The Journal of Animal & Plant Sciences, 27(3): 2017, Page: 920-928
ISSN: 1018-7081

FORECASTING THE PRODUCTION OF GROUNDNUT IN TURKEY USING ARIMA MODEL

S. Celik[1], K. Karadas[2] and E. Eyduran[*3]

[1]Bingol University, Agricultural Faculty, Department of Animal Science, Bingol, Turkey
[2]Igdir University, Agricultural Faculty, Department of Agricultural Economics, Igdir, Turkey
[3]Igdir University, Agricultural Faculty, Department of Animal Science, Igdir, Turkey
[*]Corresponding Author's email address: ecevit.eyduran@gmail.com, ecevit.eyduran@igdir.edu.tr

ABSTRACT

The aim of this study was to model groundnut production in Turkey using data from the period 1950- 2015 in Turkey in an effort to forecast groundnut production amounts between the years 2016 and 2030 by using Autoregressive Integrated Moving Average (ARIMA) models. In the study, ARIMA (0,1,1) was found to be the most appropriate model among six studied ARIMA models in forecasting the groundnut production amounts for the next 15 years. We forecasted that annual amount of groundnut production obtained in the year 2016 was 138,98 thousand tons and it reached to 167,28 thousand tons in the year 2030 with a significant acceleration for groundnut production. Forecasting results of the ARIMA (0,1,1) illustrated an increasing trend in the amount of groundnut production, and they might help to determine a better policy for increasing groundnut production in Turkey.

Key words: ARIMA, Groundnut Production Forecasting, Time Series Models, Autocorrelation Function, Partial Autocorrelation Function, Production Planning and Control.

INTRODUCTION

Groundnut, which is beneficial in human and animal nutrition, is an important product since it is used in the production of many foods and ranked at 4th among oil-seed plants after cotton, soy bean and rape. Groundnut also has a particular economic value since its oil, kernels, shell and straw can be used commercially. Production and yield amounts of the groundnut for Turkey and the world based on harvested area, according to FAOSTAT database records, are shown in Table 1. Although the groundnut yield in Turkey is more than twice that of the World average, there is little awareness among producers about this comparative advantage created by high yield of Turkish groundnut production. The contribution of groundnut production to the gross domestic product (GDP) of Turkey is rather minimal. Given the increasing need for more protein resources due to increase in human population, groundnut production in Turkey should be increased in order to meet the growing demand. On the other hand, increasing demand for human consumption decreases the available quantity of groundnut for oil industry, which creates contraction in the capacity of agriculture-based industry (Pal and Mazumdar, 2015). Increasing production would also be an important tool in development of rural areas by increasing producer revenues. Thus, it is important for Turkey to formulate policies aiming at increasing groundnut production for future sustainability of the oil industry, export revenues and food safety. In this regard, projection studies are inevitable about forecasting groundnut amounts and determining proper policies in the next years.

Pal and Mazumdar (2015) employed monomolecular and logistic non-linear growth models for forecasting groundnut production during the period 1950-1951 and 2011-2012. However, more attention on time series models has been drawn for economically forecasting some agricultural productions instead of non-linear functions in recent times (Borkar, 2016). Among various time series models, the ARIMA model has been widely implemented to forecast domestic consumption and exports as a part of agricultural production (Amin *et al.,* 2014). For example, Celik (2015) applied sixteen ARIMA forecasting models for time series data in order to forecast annual amounts of honey production from the years 1950 and 2015 in Turkey. Celik (2013) used ARIMA forecasting models, which are described as a hybrid of autoregressive (AR) and moving average models, in an attempt to forecast production amounts of some nuts (walnut, chestnut, almond, pistachios and hazelnuts). Amin *et al.* (2014) applied ARIMA (1,2,2) forecasting model in time series modeling of wheat production for the period 1902-2005 in Pakistan. Ozer and Ilkdogan, (2013) forecasted the world cotton price through ARIMA model, also recognized as Box Jenkins model. Borkar (2016) found the ARIMA (0,1,1) model as the best forecasting model for the groundnut production over a time period between 1950-1951 and 2013-2014 in India. Despite ARIMA forecasting models are used to estimate production in several agricultural products, time series modeling has not yet been applied to the projection of groundnut production in Turkey. The main aim of this study was to model the groundnut production in Turkey through Autoregressive Integrated Moving Average Models (ARIMA) for the period 1950 to 2015 in order to forecast the ground production amounts for the period 2016 to 2030. This study is important for performing groundnut production planning in the next 15 years.

Table 1. The long term groundnut statistics (Average).

