrative and attractive market for Mexico is the European Union, which demands organic and transgenic-free honey produced without pollutants; consequently, Mexican honey has positioned itself as a highly appreciated product and the price trend is upward in the medium term (Figure 1).

The Cox-Box test produced a λ-0.5, so the AAPH series was transformed into natural logarithms to keep the variance constant (Vélez *et al*., 2015); now the series was renamed AAPHL. Series Y_1, Y_2, ..., Y_n show that there is still a certain trend in time (Figure 2), but, through the first difference (∇) —*i.e.*, (1-*B*) AAPHL$_t$— a stationary series is obtained. Therefore, $d = 1$.

The AAPH series, differentiated and transformed into natural logarithms, was renamed AAPHL (1); it is intuitively known *a priori* that the series already has a stationary mean (there is no trend) and variance. However, outliers were recorded in 1979, 1981, 1985, and 1991. Therefore, working the series with two methods (with and without intervention) was necessary. Box *et al.* (2015) pointed out that, in order to obtain better forecasts, the series to be studied must have a constant variability throughout time and must not have a trend (Figure 3).

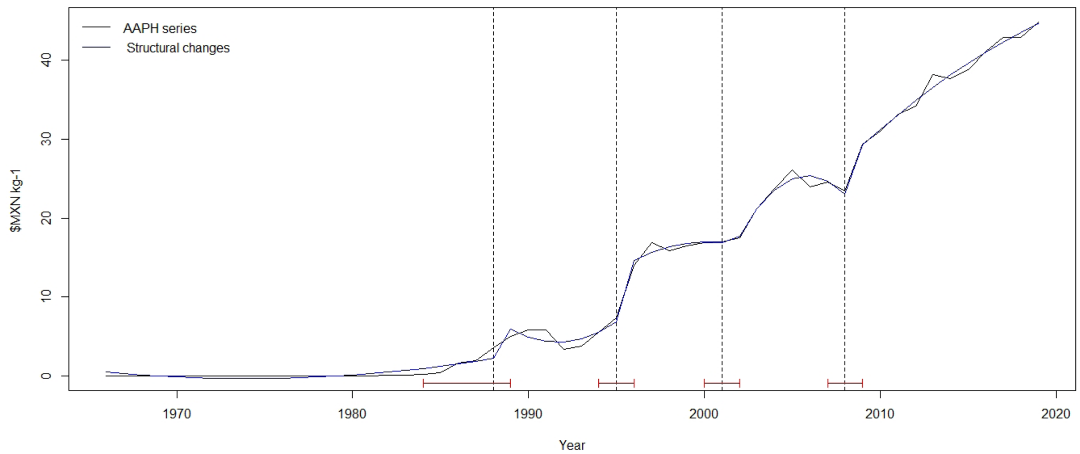

Figure 1. Original behaviour of the AAPH series (in MXN \$ kg^{-1}) and its structural changes (1966 – 2019).

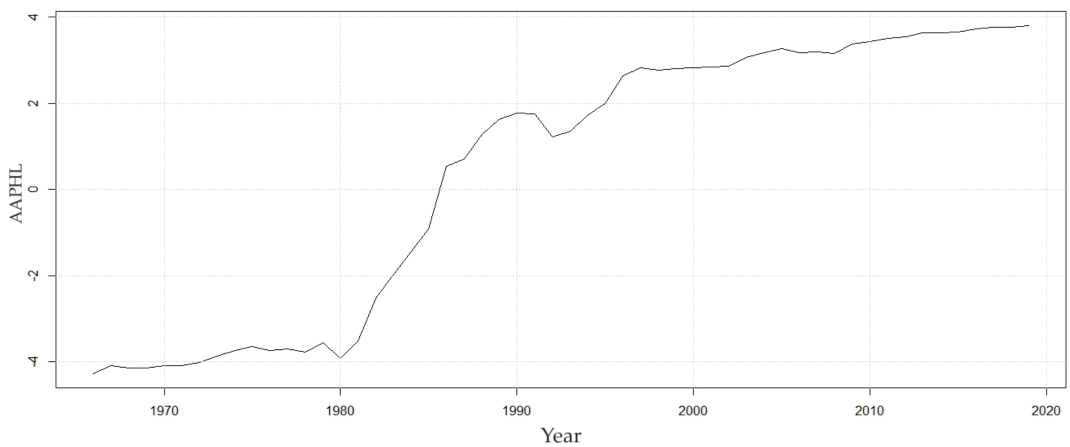

Figure 2. Behaviour of the AAPH series transformed into natural logarithms (AAPHL).

