# Prediction of Combine's Number By Using ARMAX Model in Turkey

**Gökhan Unakıtan**[1] **Bahattin Akdemir**[2]

[1]Namik Kemal University, Faculty of Agriculture, Department of Agricultural Economics, 59030, Tekirdag, Turkey.

[2]Namik Kemal University, Faculty of Agriculture, Department of Biosystem Engineering, 59030, Tekirdag, Turkey.

**Abstract:** The combine is the most expensive machine in agricultural production. Prediction of the demand of the combine will be important to plan of the future decision of producers and distributors. The aim of the study was to estimate the number of combines in the 2009-2020 periods in Turkey. This study is based on data concerning the number of combines and real agricultural gross domestic product data in Turkey for the period 1971-2008. ARMAX model was used to predict the variation of the number of the combine for the period of 2009-2020 in Turkey. According to the model results, the number of combines is estimated to reach 13292 units in 2020. The annual increase in the future number of the combines is estimated to be 54–55 units. The results from the model deviate an average 4.11% from the observed data, which may be regarded as definitely within an acceptable range. It is assumed that the estimated results created by using ARMAX for the period 2009-2020 will be useful for combine producers and distributors.

**Key words:** agricultural machinery; Box-Jenkins method; combine; forecasting

**JEL Kodları:** C22; C53; Q18

## Türkiye'de Biçerdöver Parkının ARMAX Model ile Tahmini

**Özet:** Biçerdöverler tarımsal üretimde kullanılan en pahalı makinelerdir. Bu nedenle biçerdövelerin talep tahmini üreticiler ve dağıtım firmalarının gelecek ile ilgili kararları açısından önem taşımaktadır. Çalışmanın amacı Türkiye'de 2009-2020 yılları dönemi için biçerdöver sayısının tahmin edilmesidir. Çalışmada kullanılacak veriler Türkiye'de 1971-2008 yıllarını içeren biçerdöver sayısı ve reel tarımsal gayri safi üretim değerine dayanmaktadır. Türkiye'nin 2009-2020 yılları arasında biçerdöver sayısındaki değişimler ARMAX model yardımıyla tahmin edilmiştir. Model sonuçlarına gore, 2020 yılında biçerdöver sayısının 13292 adete ulaşacağı tahmin edilmektedir. Biçerdöver sayısındaki artışın yıllık 54-55 adet olacağı tahmin edilmektedir. Model sonuçlarının gözlem değerlerinden sapmasının ortalama oranı %4,11 olarak hesaplanmıştır ve bu kabul edilebilir bir aralıktadır. 2009-2020 yılları için tahmin edilen ARMAX modelin sonuçlarının biçerdöver üreticisi ve dağıtım firmaları için yararlı olacağı öngörülmektedir.

**Anahtar kelimeler:** tarım makineleri, Box-Jenkins yöntemi, biçerdöver, tahminleme

**JEL Codes:** C22; C53; Q18

## 1. Introduction

Machines and equipment used in agricultural production have made the process more effective. Among these mechanized machines, combines have priority in terms of harvesting (Kaygısızel, 2006).

The number of combines in Turkey rapidly increased up to the 1980s; however, increases in the 1990s in the number of combines had small fluctuations. In 2008 there were 13084 combines in Turkey (FAO 2014). Given the grain production areas in Turkey, it is obvious that there is still not enough machinery. Currently, 24% of Turkey's population is rural; this rate was 75% in the 1950s (TSI 2014). Development in agricultural mechanization has played a considerable role in reducing the agricultural population, especially an increased number of tractors in agricultural production, reducing the need for labor.

The aim of the study was to determine development of combine usage and to estimate the number of combines in the 2009-2020 periods in Turkey. Apparently, no study has been published in the literature comparable to our study; therefore, this study is a first for Turkish agriculture. Data obtained from the study will guide investigators and agricultural policy makers in solving structural problems in agriculture. This study will play an important role for market-leading manufacturers and importers of combines.

The Box-Jenkins (1970) method, also known as the method of ARIMA models with enhanced type AR (I) MAX models, was used. In many studies, time series analysis utilizing tractors and agricultural equipment and machinery has been conducted (e.g., Madan,1989; Biondi et al 1998; Pawlak, 1999; Unakitan and Akdemir, 2007). There is not any research for prediction of the number of combine in the literature. This research will be first for this subject.