		Area Harvested (ha)	Production (thousand tons)	Yield(HG/HA)
1961-1971	World	17373490	150819,25	8681
	Turkey	12075	29,73	24618
1972-1982	World	17727820	166907,42	9415
	Turkey	20832	48,39	23227
1983-1993	World	15994560	184273,31	11521
	Turkey	24907	59,72	23976
1994-2004	World	20922500	299735,59	14326
	Turkey	30027	79,27	26523
2005-2014	World	24314700	396111,55	16291
	Turkey	28424	98,66	34710

MATERIALS AND METHODS

Data on annual groundnut production of the period 1950-2015 under investigation were provided by “Statistical Indicators” book published by TUIK 2014. Statistical data of the subsection “cereals and other herbal products/oil seeds” of Agricultural Statistics were also obtained from TUIK database (TUIK, 2015). All units of the production were in thousand tons.

Unit Root test: Level of stability series expressed by Dickey and Fuller (1981) is defined by the Augmented Dickey Fuller (ADF) unit root test (Seddighi *et al.*, 2000) (Eq. 1).

$$\Delta X_t = \beta_0 + \beta_1 t + \gamma_1 X_{t-1} + \sum_{i=1}^{h} \gamma_i \Delta X_{t-i} + e_t$$

(Eq.1)

Perron (1989) argued that inference drawn from the Dickey-Fuller unit root tests may be misleading if the underlying model ignores a break in the mean or trend of the time series that may result from some major events. In order to find out appropriate ARIMA (p,d,q) model, autocorrelation and partial autocorrelation functions of time series were evaluated and significance of parameter estimates was tested statistically. Parameters p and q denote the order of AR and MA, whereas d shows the degree of differencing the time series in order to provide stationarity (Borkar, 2016). In the selection of the appropriate ARIMA model, AIC (Akaike Information Criterion) and BIC (Schwartz’s Bayesian information criterion) were used. The model that produced the best results among ARIMA models tested in the study was selected as a forecasting model and presented the estimates made for next 15 years.

Time series models consist of autoregressive, moving average, autoregressive moving average models. Among these models, an autoregressive AR (p) model can be defined in Eq.2 (Wei, 2006).

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + e_t$$

(Eq.2)

A moving average MA (q) model is then expressed as Eq.3 (Cooray, 2008).

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q}$$

(Eq.3)

Autoregressive moving average model, ARMA (p,q), comprising AR (p) and MA(q) components (Cryer, 1986; Sevuktekin and Nargelecekenler, 2010; Enders, 2010) can be written as in Eq.4.

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$$

(Eq.4)

If X_t is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then X_t is stationary with the same covariance function as the independent identical distribution white noise. This is indicated by the notation $X_t \sim WN(0,\sigma^2)$ (Brockwell and Davis, 1996). Non-stationary time series is made stationary after the first difference (Box and Jenkins, 1976). Generally, ARIMA (p,d,q) model is expressed as Eq.4 (Kadilar, 2009).

The autocorrelation function of a time series (ACF) is described as Eq.5 (Shumway and Stoffer, 2006).

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

(Eq.5)

Here, $\gamma(h) = \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) / n$ and

$\gamma(0) = \sum_{t=1}^{n} (X_t - \bar{X})^2 / n$ are variance

h^{th} partial autocorrelation (PACF) is expressed as Eq.6 (Wei, 2006)

$$P_h = \frac{\gamma(h) - \alpha_1\gamma(h-1) - \alpha_2\gamma(h-2) - ... - \alpha_{h-1}\gamma(1)}{\gamma(0) - \alpha_1\gamma(1) - \alpha_2\gamma(2) - ... - \alpha_{h-1}\gamma(h-1)}$$

$$= \frac{\rho(h) - \alpha_1\rho(h-1) - \alpha_2\rho(h-2) - ... - \alpha_{h-1}\rho(1)}{1 - \alpha_1\rho(1) - \alpha_2\rho(2) - ... - \alpha_{h-1}\rho(h-1)} \quad \text{(Eq.6)}$$

Selection criteria such as AIC and BIC used in the description of the most suitable model are formulated in Eq.7 and Eq.8, respectively as also mentioned by some authors (Wei, 2006; Cooray, 2008)

$$AIC = n\ln\hat{\sigma}_e^2 + 2M \quad \text{(Eq.7)}$$

$$BIC = n\ln\hat{\sigma}_e^2 + M\ln n \quad \text{(Eq.8)}$$

Where, M is parameter number of the model and expressed as M=p+q+1.

To reveal the most appropriate one among the evaluated ARIMA models, several goodness of fit criteria described by Makridakis *et al.* (2003), Takma *et al.* (2012), Grzesiak and Zaborski (2012) and Ali *et al.* (2015) were used. The goodness of fit criteria can be formulated in the following Equations (9) - (12):

Root Mean Square Error, $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}} \quad \text{(Eq.9)}$$

Mean Absolute Percentage Error, $$MAPE = \frac{\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}}{n} \quad \text{(Eq.10)}$$

Maximum absolute percentage error, $$MaxAPE = \max_i\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right)*100, \quad i=1,2,\ldots,N \quad \text{(Eq.11)}$$

Mean absolute error (MAE), $$MAE = \frac{1}{n}\sum_{i=1}^{M}\left|y_i - \hat{y}_i\right| \quad \text{(Eq.12)}$$

Mean percentage error (MAPE) is expressed as a measure of how much a dependent series changes from its model-predictive performance level. The root mean square error (RMSE) has been used as a goodness of fit criterion to measure model performance in various research studies. Maximum Absolute Percentage Error (MaxAPE) represents the largest forecasted error, expressed as a percentage. Statistical analysis of the groundnut production data was carried out by using IBM SPSS program (version 23).