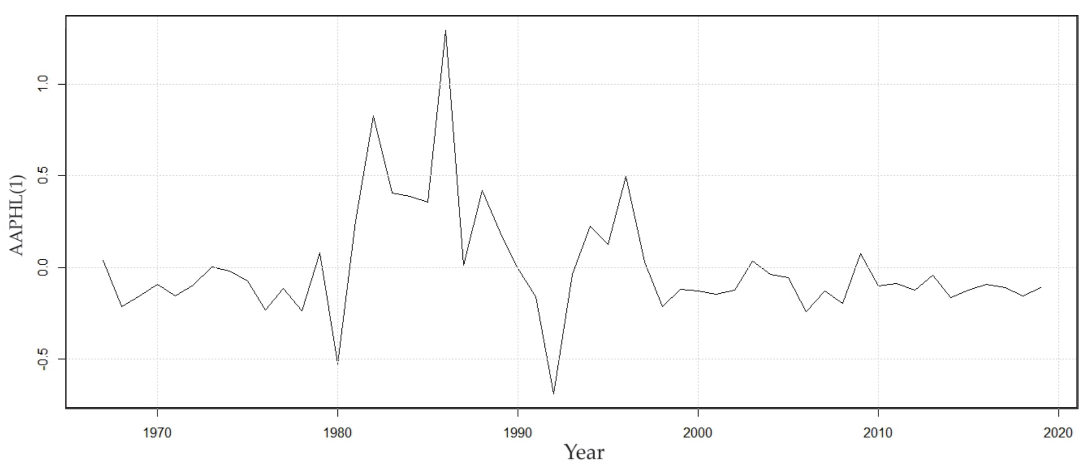

Figura 3. Differentiated AAPHL series, transformed with natural logarithms and with no apparent trend.

To statistically verify the stationarity or non-stationarity of the time series, the Augmented Dickey-Fuller unit root test (ADF) was carried out (Dickey and Fuller, 1981). This test includes lags from the first Y_t difference in the test regression, in order to include the possible existence of serial autocorrelation. Therefore, the following hypothesis is proposed: Ho, the series is not stationary (ρ = 1) and has a unit root *vs*. Ha, the series is stationary ($\rho \neq 1$) and does not have a unit root. Decision rule: Ho is rejected if p-value $\leq \alpha = 0.05$. Since the p-value of the calculated F (0.001) is lower than α = 0.05, the null hypothesis is rejected H_0: δ0 (ρ = 1), reaching the conclusion that AAPHL(1) series does not have a unit root; therefore, it is stationary. Consequently, it have a constant variance and mean over time (Table 1).

For the AAPHL(1) time series, a model was fitted using the PROC ARIMA process (SAS Institute Inc., 2014); the AR1,1 (ϕ_1) and the moving average component MA1,1 (θ_1) parameters were calculated using maximum likelihood. Because this method assumes that its estimators are asymptotically optimal, when the size of the series is large, they are considered to be centered or unbiased, and efficient, and that their distribution is normal (Montemayor, 2013).

Out of the 15 proposed models, the one that best meets the significance of parameters and white noise was identified. The AR and MA coefficients were chosen because the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) simultaneously present coefficients other than zero. Likewise, they were estimated considering different choices of p and q, as well as the values of SBC, AIC, and the variance δ_ε^2 for the four best ARIMA models fitted to the AAPHL(1) series.

The first difference is often enough (d = 1); therefore, it was established in all models. The model with the lowest SBC and AIC value for this data set was ARIMA (1, 1, 1) (Table 2).

The ARIMA (1, 1, 1) model is considered the best moderate model without intervention, since, according to Box *et al.* (2015), the absolute t statistic must be higher than 2 and the p-values of the parameters must be lower than 0.05. Not only is this model parsimonious, it sufficiently fits the old data (Table 3).

For the calculation, the equation of the ARIMA (1, 1, 1) model must be supported by the coefficients in Table 3 (without including the outliers) and by the theoretical

Table 1. Augmented Dickey–Fuller test (ADF) for the differentiated series of logarithms of annual average prices of honey in Mexico (AAPHL(1)).

Kind	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero mean	0	-25.8205	<.0001	-4.12	<.0001		
	1	-14.5181	0.0058	-2.64	0.0092		
	2	-11.4307	0.0153	-2.2	0.0283		
Simple mean	0	-32.2686	0.0005	-4.74	0.0003	11.22	0.001
	1	-21.1904	0.0042	-3.16	0.0282	4.99	0.0418
	2	-18.8792	0.0087	-2.71	0.0792	3.67	0.1653

Table 2. Values of AR, MA, SBC, and AIC of the identified ARIMA models (p, 1, q) and estimators for $\delta_{\in}$.