## 2. Material and methods

This study is based on data concerning the number of combines and real agricultural gross domestic product (AGDP) data in Turkey for the period 1971-2008. The data were obtained from the Food and Agriculture Organization (FAO) and Turkish Statistical Institute (TSI). According to the data, the number of combines has increased from 8662 to 13084 from 1971 to 2008 in the Turkish agriculture sector (FAO, 2014). The data were transformed into natural logarithms.

ARMAX model was used to predict the variation of the number of the combine for the period of 2009-2020 in Turkey. The ARMAX model is an extension of the Box-Jenkins autoregressive moving average (ARMA) model with explanatory exogenous variables (X) (Lim et al. 2007). The ARIMA method, or autoregressive integrated moving average, is one of the models used in time series forecasting analysis (Ho and Xie, 1998; Zhang, 2001; Ho et al. 2002). The ARIMA method originated from the autoregressive model (AR), the moving average model (MA), and a combination of the AR and MA, the ARMA model, which were introduced in 1926, 1937, and 1938, respectively (Blanchard and Decrochers 1984; Brown et al. 1984; Kamal and Jafri 1997; Saab et al. 2001). ARIMA models allow each variable to be explained by its own past or lagged, values and stochastic error terms.

A reliable forecast is needed that can be done accurately via somewhat sophisticated techniques, such as autoregressive (integrated) moving average cause effect (AR(I)MAX), rather than the simple cause-effect regression technique. The cause-effect regression technique does not recover lagged systematic effects or unexpected changes for an accurate forecast, but an AR(I)MAX model includes (a) autoregressive filters to account for systematic effects and (b) moving average filters to account for shock effects in itself in addition to explanatory variables in the cause-effect regression model. Therefore, the AR(I)MAX technique is able to outperform the simple cause-effect technique in terms of forecast accuracies (Akal, 2004).

An ARMAX model includes dynamic autoregressive and moving average components in addition to theoretical explanatory variables that explain variations in endogenous variables. The ARMAX model accounts for influences other than theoretical explanations; therefore, the ARMAX technique corrects the deficiencies of the econometric cause-effect technique by using dynamic filters to explain variations in endogenous variables. An explanatory part is integral to the ARMA process to construct the ARMAX technique. The ARMA part is considered a special case of ARMAX with no regressor by Greene (2000). In other words, an ARMAX (p, d, q, X) model can be explicitly represented in Equation 1,

$$y_t = \mu + p_1 y_{t-1} + p_2 y_{t-2} + \cdots + p_p y_{t-p} + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} + \varepsilon_t - q_1 \varepsilon_{t-1} - q_2 \varepsilon_{t-2} - \ldots - q_q \varepsilon_{t-q}$$

(Equation.1)

where $\mu$ is the constant term, $\beta$ parameters are the regressors for lagged distributed *x* explanatory variables, *p* parameters are the autoregressive parameters for lagged distributed *y* exogenous dependent variables, *q* parameters are the moving average parameters for lagged distributed $\varepsilon$ stochastic variables, and *d* is the degree of differencing. The same lag structure required in autoregressive distributed lag models is not necessarily applied to $y_t$ and $\varepsilon_t$. $\varepsilon_t$ is the serially undistributed constant variance random variable. Harvey (1990) and Franses (1991) treat the ARMAX problem as an extension of ARIMA modeling because the disturbances are generated by an ARMA (p, q) process (Akal 2004). Extensive discussion of ARMAX modeling and estimation can be found in Franses (1991) and Greene (2000).

The important point to note in ARIMA modeling is that we must have either a stationary time series or a time series that becomes stationary after one or more differencing to be able to use it.

Once the forecasts of the model have been obtained, appropriateness is checked by nonparametric methods (*root mean square error,* *mean percentage error*) as well as parametric methods. If the model has been designed for forecasting purposes, the e*x post rms (root mean square) forecast error* is one of the most important criteria for performance. In an *ex post* forecast, the forecast result can be compared over the forecast range, providing a measure of the model's ability to forecast. A useful simulation statistic related to *rms* simulation error and applied to the evaluation of historical simulation or *ex post* forecast is *Theil's inequality coefficient*. Note that the numerator of *U* is simply the *rms* simulation error, but the scaling of the denominator is such that *U* will always fall between 0 and 1. If *U* = 0, the predictive performance of the model is perfect (Pindyck and Rubinfeld 1997).