RESULTSANDDISCUSSION

Groundnut occupies an important place in Turkish economy as an important oil plant. Figure 1 depicts the graph of annual production amount of groundnut for the period 1950-2015. An increasing trend in annual groundnut production can be seen in the graph. In order to have a good understanding of the stationarity of the time series, ACF and PACF graphs are depicted in Figures 2 and 3, respectively. The ACF graph in Figure 2 confirms the non-stationarity of time series because of significant ACF values for large number of lags. In order to make the data stationary, the first differences of the time series was taken and ACF and PACF of the first difference series are plotted in Figures 4 and 5, respectively. It was obvious that the relationship value of the first lag in ACF graph of the first degree difference series in Figure 4 surpassed confidence limits. However, other lags were within confidence limits. This means that the stationarity in time series was obtained since the trend was removed for the first degree difference series. Next, the generalized Dickey-Fuller (ADF) unit root test, which is a formal test for stationarity developed by Dickey and Fuller(1981), was applied to the time series and the related results are reported in Table 2. The non-significant value of ADF test statistic (1.102) for level and significant value (-12.777) for the first differences confirmed that the first difference time series is stationary. The ideal model was determined by jointly examining ACF and PACF graphs of the first difference time series under the current time series study. After the first lag in ACF graph, magnitude of relationships decreased rapidly and thereafter was near-zero. The first lag obtained for PACF graph in Figure 5 was significant, and magnitude of subsequent values decreased slower. This means that moving average model was the ideal model for the first difference time series. In the ACF graph, q=1 because the relationship concerning the first lag was found to be significant. For taking the first difference of the time series, d=1 and p=0 were specified. In brief, the most suitable model for first difference time-series was determined as ARIMA (0,1,1), also called the first degree integrated moving average model, under the time series investigation.

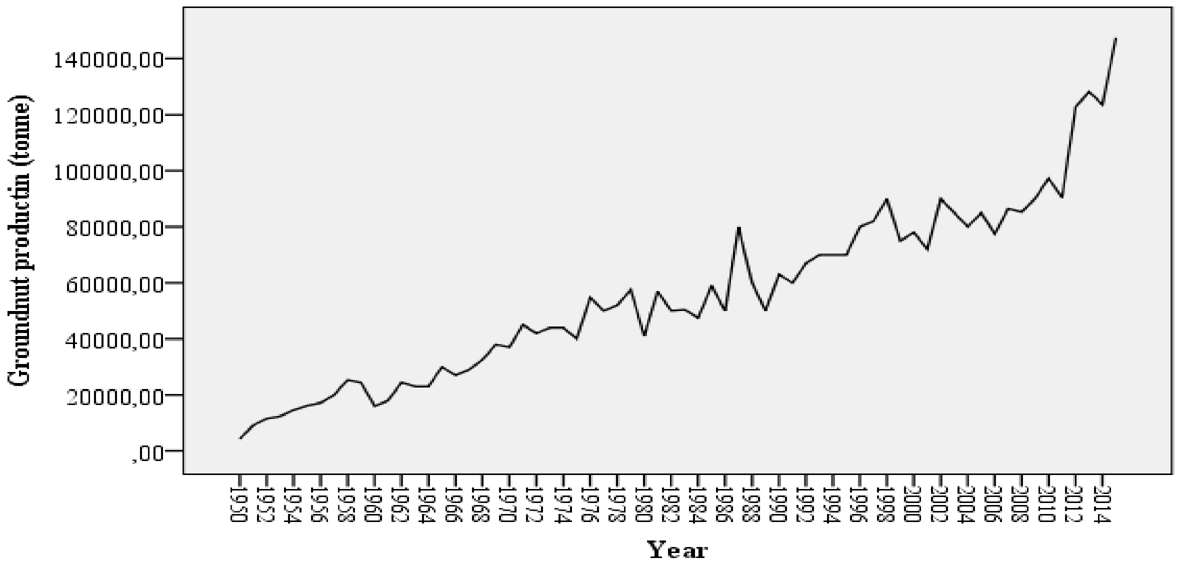

Figure 1. Annual production of the groundnut (tons)