Best ARIMA[†] models	Lags		Coefficients		SBC[Þ]	AIC[¤]	$\delta_{\in}$
	AR	MA	AR1,1[¶] (p)	MA1,1[§](q)			
1.- (1, 1, 0)	1		0.49727	-	23.75236	21.78207	0.293619
2.- (1, 1, 1)	1	1	0.82419	0.46266	23.67558	19.73499	0.285018
3.- (2, 1, 1)	2	1	0.37660	-0.36982	25.32337	21.38279	0.289676
4.- (1, 1, 2)	1	2	0.42923	-0.15275	26.26991	22.32933	0.292387

[†]ARIMA: Autoregressive Integrated Moving Average process, [¶]AR: Autoregressive coefficient of order (p), [§]MA: Moving average coefficient of order (q), Þ SBC: Schwarz Bayesian Criterion, AIC: Akaike Information Criterion. $\delta_{\in}$: Standard error of estimate.

Table 3. Model estimation for the AAPHL(1) time series by maximum likelihood without intervention.

Parameter	Estimation	Standard error	t-value	Aprox Pr > |t|	Lag
MA1,1	0.46266	0.21077	2.20	0.0282	1
AR1,1	0.82419	0.13183	6.25	<0.0001	1

approach established by Box *et al.* (2015). The following equation was obtained when the model was developed:

$$\text{ARIMA } (1,1,1) = (1 - \phi_1 B_1)\,(1 - B_1)\, Y_t = (1 - \theta_1 B_1)\,\alpha_t$$

$$\text{ARIMA } (1,1,1) = Y_t = Y_{t-1} + \phi_1\, Y_{t-1} - \phi_1\, Y_{t-2} - \theta_1 \alpha_{t-1} + \alpha_t$$

$$\text{ARIMA } (1,1,1) = Y_t = Y_{t-1} + 0.82419\, Y_{t-1} - 0.82419\, Y_{t-2} - 0.46266\, \alpha_{t-1} + \alpha_t$$

The ARIMA (1, 1, 1) model with intervention

Because the study series included level shift (LS) outliers in 1979, 1981, 1985, and 1991, these data were included into the the original ARIMA (1, 1, 1) model in order to improve it. To respect the assumption of parsimony and the significant statistical value of the parameters, this new model is known as the model with intervention (Box *et al.*, 2015). The results showed that the moving average coefficient was not significant; consequently,the ARIMA (1, 1, 0) model with intervention was chosen. In addition to a significant coefficient, there was a significant decrease in the standard error (51.38 %), compared to the ARIMA (1, 1, 1) model without intervention (Table 4).

The ARIMA (1, 1, 0) model was considered the best moderate model with intervention, since —in addition to meeting the assumptions made by the Box-Jenkins methodology— it includes outlier data (Table 5).

Table 4. AR, MA, SBC, and AIC values of the identified ARIMA models with intervention (p, 1, q) and estimators for $\delta_{\in}$.

Best ARIMA[†] models	Lags		Coefficients		SBC[þ]	AIC[¤]	$\delta_{\in}$
	AR	MA	AR1,1[¶] (p)	MA1,1[§] (q)			
1.- (1, 1, 1) [†]a	1	1	0.82419	0.46266	23.67558	19.73499	0.285018
2.- (1, 1, 1) [†]b	1	1	0.88111	0.25647	-38.3316	-52.1237	0.137813
3.- (1, 1, 0) [†]c	1	0	0.79943	-	-40.6168	-52.4386	0.138568

[†]ARIMA: Autoregressive Integrated Moving Average Process, [†]a ARIMA (1, 1, 1) model without intervention, [†]b ARIMA (1, 1, 1) model with intervention, [†]c ARIMA (1, 1, 0) model with intervention [¶]AR: Autoregressive coefficient of order (p), [§]MA: Moving average coefficient of order (q), [þ]SBC: Schwarz Bayesian Criterion, [¤]AIC: Akaike Information Criterion. $\delta_{\in}$: Standard error of estimate.

Table 5. Model estimation for the AAPHL(1) time series by maximum likelihood with intervention.

Parameter	Estimator	Standard error	Value of t	Aprox. Pr > |t|	Lag	Variable	Displacement
AR1,1	0.79943	0.08486	9.42	<.0001	1	AAPHLog	0
NUM1	-0.68236	0.10721	-6.36	<.0001	0	LS_15	0
NUM2	0.51015	0.10722	4.76	<.0001	0	LS_17	0
NUM3	1.11607	0.10796	10.34	<.0001	0	LS_21	0
NUM4	-0.58917	0.10724	-5.49	<.0001	0	LS_27	0
NUM5	0.42211	0.10733	3.93	<.0001	0	LS_31	0

The equation of the ARIMA (1, 1, 0) model with intervention is expressed as:

$$Y_t = -0.68236\xi_{1t} + 0.51015\xi_{2t} + 1.11607\xi_{3t} - 0.58917\xi_{4t} + 0.42211\xi_{5t} + \frac{(1\text{-}0.79943B_1)}{\alpha_t}$$