## 3. Results

In the first stage of the analysis, the stationarity of combines was tested by Augmented Dickey-Fuller (ADF) test, developed by Dickey and Fuller (1979). MacKinnon's critical value (1991), is used to give a decision about stationary of the data (Gujarati, 1998). As shown in Table 1, the absolute value of the ADF test statistic is higher than that of the MacKinnon critical value at 99% significance level. Accordingly, the series is determined to be stationary. Its means, d, is equal to zero.

A correlogram is a guiding property for determining autoregressive (AR) and moving averages (MA) processes. Furthermore, Akaike information criteria (AIC) and Schwartz Bayesian criteria (SBC) are widely used instead of correlograms for successfully choosing a model regarding goodness of fit. In the study, a model was chosen that provided the smallest AIC and SBC values. According to this, the AR and MA processes were determined as AR (1) and MA (5). Moreover, the model was determined to be ARMAX (1,5,X) because the explanatory variable was agricultural gross domestic product. In line with the conclusion for model identification, our model is shown in Equation 2.

$$\Delta^d Y_t = v_t + \alpha_8 Y_{t-8} - \beta_{13} v_{t-13} + u_t$$
$$y_t = \mu + p_1 y_{t\text{-}1} + \beta_1 x_{t\text{-}1} + q_5 \varepsilon_{t\text{-}5} \quad \text{(Equation.2)}$$

In this stage, the nonlinear least squares method of estimation was considered to estimate the parameters (Equation 3). We obtaied the estimates shown in Table 2.

$$y_t = 3.51 + 0.35 y_{t\text{-}1} + 0.85 + 0.94 \varepsilon_{t\text{-}5}$$

(Equation.3)

The ex post simulation errors are given in Table 3. The predicted values show that these values are very close to the actual ones. $U^M$, obtained through decomposition of the Theil *U* value, indicates that the model does not have systematic error. On the other hand, a low $U^S$ value shows how well our forecast replicates the volatility of the actual series. $U^C$ is the covariance proportion and offers a measure of unsystematic error in the forecast. Smaller values for $U^M$ and $U^S$ and larger values for $U^C$ suggest good predictions. These *U* coefficients imply that future numbers of combines serve as the best forecasting tool.

Table 1. Results of ADF test for lncombine

| ADF test statistics | | Critical value | | |
|---|---|---|---|---|
| | | 1%* | 5% | 10% |
| Ln(combine) | -3.8993 | -3.6353 | -2.9499 | -26133 |

*MacKinnon critical values for rejection of hypothesis of a unit root

Table 2. Model results

| Dependent variable: Ln(Number of combines) | | | |
|---|---|---|---|
| Independent variables | Coefficient | t-statistic | Probability |
| Constant | 3.5161 | 2.3153 | 0.0276 |
| LnAGDP(-1) | 0.3571 | 3.8772 | 0.0005 |
| AR(1) | 0.8582 | 8.2159 | 0.0000 |
| MA(5) | 0.9471 | 23.035 | 0.0000 |
| Adjusted $R^2$ | | | 0.72 |
| Log likelihood | | | 65.05 |
| Akaike info criterion | | | -3.59 |
| Schwarz criterion | | | -3.41 |
| F-statistic | | | 30.41 |

Table 3 Results of simulation error of ARMAX model

| | |
|---|---|
| Theil inequality coefficient (U) | 0,0062 |
| Bias proportion ($U^M$) | 0.0120 |
| Variance proportion ($U^S$) | 0.2467 |
| Covariance proportion ($U^C$) | 0.7413 |

One of the indicators is validation of the estimated data for the determination of suitability.The truth of this observation for the last five years of data was compared with predicted values. The results from the model deviate an average 4.11% from the observed data, which may be regarded as definitely within an acceptable range (Table 4).

According to the model results, the number of combines is estimated to reach 13292 units in 2020 (Table 5). In Figure 1, the number of combine observations for Turkey, and estimated values are given. The annual increase in future number of the combines is estimated to be 54–55 units.