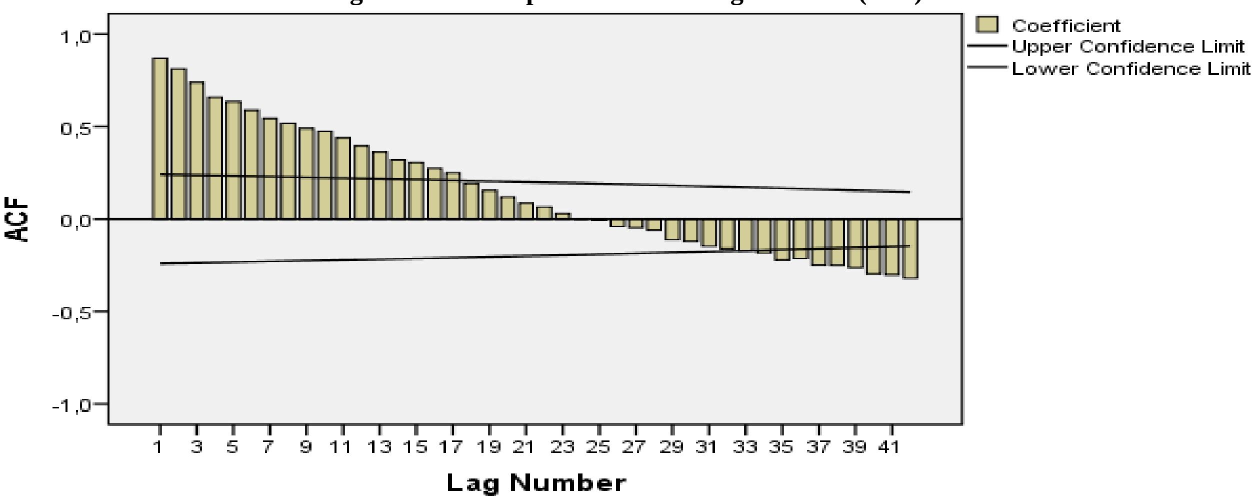

Figure 2. Graph of Autocorrelation function (ACF)

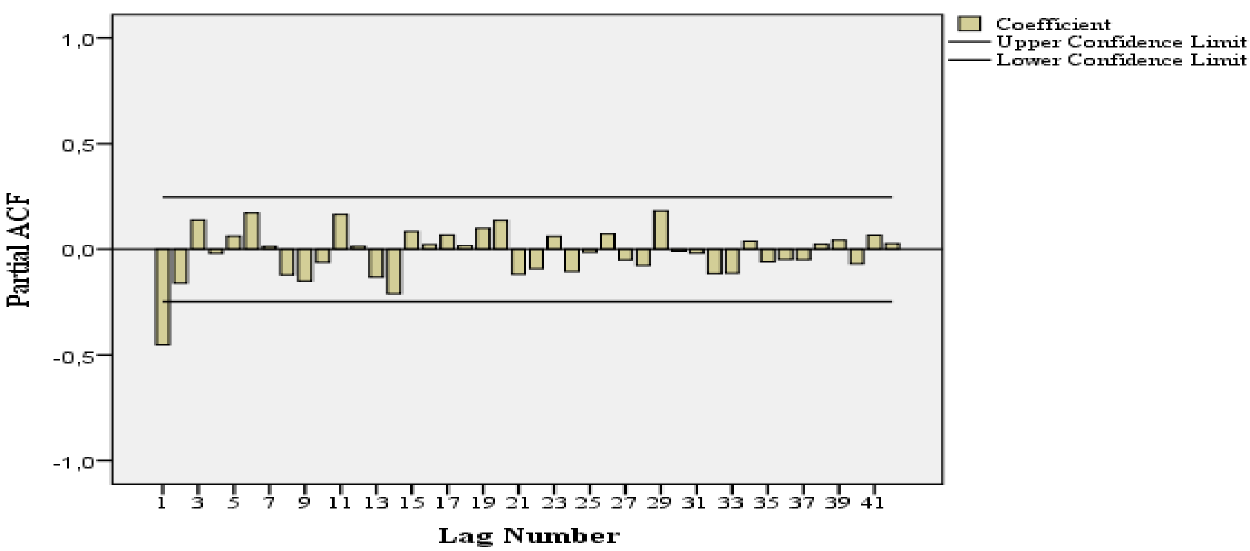

Figure 3. Graph of Partial Autocorrelation function (PACF)

Table 2. Augmented Dickey-Fuller Unit Root Tests for Groundnut Production Series.

		Level	First difference
Augmented Dickey-Fuller test statistic		1.102	-12.777
Test critical values:	1% level	-3.537	-3.537
	5% level	-2.908	-2.908
	10% level	-2.591	-2.591
	Prob. *	0.997	0.001

*MacKinnon (1996) one-sided p-values.