$\xi_{1t} = 1$ si $t \geq 15$ or other wise
$\xi_{2t} = 1$ si $t \geq 17$ or other wise
$\xi_{3t} = 1$ si $t \geq 21$ or other wise
$\xi_{4t} = 1$ si $t \geq 27$ or other wise
$\xi_{5t} = 1$ si $t \geq 31$ or other wise

To verify the overall sufficiency of the Box-Jenkins model, the residuals obtained from the models without and with intervention were analysed. The Ljung-Box Q^+ (LBQ) statistic and its associated p-value proved the H_0: $\varepsilon_t \sim RB\ (0, \sigma^2)$ null hypothesis. The autocorrelations up to a lag k are equal to zero for k values equal to 6, 12, 18, 24, and 30. The random and independent data values —up to a certain number of lags— *vs*. the H_α: ε_t are not white noise. Abdulhafedh (2017) suggest that, if the Ljung-Box Q (LBQ) statistic is higher than a specified critical value, the autocorrelations for one or more lags could be significantly different from zero, indicating that the values are neither random nor independent in time.

The decision rule is the following: if the p-value < 0.05, H_0 is rejected, but if the p-value > 0.05, H_0 is not rejected. The first 6 k' have p-values > 0.05; therefore, H_0: $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = \rho_6$ is not rejected. An α=0.05 value means that the process is purely random or white noise —this is, the residuals have a mean equal to zero. Therefore, a σ^2 = 0.01668 constant variance means that there is no longer information about the dependence of some data on others over time (Table 6).

After estimating the parameters of both models, they were validated by residual analysis (Yafee and McGee, 2000). The estimated standardized residuals of these models should behave as an independent and identically distributed sequence, with a mean equal to zero and constant variance. However, the residuals of the model without intervention still present outliers in the ± 0.5 band, a sign that this model is being affected by outliers (Figure 4A). When the intervention is included in the model, the residuals oscillate by ±0.2, substantially improving the mean and constant variance (Figure 4B). The distribution of residuals without intervention approximates a normal slightly left-skewed leptokurtic distribution (Figure 4C). The model with intervention approximates a normal distribution, which indicates a great affinity of the data, regardless of their magnitude (Figure 4D).

The ACF of the residuals of the ARIMA (1, 1, 1) model showed data that fell outside the confidence band, a sign that there is still data dependency that can be modeled (Figure 5A). The ACF of the residuals of the ARIMA (1, 1, 0) model with intervention shows that the autocorrelations fall within the confidence band (this is, they are close to zero). Consequently, the residuals did not show a significant deviation from a process of zero white noise and are random. Therefore, there is no longer information about the dependence of some data on others over time (Figure 5B).

The models estimated with and without intervention were used to make out-of-sample predictions for the seven years following the last observation and to predict the montly AAPH values for the years 1967 to 2019 with great accuracy regarding the observed values; these values are located within the confidence band (± 95 % estimate). According to the model estimation, the average prices of honey in Mexico paid to the producer in the medium term will have an upward behaviour and an average annual growth rate (AAGR) of 1.33 %. The AAPHs will range from MXN $ 46.69 to MXN $

Table 6. Verification of autocorrelation of white noise in the residuals of the AAPHL(1) series with intervention.

To lag	Chi-squared	DF	Pr > ChiSq	Autocorrelations					
6	7.65	6	0.2652	-0.195	0.082	-0.165	0.036	-0.115	0.207
12	9.81	12	0.6326	-0.116	0.036	0.075	-0.086	-0.029	-0.065
18	17.97	18	0.4576	-0.001	-0.015	0.250	-0.080	-0.064	-0.174
24	24.51	24	0.4326	0.039	-0.120	0.232	-0.037	-0.030	-0.009
30	30.07	30	0.4623	-0.036	-0.100	0.022	0.065	-0.071	0.159

