Estimation of Number Of Small Cattle Through ARIMA Models in Turkey

Şenol ÇELİK

Bingöl University Of Faculty Of Agriculture Department Of Animal Science, Bingöl, Turkey.

Abstract: In this study, the number of sheep and goats in Turkey were analysed by time series analysis method, and the number of great cattle for next years predicted through the most appropriate time series model.Time series was formed using the data on the number of sheep and goats belonging to the period between 1930 and 2014 in Turkey It was determined through autocorrelation function graphic that the series weren't stationary at first, but they became stationary after their first difference were calculated. A stagnancy test was performed through extended Dickey-Fuller test. So as to determine the suitability of the model, it was reviewed if autocorrelation and partial autocorrelation graphs were white noise series and also the results of Box-Ljung test were reviwed. Through the tested models, the model estimations, of which parameter estimates were significant and Akaike information criterion (AIC) was the smallest, were performed. The most appropriate model in terms of both the number of sheep and goats is first-level integrated moving average model stated as ARIMA(0,1,1). In this model, it was estimated that there would be an increase in the number of sheep and goats in Turkey between the years of 2015 and 2020, however, the increase in the number of sheep would be more than the increase in the number of goats.

Keywords: ARIMA Models, autocorrelation, the number of sheep, the number of goats.

1. Introduction

Time series are called as a Box-Jenkins or ARIMA model if they contain discrete, linear and stochastic processes. Autoregressive (AR) models, moving average(MA), and autoregressive moving average (ARMA) models, which is a mixed model of AR and MA models were developed respectively by Yule (1927), Slutsky (1927) and Wold (1954) for the first time [1-3]. The models applied to the models which weren't stationary but made stationary as a result of the process of difference calculation, were called as Merged Autoregressive Moving Average Model (ARIMA). This model is also referred to as Box-Jenkins technique.

As well as time series have many application areas in daily life, they also can be applied in livestock production area. As in all countries of the world,

livestock has taken place significiantly in Turkey in terms of meeting the needs of the growing population and being used as industrial raw material in many areas as well. Despite the population increases every year in Turkey, the number of animals doesn't increase every year.

According to the TSI data, Turkey's population in 2014 was 77 695 904 with an increase of 13.3 per thousand comparing to the previous year[4].The number of sheep reaching a maximum number of 49 636 000 in 1982, decreased in subsequent years with bumpy trend resulting in 21 749 508 in 2009. The number of sheep rose again after this period, increasing up to 31 115 190 in 2014.The number of goat reaching the highest number of 25 077 700 in 1959 decreased in subsequent years. The number of goat exhibiting a bumpy trend in the next period resulted in the lowest number of 5 128 285 in 2009. The number of goat rising a little in the next period, increased up to 10 347 159 in 2014 [5,6].

Corresponding author: Şenol ÇELİK, Asst. Prof., Bingöl University of Faculty of Agriculture Department of Animal Science, Bingöl, Turkey. E-mail: senolcelik@bingol.edu.tr.

According to the statistics of FAO in 2013, total number of sheep in the world is 1 162 875 535. The greatest sheep raising have been performed in China in the world with number of 175 000 240. India is in the second rank after China with sheep number of 63 800 000 in the world. Turkey is in the 9 th rank in sheep raising in the world. Total number of goat in the world is 975 803 263. The greatest goat raising have been performed in China in the world with number of 182 870 680 in 2013. This country was respectively followed by India with number of 134 000 000 and Pakistan with number of 64 858 00. Turkey is in 22 nd rank in goat raising in the world [7].

The increase in the number of small cattle also causes an increase in animal products. The amount of sheep milk is 1 113 130 tons, the amount of mutton is 98 978 tons in Turkey in 2014. The amount of goat milk is 463 394 tons, the amount of goat meat is 26 770 tons in Turkey at the same period of time [6].

Turkey has primarily undertaken the target of industrialization as developed countries did since the first years of the Republic, however agriculture and livestock couldn't be benefited in terms of economic development. In this period, agricultural and livestock sectors were the most being influenced sectors by economical crisis and political intervention. During periods of planned development, livestock sector was targeted to be developed in accordance with economical and ecological conditions. However all developments in all planned periods were under the targeted figures [8]. The aim of this study is to determine the most appropriate time series model through the time series analysis of the numbers of small cattle (sheep and goats),and estimating the numbers of great cattle by 2020 through prediction according to the determined model.

2. Literature Review

There are studies performed on agriculture and livestock using time-series analysis. In the studies performed by Gürcan and Cenan (2011), the numbers of great cattle and small cattle in respect of their species were researched by Box Jenkins method for the time period between 1936 and 2005,and accordingly animal production amounts through time-series analysis were predicted for the time period between 2006 and 2015. It was proven by unit root test results that animal number series weren't stationary. When the realized values relating to the time period between 2006 and 2009 was compared to the predicted values obtained by ARIMA models, it was seen that predicted values were calculated higher than the realized values in general. The predictions obtained by 2015 indicates that the average annual increase rate of total animal numbers would be two per thousand. It was predicted that total number of goats would increase at first, then it would decrease. Average annual decrease rate of this series for time period of prediction was calculated as 7 per thousand. The predicted values relating to sheep and goat numbers covering the time period by 2015 indicated that average annual increase rates would be respectively 6 per thousand and 3 per thousand. It was predicted that total number of goats showed a declining tendency constantly, and the average annual decrease rate would be 3 per thousand[9]. Okumuş (2012 applied time-series analysis on cotton prices in Turkey. In the study, time-series analysis was applied to the relation between cotton production, the price of cotton on stock market and price of diesel fuel. According to the findings obtained, it was seen that the production amount of the previous year and the price of cotton were key factors on the production. A strong relationship was determined between diesel price and production by causality analysis. This situation was expressed by increased cotton demand at textile sector because of that an increase in fuel oil price caused the cost of artificial woven products, which is an alternative to cotton in the textile industry, to be increased and accordingly the demand of these products to be decreased [10]. Özer and İlkdoğan

(2013) examined world cotton prices through ARIMA model. In a study performed using world cotton prices for 102 months covering the period of January 2004 to June 2012, ARIMA(1,1,1) $(1,0,1)_{12}$ seasonal model was identified as the most convenient model. According to the prediction model, it was predicted that the average world cotton prices would be 1.49 dollars for the 2012 to 2013 season, 1.57 dollars for the 2013 to 2014 season, 1.55 dollars for the 2014 to 2015 season [11].