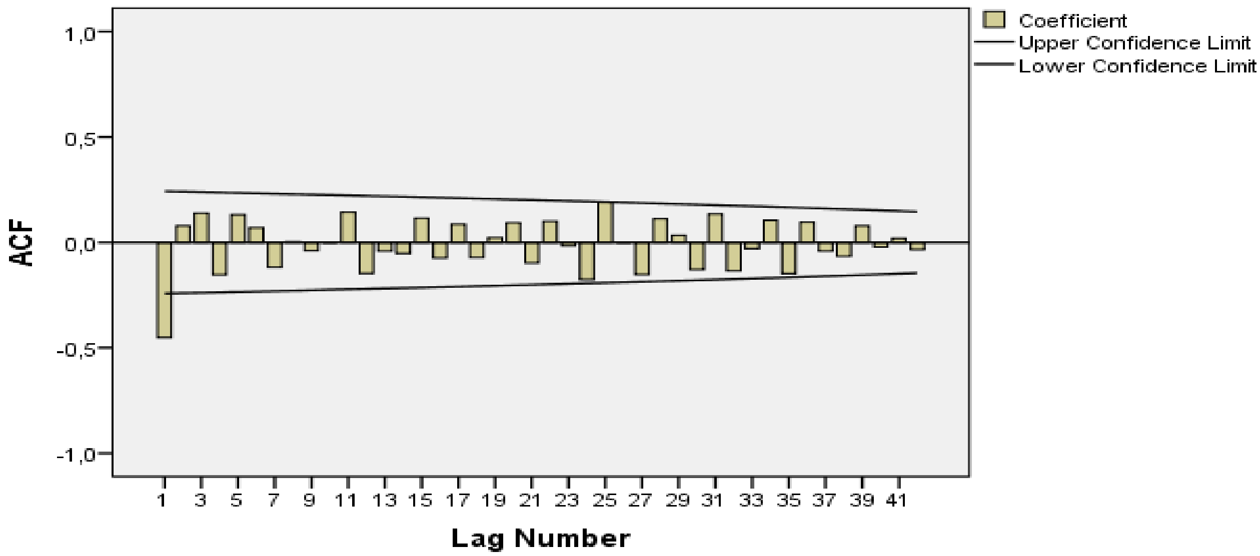

Figure 4. Autocorrelation graph of the first difference series (ACF)

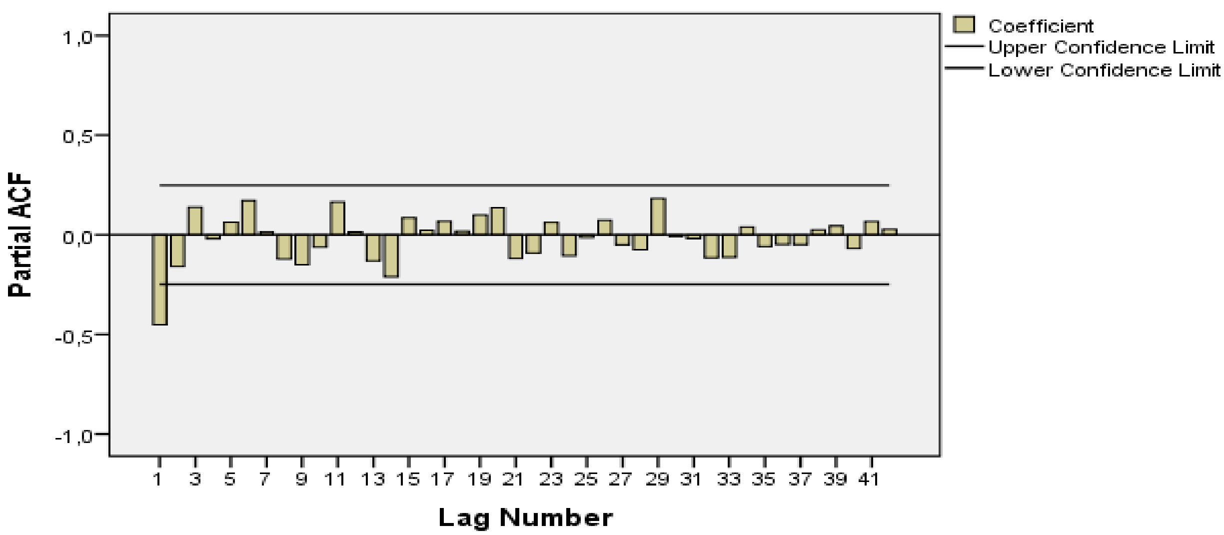

Figure 5. Partial Autocorrelation Graph of the first difference time series (PACF)

In addition, the appropriateness of the tested model was investigated by using various goodness of fit criteria. The results taken from various criteria are reported in Table 3, and also demonstrated that ARIMA (0,1,1) was found to be the most suitable model among the candidate ARIMA models. Mean percentage error (MAPE) for the best model is found as 11.090 per cent. The values of RMSE=8307.933, MaxAPE=69.212, MAE=5713.189 and MaxAE=27484.589 were recorded for the model. Besides, the Ljung-Box statistics value of

14.253 is found significant at 5% level of significance. Parameter estimates in Table 4 were also found significant for the ARIMA (0,1,1) model. Here, the parameter for ARIMA (0,1,1) model was estimated as $\theta=$0.472. By using Back-Shift operator, the ARIMA (0,1,1) model can be expressed as follows:

$$X_t = X_{t-1} - 0.472e_{t-1} + e_t$$

Residual graph of the selected model is depicted in Figure 6. The first difference series suitable for the model was white noises since all residuals were placed within intervals described in the time series analysis. From Figure 7, it was understood clearly that the original series from the observed values in the time series graph were consistent with the series from the forecasting values. Forecasting values obtained for the year 2016 are given in Figure 7.

