FORECASTING THE PRODUCTION OF GROUNDNUT IN TURKEY USING ARIMA MODEL

S. Celik¹, K. Karadas² and E. Eyduran*³

¹Bingol University, Agricultural Faculty, Department of Animal Science, Bingol, Turkey
²Igdir University, Agricultural Faculty, Department of Agricultural Economics, Igdir, Turkey
³Igdir 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).

<table>
<thead>
<tr>
<th>Area Harvested (ha)</th>
<th>Production (thousand tons)</th>
<th>Yield(HG/HA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961-1971 World</td>
<td>17373490</td>
<td>150819.25</td>
</tr>
<tr>
<td>Turkey</td>
<td>12075</td>
<td>29.73</td>
</tr>
<tr>
<td>1972-1982 World</td>
<td>17727820</td>
<td>166907.42</td>
</tr>
<tr>
<td>Turkey</td>
<td>20832</td>
<td>48.39</td>
</tr>
<tr>
<td>1983-1993 World</td>
<td>15994560</td>
<td>184273.31</td>
</tr>
<tr>
<td>Turkey</td>
<td>24907</td>
<td>59.72</td>
</tr>
<tr>
<td>1994-2004 World</td>
<td>20922500</td>
<td>299735.59</td>
</tr>
<tr>
<td>Turkey</td>
<td>30027</td>
<td>79.27</td>
</tr>
<tr>
<td>2005-2014 World</td>
<td>24314700</td>
<td>396111.55</td>
</tr>
<tr>
<td>Turkey</td>
<td>28424</td>
<td>98.66</td>
</tr>
</tbody>
</table>

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).

\[ \Delta X_t = \beta_0 + \beta_1 t + \gamma_1 X_{t-1} + \sum_{i=1}^{h} \gamma_i \Delta X_{t-i} + \varepsilon_t \]  

(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 (Schwarz’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).

\[ X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t \]  

(Eq.2)

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

\[ X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \]  

(Eq.3)

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

\[ X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \]  

(Eq.4)

If \( X_t \) is a sequence of uncorrelated random variables, each with zero mean and variance \( \sigma^2 \), 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)

\[ \gamma(h) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(X_{i+h} - \bar{X})}{n} \]  

Here, \( \gamma(0) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \) are variance

Celik et al.,

hth partial autocorrelation (PACF) is expressed as Eq.6 (Wei, 2006)

\[ p_h = \frac{\gamma(h) - \alpha_1\gamma(h-1) - \alpha_2\gamma(h-2) - \ldots - \alpha_h\gamma(1)}{\gamma(0) - \alpha_1\gamma(1) - \alpha_2\gamma(2) - \ldots - \alpha_h\gamma(h-1)} \]

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)

\[ \text{AIC} = n\ln\hat{\sigma}^2 + 2M \] (Eq.7)

\[ \text{BIC} = n\ln\hat{\sigma}^2 + M\ln n \] (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):

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n}} \] (Eq.9)

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \] (Eq.10)

\[ \text{MaxAPE} = \max_{i=1,2,..,N} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \] (Eq.11)

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i| \] (Eq.12)

Maximum absolute error (MAE) 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).

RESULTS AND DISCUSSION

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.

<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey-Fuller test statistic</td>
<td>1.102</td>
<td>-12.777</td>
</tr>
<tr>
<td>Test critical values:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1% level</td>
<td>-3.537</td>
<td>-3.537</td>
</tr>
<tr>
<td>5% level</td>
<td>-2.908</td>
<td>-2.908</td>
</tr>
<tr>
<td>10% level</td>
<td>-2.591</td>
<td>-2.591</td>
</tr>
<tr>
<td>Prob. *</td>
<td>0.997</td>
<td>0.001</td>
</tr>
</tbody>
</table>


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.

<table>
<thead>
<tr>
<th>Fit Statistic</th>
<th>Stationary R-squared</th>
<th>R-squared</th>
<th>RMSE</th>
<th>MAPE</th>
<th>MaxAPE</th>
<th>MAE</th>
<th>MaxAE</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.218</td>
<td>0.930</td>
<td>8307.93</td>
<td>11.09</td>
<td>69.212</td>
<td>5713.19</td>
<td>27484.6</td>
<td>20.984</td>
</tr>
</tbody>
</table>

Table 4. Model parameters.

<table>
<thead>
<tr>
<th>Model Parameters</th>
<th>Estimate</th>
<th>S.E.</th>
<th>t</th>
<th>Sig. (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA(1)</td>
<td>0.472</td>
<td>0.126</td>
<td>3.757</td>
<td>0.001</td>
</tr>
</tbody>
</table>

S.E.: Standard Error

Figure 6. ACF and PACF graphs of the residual series
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 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 and 167,28 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**


