Journal of Mathematics and Statistics

Original Research Paper

Prediction of Grain Products in Turkey

Özlem Akay, Gökmen Bozkurt and Güzin Yüksel

Department of Statistics, The Faculty of Science and Letters, Çukurova University, Adana, Turkey

Article history
Received: 09-12-2016
Revised: 14-04-2017
Accepted: 04-07-2017

Corresponding Author:
Özlem Akay
Department of Statistics, The Faculty of Science and Letters, Çukurova University, Adana, Turkey
Email: oakay@cu.edu.tr

Abstract: Marketing of agricultural products starts with the planning of production on the farm and ends with the sale of food or other goods to manufacturers or consumers. Overall, marketing is a main part of successful agriculture but its significance is usually underestimated, particularly in developing countries. In Turkey, annual variations in grain production are reasonable and result primarily from changes in yields. Yield variation attend a time trend, commonly taken to be the result of climatic fluctuations and technology. Hence grain growers and the government frequently need to estimate grain yields to make decisions about the future. In this study, production amounts of grain species (wheat, rice and rye) are analyzed by using time series analysis including the Box-Jenkins method, the Exponential Smoothing method and the Regression method for the years 1991-2012. Each time point in the series represents the annual amounts of grain species in tonnes. After the data are stationary, Seasonal Autoregressive Integrated Moving Average models (ARIMA(0,0,1)(1,0,0)$_3$) production of wheat, Power model production of rice and Holt Exponential model of rye were defined as the fitting models for this data. The forecasts are proposed for the years 2013 and 2014, while the increase and decrease in products are determined via the predicted values of grain production by examining changes in recent years.

Keywords: Time Series, Box Jenkins Models, Prediction, Marketing of Grain Products

Introduction

The agriculture sector is an important element of economic development in general. Economic and social development of a country that aims to increase agricultural efficiency, is affiliated with obtaining efficient yield values. In order to sustain human life on the volume of vital agricultural activities, the variation from year to year should be follow-up by national policies helping creation of detailed systematic and consistent statistics.

Many models or forecasting methods have been introduced by researchers for time series prediction data; these include the ARIMA model, Exponential Smoothing Methods, Regression Models and others.

Bornn and Zidek (2012) developed a model that describes and predicts its relationship to wheat yield and climate variables by describing the outcome of a project coordinated by Agriculture and Agri-foods Canada. It was needed as a feature of the model, as it was the ability to balance the effects of noisy measurements with online application plans in the future. The weekly rainfall data and number of rainy days recorded at the main Dry farming research station from 1958 to 1996 (39 years) are collected by Raorane and Kulkarni (2012). Regression and Correlation studies were performed to use yield as dependent variable and rainfall as the independent variable to develop yield prediction model for important crops and to derive information on the rainfall-yield relationship. Varied techniques that are available to estimate crop production and crop area in farming systems are evaluated by Fermont and Benson (2011). They provided a definition as well as summary tables from a database of estimated crop yields collated from a large set of field studies over the previous decades in Uganda. Çelik (2013) analyzed production amount of nutfruit species (pistachios, walnuts, hazelnuts, almond and chestnuts) by Box Jenkins methodology for the years 1936-2011.

This study aims at estimating the average annual production amounts of wheat, rice and rye in the market of agricultural products in Turkey. According to the results obtained in grain production, in order to avoid dependence on foreign sources, importance of measures to be taken is emphasized. Quantities of production of wheat, rice and rye are taken from the data of the office of agricultural crops in Turkey. Forecasts of production were found for the years 2013-2014 by analyzing the data with the help of time series analysis.

Time Series Methods

In the section, three different classical time series approaches, namely the Box-Jenkins method, the Exponential Smoothing method and the Regression method were used to the grain products data for forecasting. These methods and related formulas are discussed.

Box-Jenkins Methods

A general methodology is proposed by Box and Jenkins (1976) for forecasting univariate series starting from a model based on the Autoregressive Integrated models and Moving Average process (ARIMA). One of the most widely used techniques is the Box-Jenkins approach owing to its structured modelling basis and acceptable forecasting performance. It is rather reliable and relatively simple.

Box-Jenkins models can be used to forecast many empirical time series, containing stationary or non-stationary ones, with or without seasonal elements in theory. If the data is seasonal, then the general seasonal $ARIMA(p,d,q)(P,D,Q)_s$ model is described by:

$$B^k Z_t = Z_{t-k} k = 0,1,2 \qquad (1)$$

Where:
B^k = The seasonality with the s delay operator
Z_t = The dependent variable (studied)

Exponential Smoothing Methods

Exponential smoothing can be easily generalized to deal with time series including trend and seasonal variation. Trend and seasonal terms updated via exponential smoothing are introduced.

The exponential smoothing approach can be effectively used to forecast stochastic time series by referring to Bowerman *et al*. (2005). A time trend forecast with the ability to easily adjust for past errors is constructed and prepared follow- on forecasts by this approach.

Holt's Linear Exponential Smoothing method can be used to deal with non-seasonal series by introducing smoothing parameters.

Holt-Winters exponential smoothing method which is extended from Holt's method can be used to forecast data including seasonality and trend. This technique has additive and multiplicative versions, basing on the features.

Holt's Method

This method is to overcome the problem of Brown's Method because it has only one smoothing constant. The obtained estimated linear trend values are sensible to random effects. Brown's Linear Exponential Smoothing technique with single parameter has some similarities with linear moving averages technique. But the difference between first and second smoothing values is added into the first smoothing value. Holt's method has two-parameters to handle data with a linear trend. The technique both smoothes the trend and the slope directly by using different smoothing constants and provides more flexibility in selecting the rates at which slopes and trend are traced.