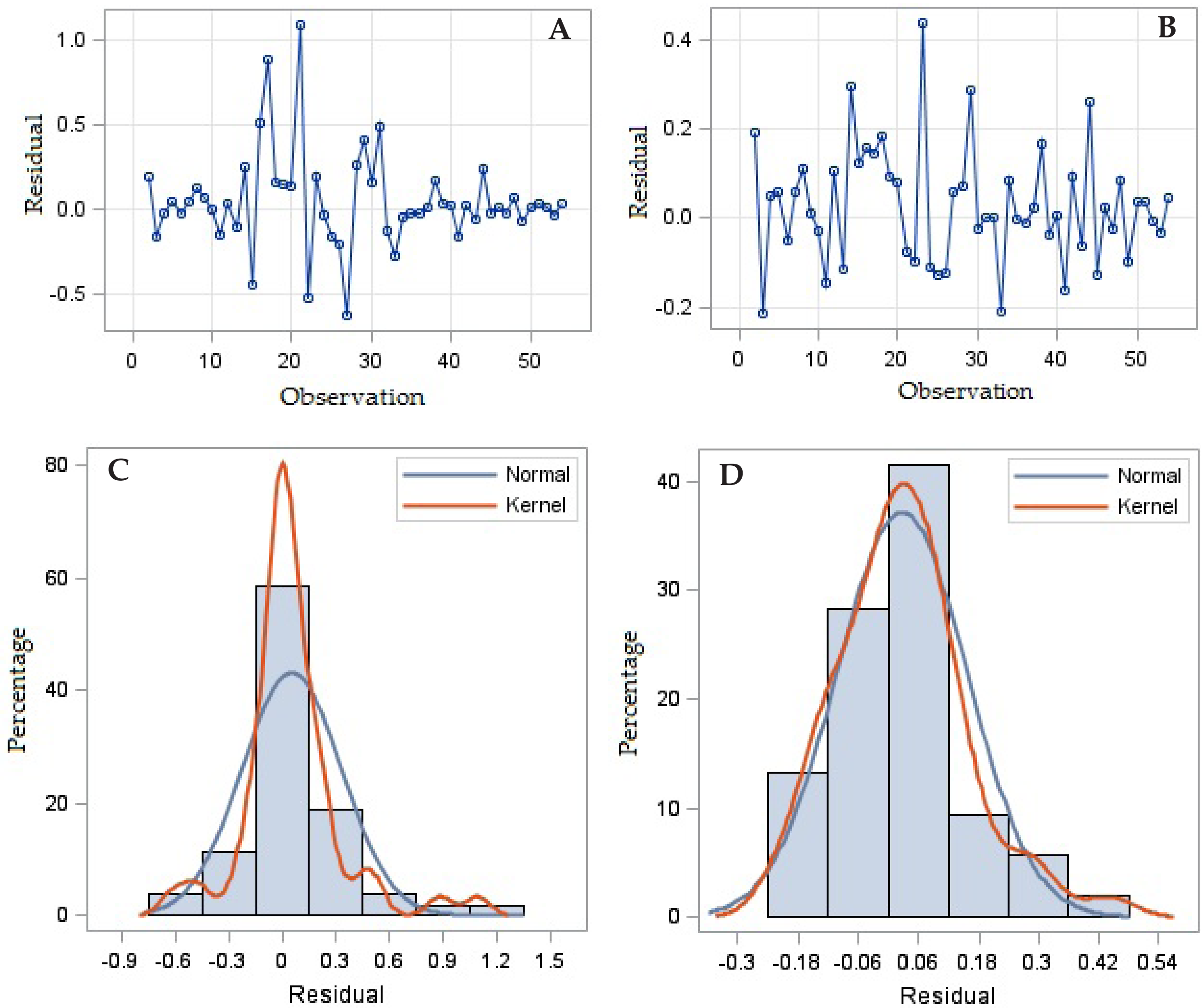

Figure 4. Graphical diagnostics used to evaluate the fit of the ARIMA (1, 1, 1) and ARIMA (1, 1, 0) models, based on the standardized residuals. A: Residuals of the model without intervention; B: Residuals of the model with intervention; C: Distribution of residuals without intervention; D: Distribution of residuals with intervention.

49.25 according to the ARIMA model without intervention. With the ARIMA model that includes the outliers, the AAPHs will fluctuate between MXN $ 47.49 and MXN $ 50.15. These results match the findings of Ramos and Pacheco (2016), who pointed out that the beekeeping sector is increasingly specialized and constantly improves the product, adding and diversifing value and, therefore, obtaining better international prices for honey. However, this implies greater incentives to marketers-exporters. In contrast, although the prices paid to producers have increased in recent years, they have not increased in the same proportion (Figure 6).