Table 4. Validation of ARMAX model

| Years | Actual consumption | Forecasted consumption | Absolute value of deviation | Deviation as a percentage of actual consumption |
|---|---|---|---|---|
| 2004 | 11519 | 12398.16 | 879.16 | 7.0910 |
| 2005 | 11811 | 12489.66 | 678.66 | 5.4337 |
| 2006 | 12359 | 12740.40 | 381.39 | 2.9935 |
| 2007 | 12775 | 12873.56 | 98.56 | 0.7656 |
| 2008 | 13084 | 1255.74 | 533.26 | 4.2488 |

Table 5. Estimation of the number of combines for period 2009-2020

| Year | Combine | Differences | Year | Combine | Differences |
|---|---|---|---|---|---|
| 2009 | 12721 | -- | 2015 | 13020 | 53.92 |
| 2010 | 12751 | 30.57 | 2016 | 13074 | 54.10 |
| 2011 | 12805 | 53.56 | 2017 | 13128 | 54.25 |
| 2012 | 12858 | 53.61 | 2018 | 13182 | 54.37 |
| 2013 | 12912 | 53.73 | 2019 | 13237 | 54.58 |
| 2014 | 12966 | 53.84 | 2020 | 13292 | 54.76 |

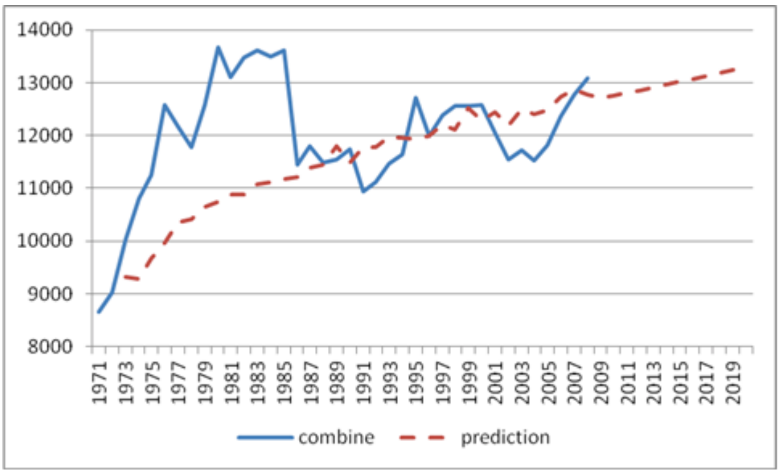


Figure 1. Predicted number of combines (units)

Table 6. Annual working day of a combine

| Year | Combine | Day | Year | Combine | Day |
|---|---|---|---|---|---|
| 2009 | 12721 | 75.81 | 2015 | 13020 | 74.06 |
| 2010 | 12751 | 75.62 | 2016 | 13074 | 73.76 |
| 2011 | 12805 | 75.31 | 2017 | 13128 | 73.45 |
| 2012 | 12858 | 74.99 | 2018 | 13182 | 73.15 |
| 2013 | 12912 | 74.68 | 2019 | 13237 | 72.85 |
| 2014 | 12966 | 74.37 | 2020 | 13292 | 72.55 |

About 13.5 million hectares of arable land are harvested by combines in Turkey (TSI 2014). According to information from AMA (Automotive Manufacturers Association) (2014), the average daily working hours and harvesting capacity of a combine is 10 hours and 1.4 ha. respectively. The number of combines was 12721 in 2009. Average usage of the combine was determined approximately 76 days per year. As a result of increasing numbers of combines, it is estimated that average usage of a combine will decrease to 72 days (Table 6).

## 4. Conclusion

In our study, the ARMAX model was used to predict the variation of the number of the combine for the period of 2009-2020 in Turkey. According to the results of the model, the number of combine increases about 55 per year and is expected to reach 13292 units in 2020. The old stock of combines and presentation of new combines on the Turkish market will keep the combine market alive for a period.

Arable land in Turkey has decreased to 21 million hectares from 25 million in the 1970s (TSI, 2014). The most important reasons for the decrease are non-agricultural activities and establishment of industrial and tourism enterprises in agricultural areas. A decrease in the amount of agricultural area and increase in the number of combines has caused to decrease yearly working hours of combines. This situation indicates that the stock of combines will be narrowed.