Consider s_t be the time series to be forecast. This technique is based on the trend of the time series and estimating smoothed versions of the level. The level plus the trend is then extrapolated forward to obtain forecast. The formulas governing the update of the trend and the level are given by:

$$l_t = \alpha s_t + (1-\alpha)(l_{t-1} + a_{t-1}) \qquad (2)$$

$$a_t = \gamma(s_t - l_{t-1}) + (1-\gamma)a_{t-1} \qquad (3)$$

where, l_t is the estimated level and a_t is the estimated trend of the time series. The forecast is given by:

$$\hat{s}_{t+m} = l_t + ma_t \qquad (4)$$

The parameters α and γ are the smoothing constants. These parameters should be optimized for minimizing the sum of squared error.

Regression Analysis

When using regression for prediction, we are often considering time series data and we are aiming to forecast the future.

Let be a time series y_t = 1,2,…, n be to display an increasing, decreasing or curvilinear trend TR_t. A common feature of time series data is a trend. Time series regression models are interested in the observed values y_t, t = 1,2,…,n of the random variable Y_t, $t \in Z$ indexed by time to some function of time. These models are generally used to indicate and so remove the trend component TR_t of a time series under the assumption that the general aspect of the series will proceed into the future. In other words, if the trend TR_t, t = 1,2,…n, is a linear function of time, it is assumed that the trend will continue to be linear in the future over a period of time, say $t = n +1$, $n +2$,…. We forecast or extrapolate future trend values by this assumption. There are different models for trend. Using regression we can model and forecast the linear trend in time series data by including $t = 1,\ldots,T$, as a predictor variable:

$$y_t = \beta_0 + \beta_1 t + \varepsilon t$$

and quadratic trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon t$$

or cubic trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon t$$

The trend should not be seen as anything other than the accumulated effect of fluctuations in some cases. It is thought that the trends and fluctuations have different sort of effects and the time series into the corresponding components are decomposed in other cases. The series must be stationary but it need not be white noise to fit a standard time series model. So, the trends must be removed. Differences of series are taken to remove the trends.

White Noise: ε_t's elements have mean zero and variance σ^2:

$$E(\varepsilon_t) = 0 \quad (5)$$

$$E(\varepsilon_t^2) = \sigma^2 \quad (6)$$

and for which the ε's are uncorrelated across time:

$$E(\varepsilon_t \varepsilon_T) = 0 \ for\, t \neq T \quad (7)$$

A process satisfying (5) through (7) is described as a white noise process.

Stationarity: The time series $\{y_t, t\in Z\}$, with index set $Z = \{0, \pm1, \pm2,\ldots\}$, is said to be stationary if:

- $E|y_t|^2 < \infty$ for all $t\in Z$
- $E(y_t) = m$ for all $t\in Z$ and
- $\gamma_y(r,s) = \gamma_y(r + t, s + t)$ for all $r, s, t\in Z$

That is, a stationary time series $\{y_t\}$ must have constant first moment, where the second moment $\gamma_y(r,s)$ depends only on $(s\text{-}r)$, not on r or s and finite variation features. Considering the last point, the auto covariance function of a stationary process is rewrited as:

$$\gamma_y(t) = Cov(y_s, y_{s+t)} \ for\ s,t \in Z$$

Additionally, when x_t is stationary, we must have:

$$\gamma_y(t) = \gamma_y(-t)$$

where, $t = 0$, $\gamma_y(0) = Cov(y_t, y_t)$ is the variance of y_t, so the autocorrelation function for a stationary time series $\{x_t\}$ is described by:

$$P_y(t) = \frac{\gamma_y(t)}{\gamma_y(0)}$$

Strict stationarity: If the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is the same as that of $(X_{t_{1+h}}, X_{t_{2+h}}, \ldots, X_{t_{k+h}})$ the time series $\{X_t, t\in Z\}$ is said to be strict stationary. That is, strict stationarity means that the joint distribution only depends on the 'difference' h, not the time $(t_1,\ldots,t_k)$.

Case Study

The data set studied here is Turkey's yearly grain products in the period of 1991-2012 obtained from Turkish Grain Board. It is a yearly time series with 22 observations in 22 years for each products. In this study, the SPSS 11.5 package software is used to apply the methods. There are three steps in usage of these methods. In the first step, the types of components in the time series are identified. After that, the values of parameter are determined. Once the parameter values are determined, the process of forecasting can be carried out. Finally, forecast values dependent on the periods are displayed.

In order to measure the accuracy of all models-ARIMA model, Exponential Smoothing Method and Regression model- in making the forecasts, their test accuracy of forecasting is determined. The accuracy of a forecasting model is determined from the size of forecasting error. The best model produces the smallest forecasting error. The criteria chosen to select the best model in making the forecasts is Mean Squared Error (MSE).

In this study, our aim is to fit the best models for grain products in Turkey. Firstly, the plot for the time series data of wheat is given below.

Time series graphs for the period of 1991-2012, are shown in Fig. 1. It is seen more clearly that a cyclic pattern exists since data exhibit rises and falls that are not of fixed period. Since the fluctuations are not of fixed period, they are cyclical behavior, which is somewhat irregular.

In the second stage, the partial autocorrelations and the autocorrelations for the time series are examined to provide a quantitative conclusion about its periodicity for the time series. It is seen more clearly if this shows the characteristics of a seasonal series. Figure 2 plots the pattern of the Autocorrelation Function (ACF) of the time series and Fig. 3 plots the pattern of the Partial Autocorrelation Function (PACF) of the time series.