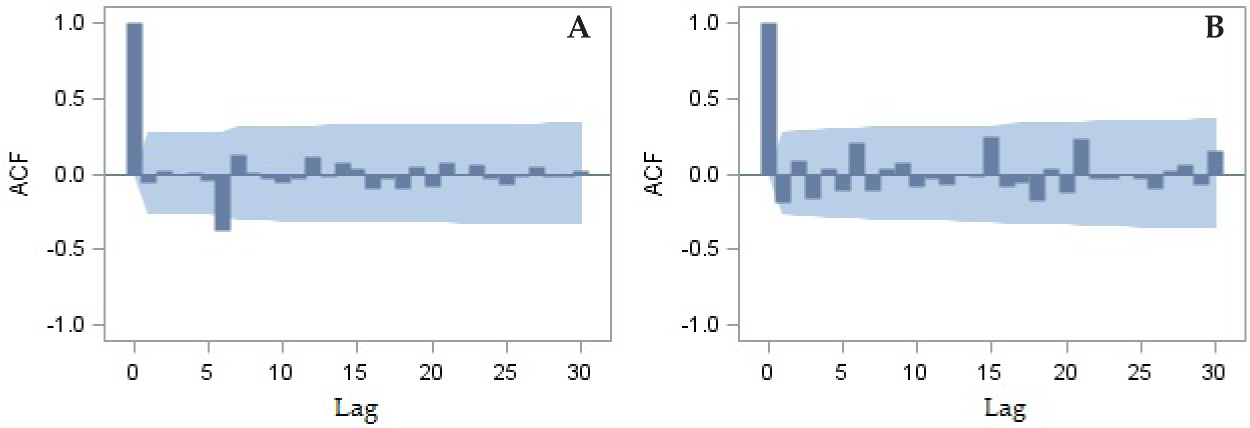

Figure 5. Graphical diagnostics used to evaluate the fit of the ARIMA (1, 1, 1) and ARIMA (1, 1, 0) models, based on the ACF of the residuals. A: Without intervention; B: With intervention.

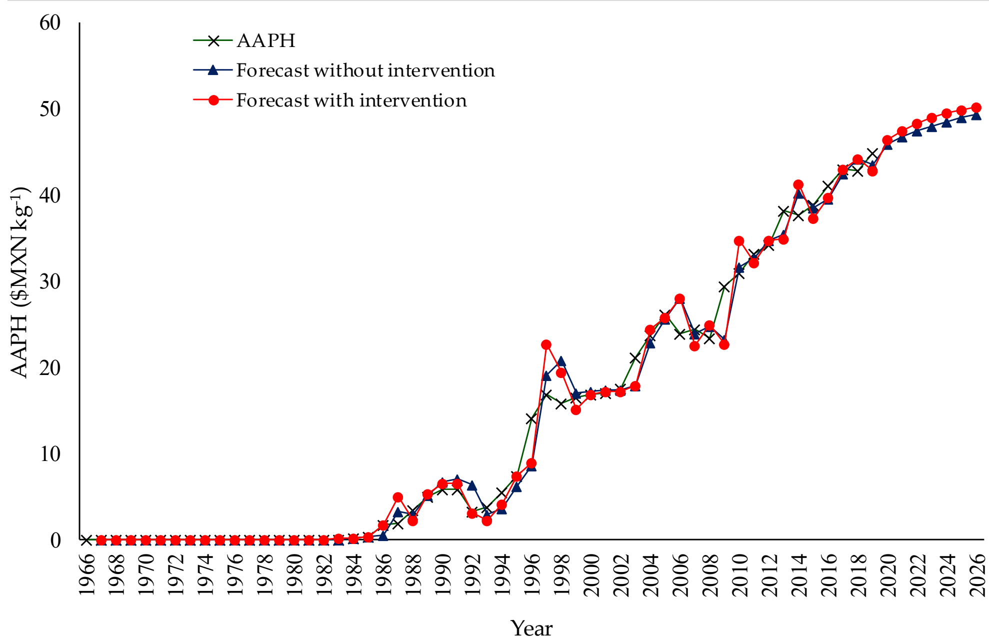

Figure 6. Average honey prices (observed and forecasted) paid to producers in Mexico, 2019 (MXN \$ kg^{-1}). They were obtained through the ARIMA (1, 1, 1) and ARIMA (1, 1, 0) models with intervention, based on information provided by SIAP (2020) and FAO (2020).

With the ARIMA models without and with intervention, the AAPHs were forecasted for the 1967-2019 period and these prices were compared with those of the AAPH series. The predictions had a mean absolute percentage error (MAPE) of 8.16 % for the model without intervention and 4.02 % for the model with intervention. This indicates that the second model, which included the outliers with special treatment that improve the statistical fit of the studied time series, improved the predictions of honey prices.

Compared to the ARIMA methodology, some research about vanilla production prediction in Mexico provided models with a 91.68 % forecast accuracy; Luis-Rojas *et al.* (2020), for example, reported an ARIMA (1, 1, 1) structure similar to the one proposed here.

Other production and price prediction models, which contemplate both ARIMA and SARIMA structures, are used to forecast white egg prices to Mexican producers, as well as the prices of vanilla and pork. Barreras-Serrano *et al.* (2014) and Luis-Rojas *et al.* (2019) point out that this methodology is only useful to establish short-term forecasts, suggesting that a greater accuracy could be achieved through the inclusion of exogenous variables through transfer function models; they also propose the use of multivariate models for long term forecasts.

Ruiz *et al.* (2019) used a SARIMA (2, 1, 0) X $(1, 1, 0)_{s=12}$ model to make a 12-month forecast of the apple price, concluding that future apple prices show an upward trend. However, the authors suggest considering the limitation of the prediction, since the economic dynamics of prices will always be complex. Nevertheless, prediction can be a useful tool for decision-making.