In order to choose the ideal model for the groundnut data, the probable values between 0 and 2 were tested as p and q values for ARIMA model with d=1. Performance results of candidate ARIMA models tested in the study are recorded in Table 5. Both the ARIMA (1,1,0) and ARIMA (0,1,1) model have produced accurate results, however, since the ARIMA (0,1,1) model had the lowest AIC and BIC values, it was found to be the best one among the candidate models. The groundnut production (in tons) for the period 2016 – 2030 were forecasted from the ARIMA (0,1,1) model and the forecasting values are presented in Table 6.

Table 3. Model fit statistics of the fitted ARIMA(0,1,1) model.

Fit Statistic	Stationary R-squared	R-squared	RMSE	MAPE	MaxAPE	MAE	MaxAE	BIC
Mean	0.218	0.930	8307.93	11.09	69.212	5713.19	27484.6	20.984

Table 4. Model parameters.

Model Parameters	Estimate	S.E.	t	Sig. (p)
MA(1)	0.472	0.126	3.757	0.001

S.E.: Standard Error

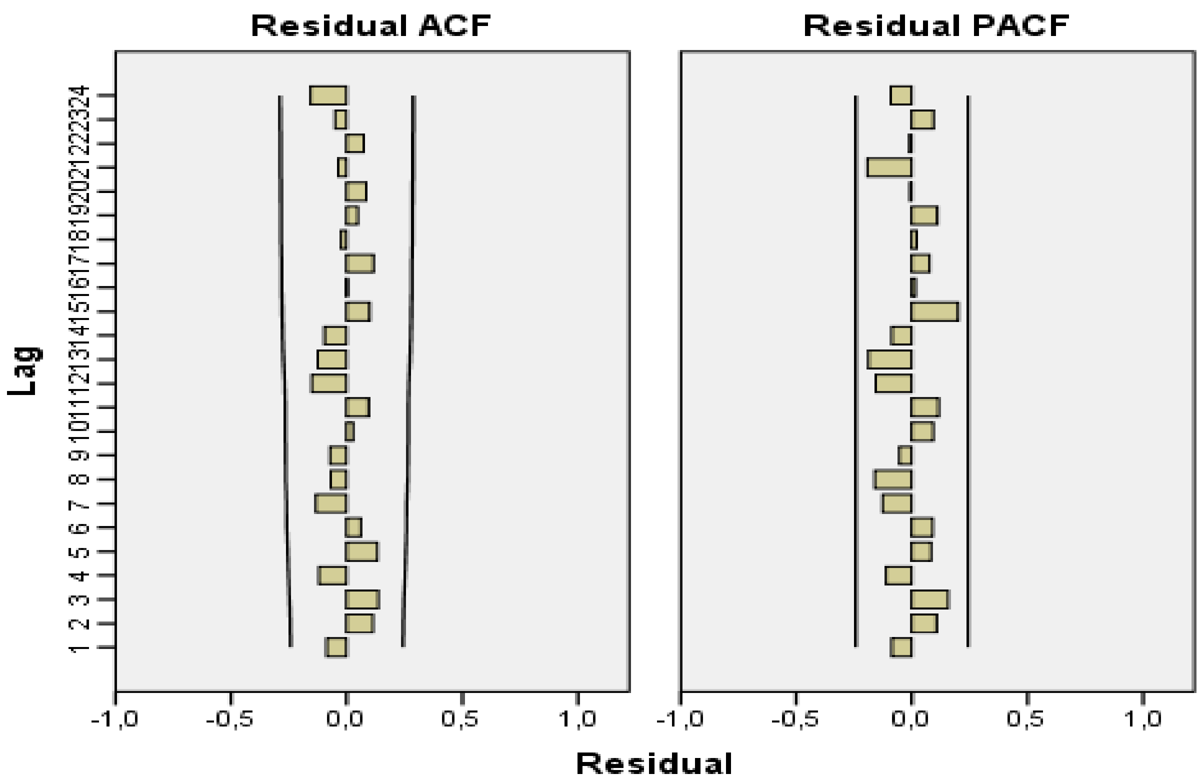

Figure 6. ACF and PACG graphs of the residual series

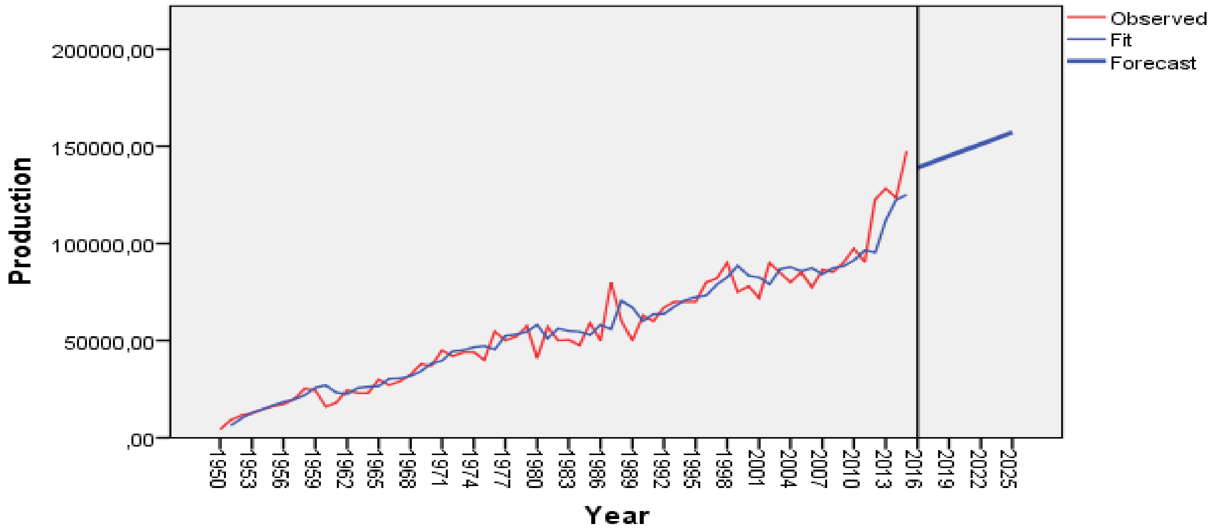

Figure 7. The graph of the observed and forecasted values in the first different time series

Table 5. Results of the goodness of fit criteria of candidate ARIMA models.

Model	AIC	BIC	Parameters and p (significant)
ARIMA(0,1,1)	20.917	20.984	$\theta = 0.472^{**}$
ARIMA(0,1,2)	20.912	21.012	$\theta_1 = 0.567 **, \theta_2 = -0.195$
ARIMA(1,1,0)	20.925	20.993	$\varphi = -0.495^{**}$
ARIMA(1,1,1)	20.934	21.036	$\varphi = -0.316, \theta = 0.247$
ARIMA(1,1,2)	20.953	21.089	$\varphi = 0.666, \theta_1 = 1.208 **, \ \theta_2 = -0.512 *$
ARIMA(2,1,0)	20.940	21.042	$\varphi_1 = -0.584 **, \varphi_2 = -0.188,$
ARIMA(2,1,1)	20.951	21.087	$\varphi_1 = -1.079 **, \varphi_2 = -0.448 **, \ \theta = -0.509$