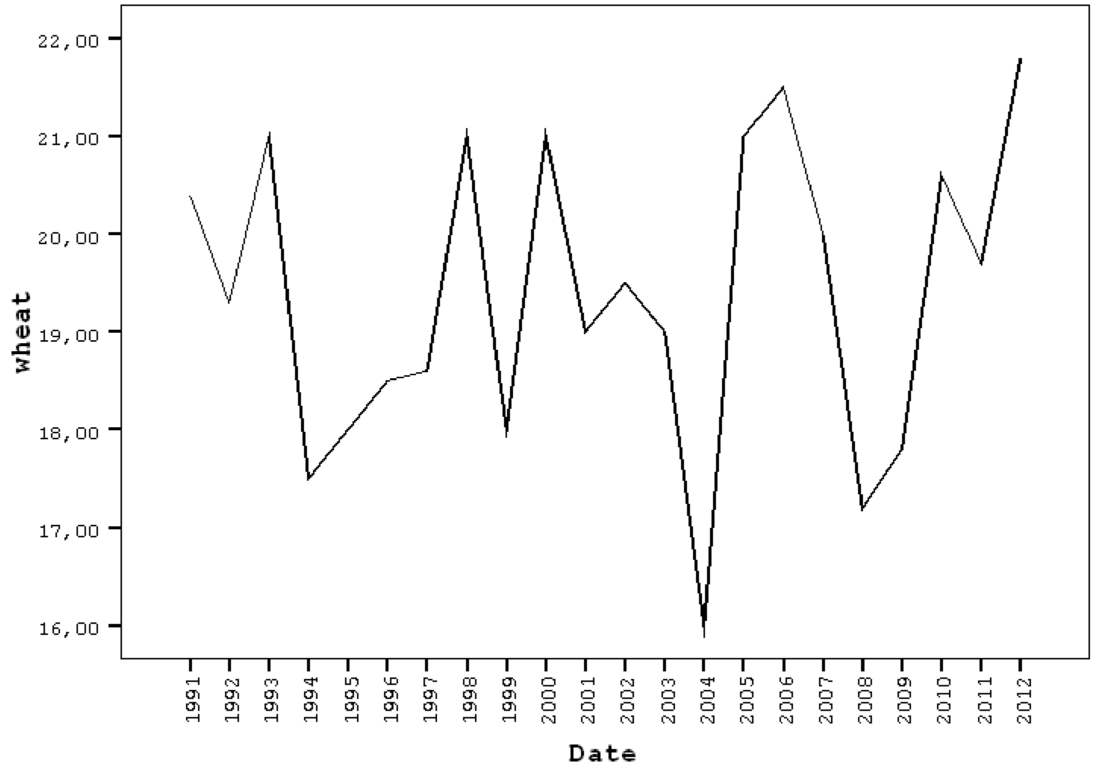

Fig. 1. Graph of wheat production over time

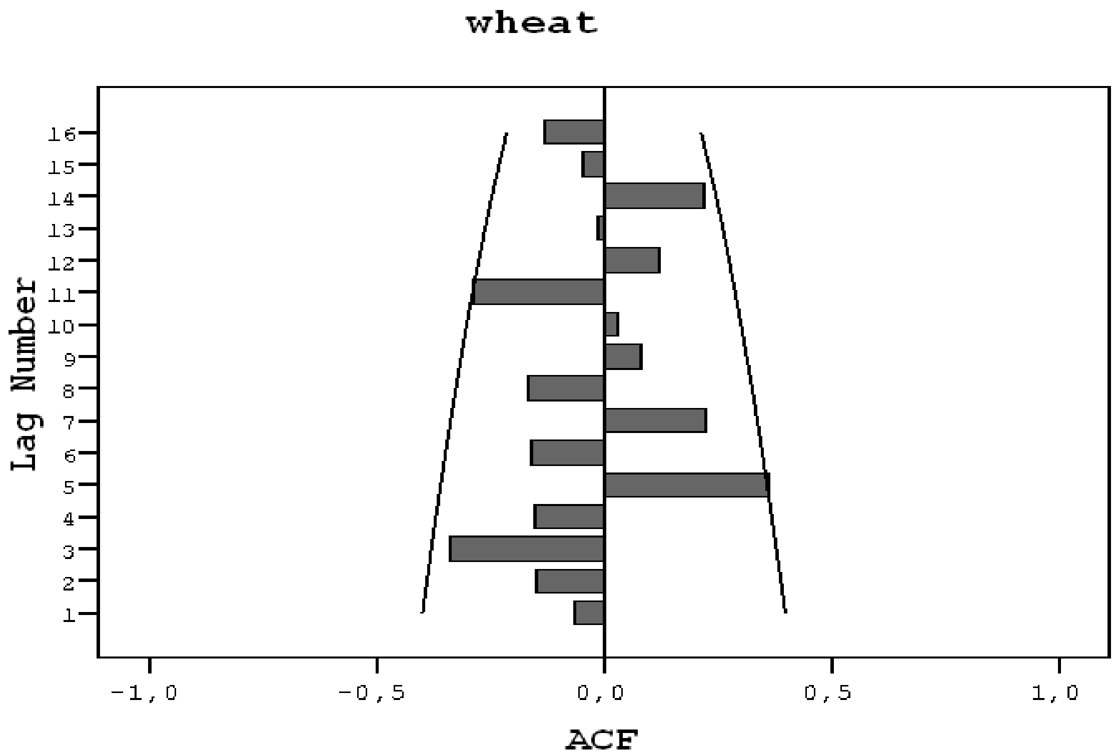

Fig. 2. The autocorrelation function ACF

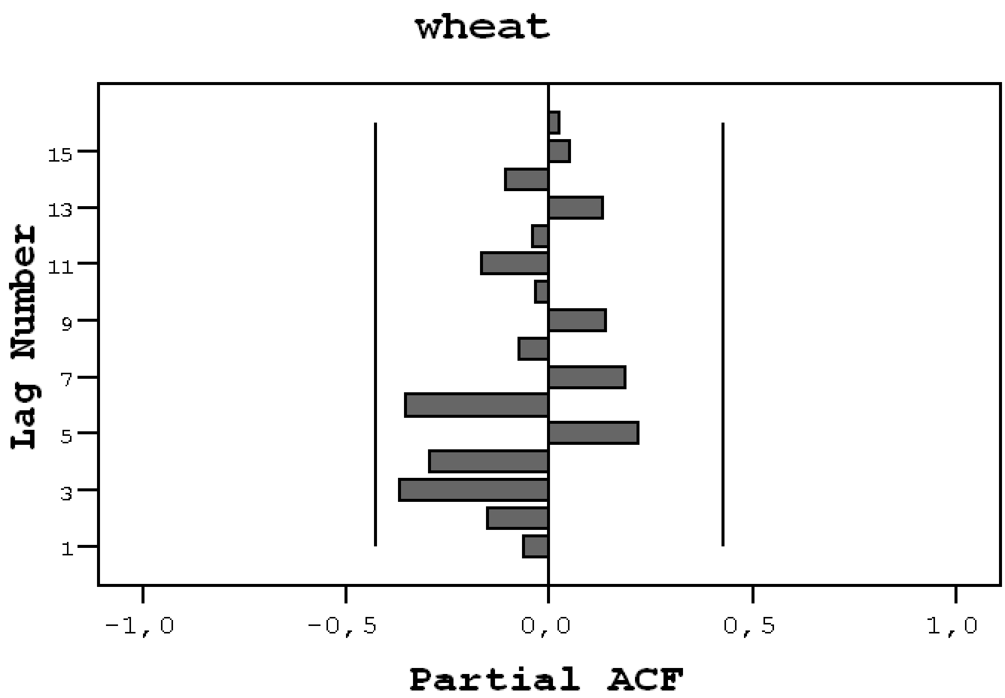

Fig. 3. The partial autocorrelation function PACF

Table 1. The table estimated parameters for the ARIMA models

Models	Parameter estimate	Significance of coefficient	Box Ljung test	AIC	AICC
ARIMA(0,0,1)	MA(1) = 0.151	0.535 > α	Q = 23.799	24.652	29.283
			P = 0.125 > α		
ARIMA(1,0,0)(1,0,0)	AR(1) = -0.337	0.163 > α	Q = 12.685	20.257	24.888
	SAR(1) = -0.529	0.025 < α	P = 0.696 > α		
ARIMA(0,0,1)(1,0,0)	MA(1) = 0.669	0.010 < α	Q = 10.862	17.787	22.418
	SAR(1) = -0.538	0.019< α	P = 0.818 > α		
ARIMA(0,0,1)(1,0,1)	MA(1) = 0.634	0.029 < α	Q = 7.756	17.690	22.321
	SAR(1) = -0.288	0.480 > α	P = 0.933 > α		
	SMA(1) = 0.408	0.373 > α			
ARIMA(0,0,0)(1,0,1)	SAR(1) = 0.018	0.974 > α	Q = 14.613	20.470	25.101
	SMA(1) = 0.576	0.293 > α	P = 0.553 > α		
ARIMA(1,0,1)(0,0,0)	AR(1) = 0.414	0.309 > α	Q = 20.573	20.711	25.343
	MA(1) = 0.987	0.799 > α	P = 0.195 > α		
ARIMA(0,0,1)(0,0,1)	MA(1) = 0.671	0.027 < α	Q = 9.305	17.171	21.803
	SMA(1) = 0.619	0.058 > α	P = 0.900 > α		

According to ACF and PACF graphs, the wheat production series do not exceed the confidence interval. Hence the dependent variable is white noise.

Various ARIMA models were tested by looking at ACF and PACF graphs to determine the model. Values for this model are given in the Table 1.

In Table 1, although ARIMA(0,0,1)(1,0,1) and ARIMA(0,0,1)(0,0,1) models have smaller than AIC and AICC values than the other models, it is suitable to choose the ARIMA(0,0,1)(1,0,0) model with three periods as the best fitting model because of the parameter significance. The value of the Box-Ljung Q statistic is 10.862 and the p-value is 0.818, thus, we can not reject the null hypothesis of no autocorrelation in the residuals. As a result, model for wheat production is given below by using Equation 1:

$$Z_t = -0.538 Z_{t-3} - 0.669 \varepsilon_{t-3} + \varepsilon_t \quad (8)$$

Hereafter, the model for rice production is investigated. The plot for the time series data of rice is as follows.

According to Fig. 4, there is a general upward trend on time series plots of rice data.