Finally, the estimation of a unique and universally accepted model to understand the future prices of bee honey in Mexico is unrealistic and perhaps unnecessary since the international market demands honey from different blooming fields and regions. Therefore, the statistical agencies should provide increasingly specific information. However, this ARIMA (1, 1, 0) model with intervention explains, to a large extent, the panorama of the prices paid to honey producers in Mexico. In coincidence with the structure of the studied series, it can be understood and analysed, because it is well specified.

CONCLUSIONS

The series presents five structural changes with a trend break of the AAPHs (1966–1985, 1986–1995, 1996–2003, 2004–2008, and 2009–2019), explained by the protection of the domestic market, the Africanization of hives, the presence of varroa, extreme climatic factors (hurricanes, droughts, etc.), and the appreciation of pollutant-free, organic, and transgenic-free honey in the foreign market.

The inclusion of outliers in the ARIMA model, which reduced the variance from 0.08123 to 0.01920, resulted in a better prediction of the AAPH than the univariate ARIMA model without intervention.

The short-term predictions of the study series showed an 8.16 % and 4.02 % difference from the data observed with the ARIMA model without intervention and with intervention, respectively; in both cases the random error was minimized. The two models proposed explain the AAPH in Mexico, based on the AAPH from a previous period, which showed an upward trend in the medium term. Therefore, the demand displacement, based on the tastes and preferences of the consumer, explains the increase in prices, although the supply increase has not been a major factor, because the quantity produced has practically been very similar in the last 20 years.

ACKNOWLEDGMENTS

This research was included in the Línea de Generación y Aplicación del Conocimiento (LGAC) - Economía Agrícola y Política Pública of the Posgrado en Socioeconomía, Estadística e Informática-Economía, Colegio de Postgraduados. The authors would like to thank the Consejo Nacional de Ciencia y Tecnología (CONACYT) of Mexico for the Sc.D. scholarship granted to Samuel Luis-Rojas. This research was funded with the resources that the Colegio de Postgraduados Campus Montecillo assigns to the first author as part of his Sc.D. studies

REFERENCES

Abdulhafedh A. 2017. How to detect and remove temporal autocorrelation in vehicular crash data. Journal of Transportation Technology 7 (2): 133−147. https://doi.org/10.4236/jtts.2017.72010

Barreras-Serrano A, Sánchez-López E, Figueroa-Saavedra F, Olivas Valdez JA, Pérez-Linares C. 2014. Uso de un modelo univariado de series de tiempo para la predicción, en el corto plazo, del comportamiento de la producción de carne de bovino en Baja California, México. Veterinaria México 45 (spe): 1–9.

Box, GE, Jenkins GM, Reinsel GC, Ljung GM. 2015. Time series Analysis: Forecasting and Control. 5th ed. John Wiley & Sons Inc: Hoboken, NJ, USA. 712 p.

Caro-Encalada MJ, Leyva-Morales CE, Chi-Tuz M. 2012. Comercialización de miel en el municipio de Espita, Yucatán. Revista de Economía 29 (78): 9−35. https://doi.org/10.33937/reveco.2012.29

CEDRSSA (Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria). 2018. La dependencia alimentaria de México. http://www.cedrssa.gob.mx/post_la_dependencia_-n-alimentaria-n-_de_-n-mn-xico-n.htm (Retrieved: February 2021).

Cervantes-Sánchez JM, Román de Carlos AM, Velázquez-Camacho BL. 2018. Evolución de la producción apícola en México (1521-2010). *In* Avances de la investigación sobre producción animal y seguridad alimentaria en México. Herrera CJ, Chay CAJ, Casanova LF, Piñeiro VAT, Marquez BL, Santillan FE, Arce MJ (eds.). Universidad Michoacana de San Nicolás de Hidalgo: Morelia, Michoacán, México. pp: 782−786.

Dickey AD and Fuller WA. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica Journal of the Econometric Society 49 (4): 1057−1072. https://doi.org/10.2307/1912517

Dolores-Mijangos G, Santiago-Cruz MJ, Arana-Coronado JJ, Utrera-Quintana F. 2017. Estudio del impacto de la actividad apícola en el Istmo de Tehuantepec, Oaxaca, México. Agricultura Sociedad y Desarrollo 14 (2): 187−203. https://doi.org/10.22231/asyd.v14i2.588

FAO (Organización de las Naciones Unidas para la Agricultura y la Alimentación). 2020. Precios al productor. Roma, Italia. https://www.fao.org/faostat/es/#data/PP (Retrieved: February 2021).

Ferruz-Agudo L, Marco-Sanjuan I, Knebel-Baggio D. 2011. Análisis de intervención de las series temporales patrimonio y flujo neto de dinero de los Fondos de Inversión Socialmente Responsables (FISR) de Brasil. Contabilidad y Negocios 6 (12): 26−35.