ARIMA(2,1,2)	20.972	21.142	$\varphi_1 = -1.076 **, \varphi_2 = -0.631 **, \theta_1 = -0.559, \theta_2 = -0.308$

* (P<0.05), ** (P<0.01)

Table 6. Forecasting groundnut production values between the years 2016-2030 (tons).

Year	Production (thousand tons)	Lower limit (thousand tons)	Upper limit (thousand tons)
2016	138,98	122,38	155,58
2017	141,00	122,23	159,77
2018	143,02	122,30	163,74
2019	145,04	122,55	167,54
2020	147,06	122,92	171,21
2021	149,09	123,40	174,77
2022	151,12	123,96	178,25
2023	153,13	124,61	181,65
2024	155,15	125,31	184,99
2025	157,17	126,07	188,27
2026	167,28	130,51	204,04
2027	161,21	127,73	194,69
2028	163,23	128,62	197,84
2029	165,26	129,55	200,96
2030	167,28	130,51	204,04

An increasing trend for groundnut production was determined after the year 2016 as an outcome of ARIMA (0,1,1) forecasting model specified for the first difference series. This trend is essential for continuing domestic consumption and exports in the production of groundnut in Turkey. In literature, number of time series modeling studies is still not inadequate. Similarly, the ARIMA (0,1,1) having the high forecasting ability under the study was also supported by Borkar (2016) who used ARIMA (0,1,1) forecasting model with similar fitting criteria as the best model for groundnut production over a time period of 1950-1951 and 2013-2014 in India. Likewise, Celik (2015) also reported that ARIMA (0,1,1) was found as the best one among sixteen ARIMA models for time series data on forecast annual amounts of honey production from the period 1950-2015 in Turkey. However, the suitability of the present model in the study was in disagreement with the ARIMA (1,2,2) model found as the best forecasting model on the basis of goodness of fit criteria like AIC, MAE, ME, MAPE, and RMSE by Amin *et al.* (2014) for forecasting annual wheat production in subsequent years of Pakistan.