DOI: 10.3844/jmssp.2017.220.230

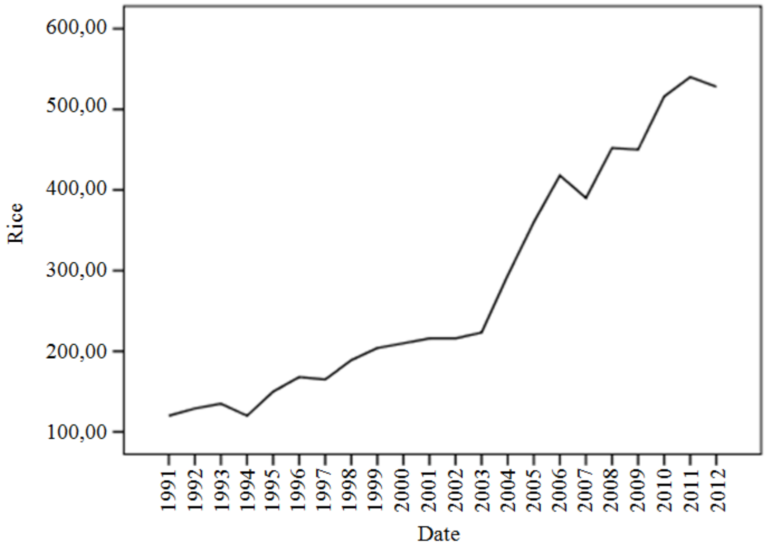

Fig. 4. The time series graph for data of Rice

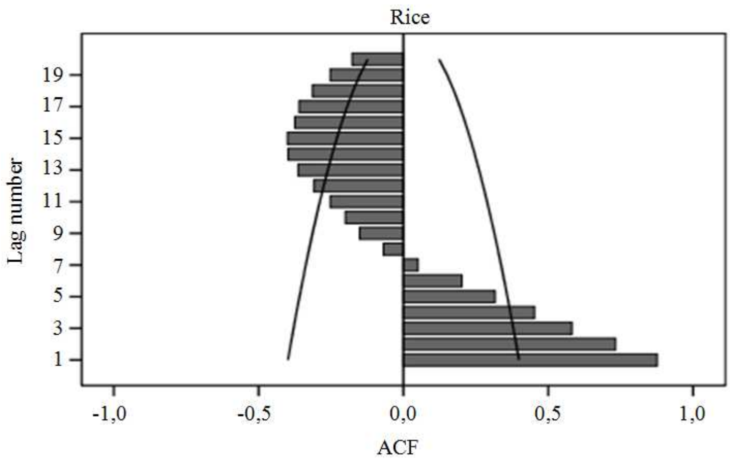

Fig. 5. The autocorrelation function ACF

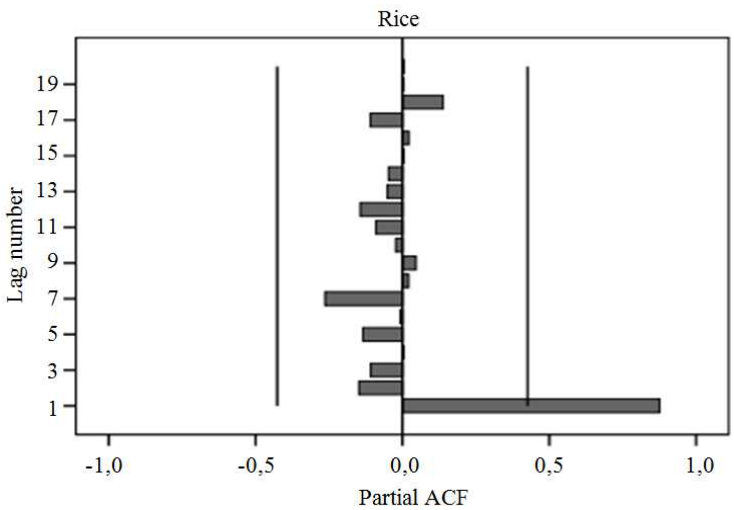

Fig. 6. The partial autocorrelation function PACF

The autocorrelation function ACF and the partial autocorrelation function Partial ACF are represented in Fig. 5 and 6. It is immediately clear from Fig. 6 that the errors are correlated with one another and hence not independent. Lags 1 is significant autocorrelations, implying that, in this model, when the error in predicting number of products increases, the increase tends to be followed by other increases in error and vice versa. It is not independent. In a situation like this (with positive autocorrelations), statistical regression models under-estimate the size of the error variance and tend to over-estimate the significance of the trend and the precision of our forecasts.

Unit root test is done to examine the stationarity of the rice production series. Stationarity is a matter of concern in two important areas. First, stationarity of regressors is assumed in the deviation of standart inference procedures for regression models. Second, a crucial question in the ARIMA modeling of a single time series is the number of times the series needs to be first differenced before on ARIMA model is fit. Each unit root requires a differencing operation.

The results of the unit root test using the models suggested by Dickey and Fuller (1981) are shown in the Table 2. In routine practice, three types of models are used while testing for unit root.

According to Dickey and Fuller (1981) unit root test results obtained for each three models, the rice production series has a unit root. The null hypothesis of a unit root is not accepted in favour of the stationary alternative in each case if the test statistic is more negative than the critical value $\left(t_{\hat{\delta}}\right)$. So, this series is taken difference in order to be stationary.

Considering ACF and PACF (Fig. 7 and 8) graphs of the rice production series taken first difference, this series do not exceed the confidence interval. So, this series are stationary.

After giving the required graphs, the parameter estimates of the models for rice data can be given as follows.

Table 3 shows that the Power model has the smallest MSE (0.062) and the smallest adjusted R^2 (0.759), therefore, this model is chosen as the fitting model for rice data. This model is as follows:

$$Z_t = 72.222t^{0.559} \quad (9)$$

The remaining of this paper is related with searching the model for rye production in Turkey. The plot for the time series data of rye is given below.

Table 2. The results of the Dickey-Fuller unit root test for the rice production.

Significance level	Model with drift and trend (τ_τ)	Model with drift, but no trend(τ_μ)	Model without drift, trend (τ)
%1	-4.380	-2.539	-2.66
%5	-3.60	-1.729	-1.950
%10	-3.240	-1.328	-1.600
DF statistics $\left(t_{\hat{\delta}}\right)$	-1.939	0.523	2.930