Luis-Rojas S, Ramírez-Valverde B, Díaz-Bautista M, Pizano-Calderón J, Rodríguez-López C. 2020. La producción de vainilla (*Vanilla planifolia*) en México: análisis y pronóstico. Revista Mexicana de Ciencias Agrícolas 11 (1): 175−187. https://doi.org/10.29312/remexca.v11i1.2065

Luis-Rojas S, García-Sánchez RC, García-Mata R, Arana-Coronado OA, González-Estrada A. 2019. Metodología Box - Jenkins para pronosticar los precios de huevo blanco pagados al productor en México. Agrociencia 53 (6): 911−925.

Magaña-Magaña MA, Tavera-Cortés ME, Salazar-Barrientos LL, Sanginés-García JR. 2016. Productividad de la apicultura en México y su impacto sobre la rentabilidad. Revista Mexicana de Ciencias Agrícolas 7 (5): 1103−1115.

Martínez-González EG, Pérez-López H. 2013. La producción de miel en el trópico húmedo de México: avances y retos en la gestión de la innovación. UACH–CIESTAAM. Texcoco, Estado de México, México. 102 p. https://doi.org/10.13140/RG.2.1.2595.6325

Maté V. 2012. China invade el mercado con miel a bajo precio. Nota periodística. El País. Economía. Madrid, España. 8 julio 2012. https://elpais.com/economia/2012/07/08/actualidad/1341776471_952938.html (Retrieved: February 2021).

Medina-Flores CA, Guzmán-Novoa E, Hamiduzzaman MM, Aréchiga-Flores CF, López-Carlos MA. 2014. Africanized honey bees (*Apis mellifera*) have low infestation levels of the mite *Varroa destructor* in different ecological regions in Mexico. Genetic and Molecular Research 13 (3): 7282−7293. http://dx.doi.org/10.4238/2014.February.21.10

Moffat IU and Akapan EA. 2019. White noise Analysis: A Measure of Time series Model Adequacy. Applied Mathematics. 10 (11): 989−1003. https://doi.org/10.4236/am.2019.1011069

Montemayor-Gallegos JE. 2013. Métodos de pronóstico para negocios. Monterrey, Nuevo León, México. 260 p. http://prod77ms.itesm.mx/podcast/EDTM/P196.pdf (Retrieved: November 2020).

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria. https://www.R-project.org/ (Retrieved: December 2019).

Ramos-Díaz AL, Pacheco-López NA. 2016. Producción y comercialización de miel y sus derivados en México: Desafíos y oportunidades para la exportación. Editorial CIATEJ-CONACYT: Mérida, Yucatan, México. 196 p.

Rodríguez-Castro JH, Correa-Sandoval A, Ramírez-de-León JA. Adame-Garza JA. 2017. Modelización de la captura y fases de desarrollo de la pesquería de la jaiba azul (*Callinectes sapidus*) en la Laguna Madre, Tamaulipas, México. Ciencia UAT 12 (1): 96-113.

Ruiz-Hernández JA, Barrios-Puente G, Gómez-Gómez AA. 2019. Análisis del precio de la manzana mediante un modelo SARIMA. Revista Mexicana de Ciencias Agrícolas 10 (2): 225−237. https://doi.org/10.29312/remexca.v10i2.509

SAS Institute Inc. 2014. SAS/ETS® 13.2 User's Guide the ARIMA Procedure. SAS Institute Inc. Cary, NC. USA. 124 p.

Segura EO y Torres V. 2014. Tratamiento de valores perdidos y atípicos en la aplicación del Modelo Estadístico de Medición de Impacto en un estudio de 90 fincas lecheras en la provincia de Pastaza, Ecuador. Revista Cubana de Ciencia Agrícola 48 (4): 333−336.

SIAP (Sistema de Información Agroalimentaria y Consulta). 2019. Panorama agroalimentario 2019. Servicio de Información Agroalimentaria y Pesquera. Ciudad de México, México. 214 p.

SIAP (Sistema de Información Agroalimentaria y Consulta). 2020. Miel avance mensual de la producción pecuaria. http://infosiap.siap.gob.mx/repoAvance_siap_gb/pecAvanceProd.jsp. (Retrieved: February 2021).

Varian H. 2010. Microeconomía Intermedia. Un enfoque actual. 8a Edición. Editorial Antoni Bosch. Barcelona, España. 864 p.

Vélez JI, Correa JC, Marmolejo-Ramos F. 2015. A new approach to the Box-Cox transformation. Frontiers in Applied Mathematics and Statistics 1(12): 1−10. http://dx.doi.org/10.3389/fams.2015.00012

Yafee RA, McGee M. 2000. Introduction to Time series Analysis and Forecasting with applications of SAS and SPSS (ed.). Academic Press: Brooklyn, NY, USA. pp: 83−89.

Zeileis A, Leisch F, Hornik K, Kleiber C. 2019. Strucchange: An R Package for Testing for Structural Change in Linear Regression Models. http://www.jstatsoft.org/v07/i02/. (Retrieved: December 2019).