When time series data of the groundnut production from the period 1950-2014 was specified instead of the years 1950-2015, ARIMA (0,1,1) model forecasted 144,93 thousand tons with a small deviance of approximately 1.64% in the production amount of the year 2015 whose real value is 147,35 thousand tons (data not shown).

Conclusion: In time series modeling of annual groundnut production amounts from the period of 1950-2015, the non-stationary time series were converted into stationary time series after taking the first difference of the data. Among candidate ARIMA models tested in the study, ARIMA (0,1,1) was found to be the best forecasting model, and forecasted up to the year 2030. According to forecasting results, annual amounts of the groundnut production in the years 2016 and 2030 were expected to be 138,98 thousand tons and167,28 thousand tons, which means a significant acceleration for the groundnut production of Turkey. Time series models are useful tools for forecasting production amounts and for effectively making agricultural planning in the next years. The groundnut production policy in Turkey should be planned according to accurate forecasts in order to increase production in upcoming years. Awareness of groundnut producers should be raised in the following years with appropriate extension programs.

REFERENCES

Ali M., E. Eyduran, M.M. Tariq, C. Tirink, F. Abbas, M.A. Bajwa, M.H. Baloch, A.H. Nizamani, A. Waheed, M.A. Awan, S.H. Shah, Z. Ahmad and S. Jan (2015). Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in Harnai sheep. Pakistan J. Zool.47:1579-1585.

Amin M., M. Amanullah and A. Akbar (2014). Time series modeling for forecasting wheat production of Pakistan. The J. Anim. Plant Sciences. 24(5):1444-1451.

Borkar, P. (2016). Modeling of groundnut production in India using ARIMA Model. International J. Research in IT and Management. 6(3):36-44.

Brockwell, P. J. and R. A. Davis (1996). Introduction time series and forecasting. Springer, New York.

Celik, S. (2013). Modelling of production amount of nuts fruit by using Box-Jenkins technique. Yuzuncu Yil J. Agr. Sci. 23(1):18-30.

Celik, S. (2015).Modelling of honey production by using time series in Turkey. SAU Fen Bil. Der. 19 (2):377-382.

Cooray, T. M. J.A. (2008). Applied time series analysis and forecasting. Narosa Publishing House Pvt. Ltd.

Cryer, J. D. (1986). Time series analysis, Duxbury Press, Belmont, pp. 269.

Dickey, D. A. and W.A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica. 49(4): 1057-1072.

Enders, W. (2010). Applied econometric time series. John Wiley and Sons Inc., USA.

FAOSTAT. (2013). Production-Crops (2013) data. Food and Agriculture Organization of the United Nations.http://faostat3.fao.org/browse/Q/QC/E

Grzesiak, W. and D. Zaborski(2012).Examples of the use of data mining methods in animal breeding.

Kadilar, C., M. Simsek and C.H. Aladag (2009). Forecasting the exchange rate series with ANN: The case of Turkey. Istanbul U. Econ. Statist. J. 9(1): 17-29.

Makridakis, S., S.C. Wheel wright and R.J. Hyndman (2003). Forecasting: Methods and applications. 3 rd Ed. John Wiley and Sons.

Pal, S. and D. Mazumdar (2015). Forecasting groundnut production of India using nonlinear growth models. J. Crop and Weed.11 (Special Issue):67-70.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. Econometrica. 57: 1361–1401.

Seddighi, H.,K. Lawlerand, and A. Katos (2000). Econometrics: A Practical Approach, Routledge, London.

Sevuktekin, M. and M. Nargelecekenler (2010).Analysis of Econometric Time Series. Nobel Publish. Turkey.

Shumway, R.H. and D.S. Stoffer (2006). Time Series Analysis and Its Applications with r examples. Springer Science and Business Media, New York.

Takma, C., H. Atil and V. Aksakal (2012). Comparison of multiple linear regression and artificial neural network models goodness of fit to lactation milk yields. Kafkas Univ. Vet. Fak. Derg. 18(6): 941-644.

Ozer, O., and U. Ilkdogan (2013). The world cotton price forecasting by using Box-Jenkins model. Tekirdag J. Agricultural Faculty. 10(2):13-20.

TUIK, (2014). Statistical Indicators (Istatistik Gostergeler) 1923-2013. Turkish Statistics Institute. Publish No: 4361, Ankara, Turkey.

TUIK, (2015).Cereals and other herbal products (oil seeds) 1988-2015.http://www.tuik.gov.tr/PreTablo.do?alt_id=1001.

Wei, W. W. S. (2006). Time series analysis. Addison Wesley Publishing Company, New York.