During this research, there are not any recorded data for the daily and annual working hours, production number of combines for Turkey. There are only recorded data for the number of exported and imported combines after year 2000. It is required that studies on the determination of working hours of the combines to establish a database, and help to another research about combine. There is not any research on daily and annual working hours of combines in Turkey. This study determined that yearly and daily working hour of combine has importance to lead the agricultural machinery market.

The combine is the most expensive machine in agricultural production. Prediction of the demand of the combine will be important to plan of the future decision of producers and distributors. It is assumed that the estimated results created by using ARMAX for the period 2009-2020 will be useful for combine producers and distributors.

## References

Abdel-all, R. E. and Al-Garni, A.Z. (1997). Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time series analysis. Energy 22: 1059-1069

Akal, M. (2004) Forecasting Turkey's tourism revenues by ARMAX model. Tourism Management 25: 565-580

AMA, (2014). Automotive Manufacturers Association, Database of agricultural machinery, İstanbul.

Biondi, P., Monarca, D. and Panaro, A. (1998). Simple forecasting models for farm tractor demand in Italy, France and the United States. Journal of Agricultural Engineering Research 71: 25-35

Blanchard, M. and Desrochers, G. (1984). Generation of autocorrelated wind speeds for wind energy conversion system studies. Solar Energy 33: 571–579

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Holden Day, San Francisco

Brown, B. G., Katz, R. W. and Murphy, A. A. (1984). Time series models to simulate and forecast wind speed and wind power. Journal of Applied Meteorology 23: 1184–1195

Dickey, D. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74: 427–431

FAO, (2014). Databases of Food and Agriculture Organisation, Rome. 02/06/2014 http://faostat.fao.org/site/576/DesktopDefault.aspx?PageID=576#ancor

Franses, P. H. (1991). Primary demand for beer in the Netherlands: An application of ARMAX model specification. Journal of Marketing Research 28(2): 240-245

Greene, W. H. (2000). Econometric Analysis. Fourth Edition. Englewood Cliffs, NJ: Prentice Hall

Gujarati, D. N. (1995). Basic Econometrics. Literatür Kitabevi, Thirth Edition, Istanbul

Harvey, A. C. (1990). The econometric analysis of time series. Cambridge, MA: MIT Press.

Ho, S. L. and Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. Computers Industrial Engineering 35 (1–2): 213-216

Ho ,S. L., Xie, M. and Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series predictions. Computers and Industrial Engineering 42: 371-375

Kamal, L. and Jafri, Y. Z. (1997). Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy 61(1): 23–32

Kaygısızel, M. (2006). Mısır Hasadında Tane Kayıpları ile Biçerdöver Marke ve Model İlişkisinin Saptanması, Ç.U. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Adana.

Lim. C., Min, J. C. H. and McAleer, M. (2007). ARMAX modelling of international tourism demand, in Oxley, L. and Kulasiri, D. (eds.), MODSIM 2007 International Congress on Modelling and Simulation, MSSANZ, Christchurch, New Zealand, pp. 1885-1891

MacKinnon, J. G. (1991). Critical Values for Cointegration Tests. Chapter 13 in R. F. Engle and C. W. J. Granger (eds.), Long-run Economic Relationships: Readings in Cointegration, Oxford University Press.

Madan, S. C. (1989). Factor demand in Indian agriculture: an econometric study. Thesis, State University of New York at Binghamton, 147

Pawlak, J. (1999). Impact of some selected factors on the sale of agricultural tractors. *Problemy Inzynierii Rolniczej* 7(4): 89-94

Pindyck, R. S. and Rubinfeld, D. L. (1997). Econometric Models and Economic Forecast, 4th Edition, McGraw-Hill Inc.

Saab, S. Badr, E. and Nasr, G. (2001). Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. *Energy* 26: 1-14.

TSI (2014). Database of Türkiye İstatistik Kurumu [Turkish Statistical Institute], Ankara. 12/06/2014 http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=193

Unakitan, G. and Akdemir, B. (2007). Tractor demand projection in Turkey. *Biosystems Engineering* 97: 19-25

Zhang, G.P. (2001). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50: 159–175