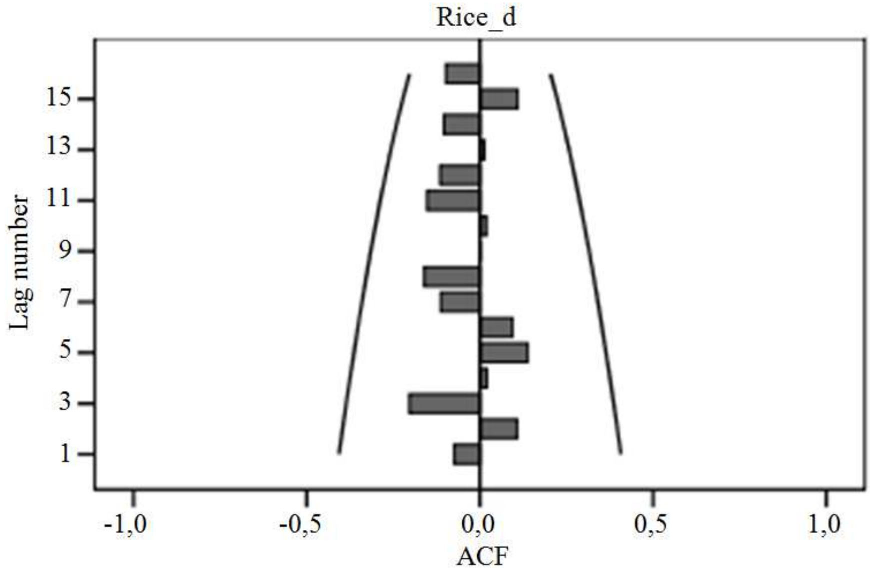

Fig. 7. The autocorrelation function ACF taken first difference

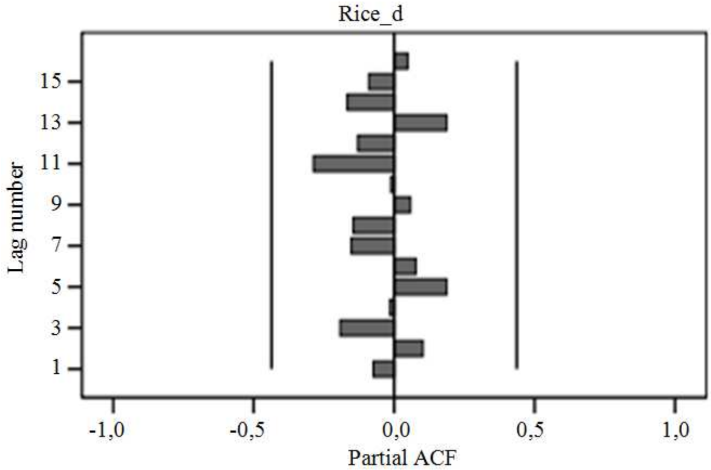

Fig. 8. The partial autocorrelation function PACF taken first difference

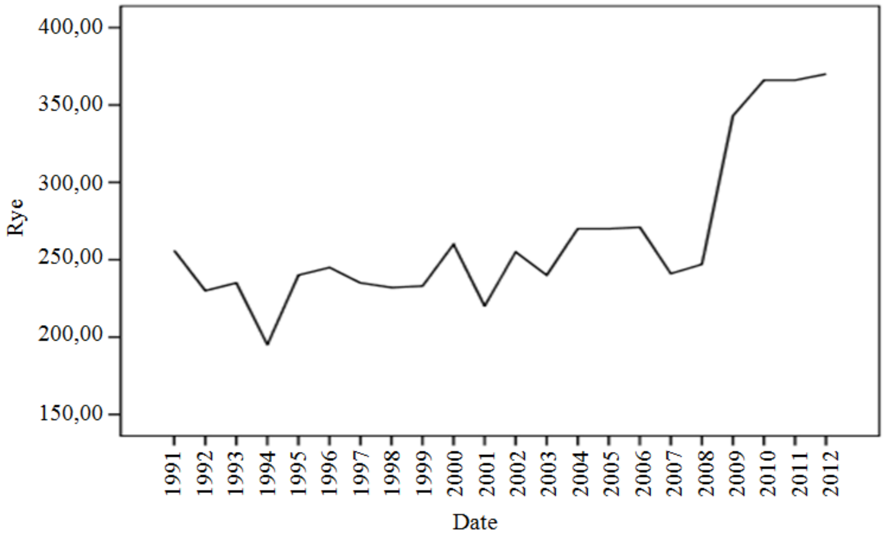

Fig. 9. The time series graph for data of Rye

Table 3. The estimated parameters for the regression models

	Parameter estimate	Significance of Coefficient	Adjusted R^2	Mean squared error
β_2	-0.123	0.974		
Quadratic β_1	0.942	0.000	0.967	701.505
β_0	120.500	0.000		
β_3	-10.641	0.269		
Cubic β_2	2.060	0.051	0.968	689.970
β_1	-0.032	0.266		
β_0	142.866	0.000		
Power β_1	0.559	0.000	0.769	0.062
β_0	72.222	0.000		
Holt Exponential	$\alpha = 0.904$	0.001	0.960	851.939
Smoothing	$\gamma = 1.200$	1.000		
ARIMA(1,1,1)	AR(1) = -0.758	0.584	0.959	924.464
	MA(1) = -0.668	0.673		
ARIMA(1,1,0)	AR(1) = -0.076	0.752	0.958	890.246
ARIMA(0,1,1)	MA(1) = -0.065	0.787	0.958	891.082

Figure 9 shows that the trend occurs after 2009 on time series plots of rye data. In Turkey, rye is predominantly used in the industry. However, in recent years, as a result of changing eating habits, this product is started to be used in human nutrition. Also, production has increased due to increase in efficiency despite the decrease of rye planting areas in Turkey.

It is immediately clear from Figure 10 and 11 that the errors are correlated with one another and hence not independent. As lags exceed the confidence interval, autocorrelations are significant.

Unit root test is done to examine the stationarity of the rye production series. The results of the unit root test using the models suggested by Dickey and Fuller (1981) are shown in the Table 4.

Since the test statistics of three models is smaller than the critical value, the null hypothesis of a unit root can not accept. So, rye production series is been stationary by taking difference of this series.

Considering ACF and PACF (Fig. 12 and 13) graphs of the rye production series taken first difference, the rye series do not exceed the confidence interval. So, it is possible to say that this series are stationary.

The information of the three different models for the rye data is shown in the Table 5.

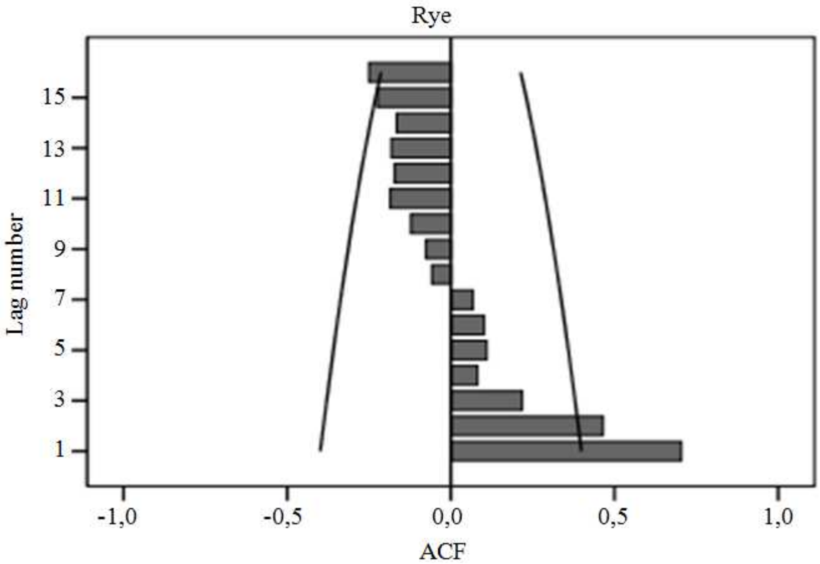

Fig. 10. The autocorrelation function ACF

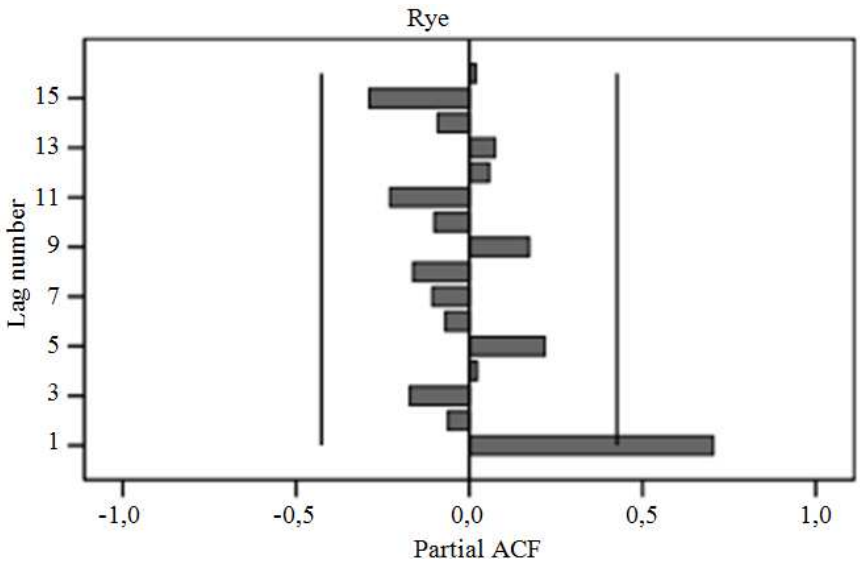

Fig. 11. The partial autocorrelation function PACF

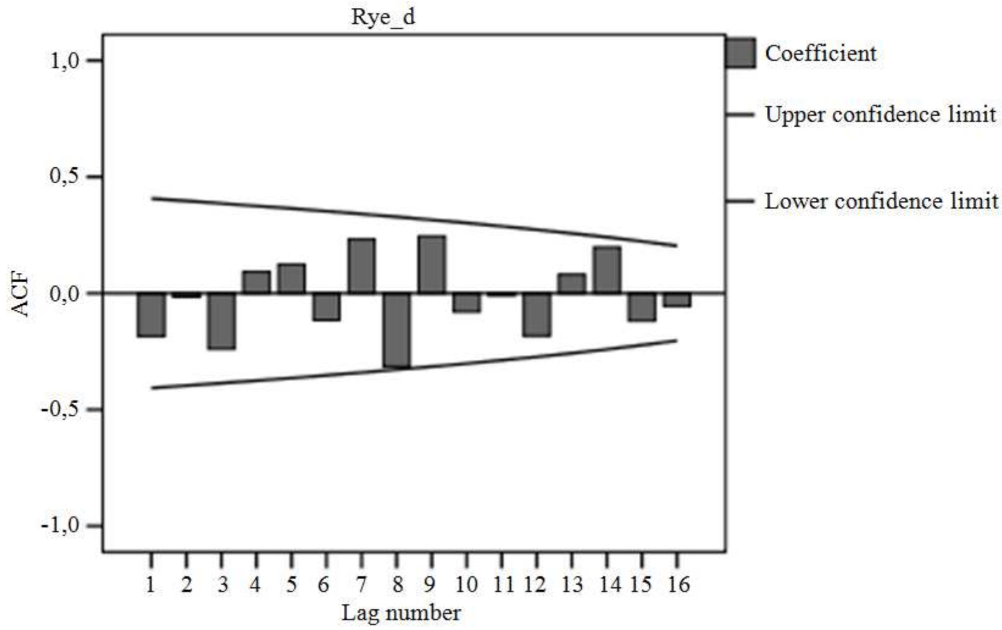

Fig. 12. The autocorrelation function ACF taken first difference

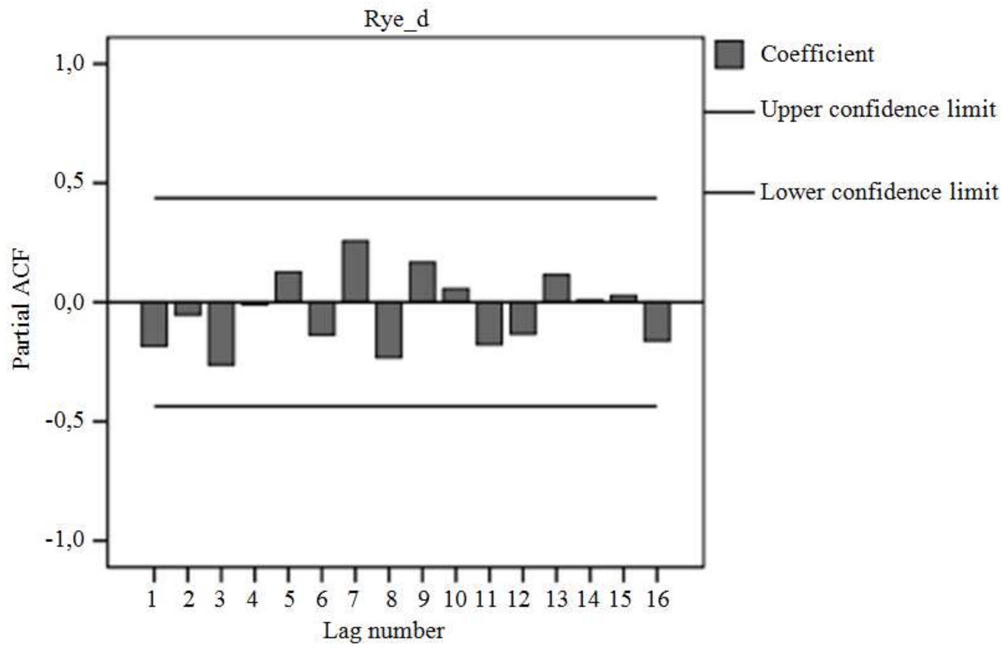

Fig. 13. The partial autocorrelation function PACF taken first difference

Table 4. The results of the Dickey-Fuller unit root test for the rye production

Significance level	Model with drift and trend (τ_τ)	Model with drift, but no trend (τ_μ)	Model without drift, trend (τ)
%1	-4.380	-2.539	-2.660
%5	-3.600	-1.729	-1.950
%10	-3.240	-1.328	-1.600
DF statistics $\left(t_{\hat{\delta}}\right)$	-2.386	-0.601	0.686

Table 5. The estimated parameters for the exponential smoothing and ARIMA models

	Parameter estimate	Significance of coefficient	Adjusted R^2	Mean squared error
Holt Exponential	$\alpha = 0.800$	0.002	0.785	29.655
Smoothing	$\gamma = 5.120$	1.000		
Winter Exponential	$\alpha = 0.900$	0.00	0.820	30.395
Smoothing	$\gamma = 2.405$	1.000		
(Additive)	$\delta = 7.310$	1.000		
Model	AR(1) = 0.550	0.100> α	0.485	28.007
ARIMA(1,0,0)(0,1,1) d = 1	SMA(1) = 0.592	0.277> α		

Table 5 shows that although ARIMA$(1,0,0)(0,1,1)_3$ has the smallest MSE value of 28.007, coefficients of this model are not significant. Since the Holt Exponential model has adjusted R^2 (0.785) and the second smallest MSE (29.655), this model is chosen as the fitting model for rye data. After computing L_t = 204.481 and b_t = 4.828 values using Equation 2 and 3, the model for rye can be given as follows:

$$\hat{y}_{t+m} = 204.481 + m(4.828) \quad (10)$$

Conclusion

After constructing the appropriate models determined by analyzing the last 22 years important variables of the market for agricultural products in Turkey, forecasts of wheat, rice and rye are found for the years 2013 and 2014.

Forecasts of wheat production computed using Equation 8 are found 17.18 and 18.98 million tons, respectively, for 2013 and 2014. A decrease of 21% is observed in wheat products in comparison to the year 2012.

Forecasts of rice production computed using Equation 9 are found 416.681 and 426.712 thousand tons, respectively, for 2013 and 2014. A decrease 28% is observed in rice products in comparison to 2012.

Forecasts of rye production computed using Equation 10 are found 334.66 and 335.627 thousand tons, respectively, for 2013 and 2014. A decrease of 8.5% is observed in rye products in comparison to 2012.

Since decline in wheat, rye and rice production in Turkey is causing dependence on foreign production, economic and social losses will follow. Therefore, production amounts of wheat and rye need to be increased.

Acknowledgment

The author would like to thank anonymous referees for their helpful comments.

Author's Contributions

Özlem Akay: Participated in all experiments, coordinated the data-analysis and contributed to the writing of the manuscript.

Gökmen Bozkurt: Contributed to acquisition of data.

Güzin Yüksel: Designed the research plan and organized the study.

Ethics

This paper is constructed based on our research and each step and all findings are original.

References

Bornn, L. and J.V. Zidek, 2012. Efficient stabilization of crop yield prediction in the Canadian Prairies. Agric. Forest Meteorol., 152: 223-232. DOI: 10.1016/j.agrformet.2011.09.013

Bowerman, B.L., R.T. O'Cannell and A.B. Koehler, 2005. Forecasting, time series and regression: An applied approach. Thomson Brooks/Cole, Belmont CA.

Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. 1st Edn., Holden-Day, San Francisco, ISBN-10: 0816211043, pp: 575.

Çelik, Ş., 2013. Modelling of production amount of nuts fruit by using box-Jenkins technique. YYÜ Tar Bil Derg, 23: 18-30.

Dickey, D. and W. Fuller, 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica, 49:1057-1072

Fermont, A. and T. Benson, 2011. Estimating yield of food crops grown by smallholder farmers. Internetional Food Policy Research Instıtute, IFPRI Discussion Paper 01097.

Raorane, A.A. and R.V. Kulkarni, 2012. Data mining: An effective tool for yield estimation in the agricultural sector. Int. J. Emerg. Trends Technol. Comput. Sci., 1: 75-79.

Turkish Grain Board, www.tmo.gov.tr