3. Methodology and Data

In this research, small cattle raising data (sheep and goat numbers) were obtained through records of Turkey Statistics institution (TSI), (Food and Agriculture Organization of the United Nations, FAO). The population data were obtained from the section of "Mid-Year Population",which took place in the section of "Population Projections" with head line "Population and Demography". These data comprises the period between 1930 and 2006. The population data relating to the period between 2007 and 2014 was provided from "address based population registration system" issued by Turkey Statistical Institute. The data formed by all variables was prepared for statistical analysis by being formed in annual time series for the period between 1930 and 2014.

ARIMA or Box-Jenkins models are a combination of the AR and MA models applied to the series, of which d th level difference were calculated. General representation of these models is ARIMA(p, d, q) and it is shown as [12].

$$\varphi(B)(1-B)^d X_t = \varepsilon_t \qquad (1)$$

The meaning of variables took place in models are as follows:

q: Moving average model level, p: Autoregressive model level, d: Non-seasonal difference calculation level.

$(1-B)^d$: d th difference calculation process. $(1-B)^d X_t$ statement can be stated by (Backshift)

operator as follows [13]:

$$BX_t = X_{t-1} \qquad (2)$$

For stagnancy, Extended Dickey-Fuller (ADF) unit root test developed by Dickey and Fuller (1981) can be applied and it has been stated in the (3) numbered equation.

$$\Delta X_t = \beta_0 + \beta_1 t + \gamma X_{t-1} + \sum_{i=1}^{h} \alpha_i \Delta X_{t-i} + e_t \quad (3)$$

If series is unit rooted in the equation, that is, if it is not stationary, its stagnancy is controlled by application of difference calculation process.

The expression of ARIMA (p, d, q) model can be defined as given in equation (4) [14, 15].

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + ... + \varphi_p X_{t-p}$$
$$+ e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} \qquad (4)$$

The meaning of variables in this equation are as follows:

φ_p : Parameter values for autoregressive operator ,
e_t : Error term coefficients,

θ_q : Parameter values for moving average operator,
X_t : Time series created by d th level difference calculation of original series. When difference calculation level d = 0, original series is stationary. In this case the ARIMA model will turn into AR, MA or ARMA model. Stagnancy means that the expected value and auto covariance of X_t time series are independent of time [16].

In order to establish a model, the stages given below must be realized [17].
• Determination of the general model class,
• Determination of temporary model.
• Estimation of parameters of temporary model,
• Temporary model convenience test
• Using the model for estimation if it is convenient for that.

In determining the general model class, autocorrelation and partial autocorrelation functions are used for the selection of the model [18, 19].

A model from AR, MA, ARMA or ARIMA model class is selected to determine a model by using autocorrelation and partial autocorrelation functions. Then, the model parameters are estimated and standard errors of coefficients are calculated and tested to determine if they are significant. In the last phase, the conveince control is performed to understand if the determined model is convenient for prediction [20]. Autocorrelation function of the temporary model assumed to be convenient is examined by examining the autocorrelation coefficients graph of its errors. If the errors are white noise according to autocorrelation function, it becomes clear that the model is convenient [18]. Autocorrelation and partial autocorrelation functions were examined to determine the model and the properties of theoretical functions of ARIMA models are given in figure 1 [21].

Autocorrelation function is as follows [22] :

$$\rho(t,s) = \frac{\gamma(X_t, X_s)}{\sqrt{\gamma(X_t, X_t)\gamma(X_s, X_s)}} \qquad (5)$$

Partial autocorrelation function is as follows [23] :

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

$$\frac{\rho(h) - \alpha_1\rho(h-1) - \alpha_2\rho(h-2) - ... - \alpha_{h-1}\rho(1)}{1 - \alpha_1\rho(1) - \alpha_2\rho(2) - ... - \alpha_{h-1}\rho(h-1)} \qquad (6)$$

The model convenience test is calculated in equality (7) as Box-Ljung statistics developed by Box and Ljung suggest [12].

$$Q = n(n+2)\sum_{h=1}^{k} \frac{\rho^2(h)}{n-h} \qquad (7)$$

Depending on this statistic, error terms of model are independent from each other and exhibit normal distribution [24]. Here, h indicates the number of delay, p and q are the levels of ARIMA model. n is the number of observation, $\rho^2(h)$ is the autocorrelation coefficient calculated for the delay [25]. For the purpose of selection of the best model, some criterias, such as Akaike Information criteria and Bayesian information criteria (BIC) were developed.

Table 1   Periodic properties of autocorrelation and partial autocorrelation functions

| Model	Autocorrelation function	Partial autocorrelation function
AR(p)	Decreases in the form of an exponential or sine wave.	It gets end after p delay
MA(q)	It gets end after q delay	Decreases in the form of an exponential or sine wave.
ARMA(p,q)	Decreases in the form of an exponential or sine wave, It gets end after p or q delay	Decreases in the form of an exponential or sine wave, It gets end after p or q delay
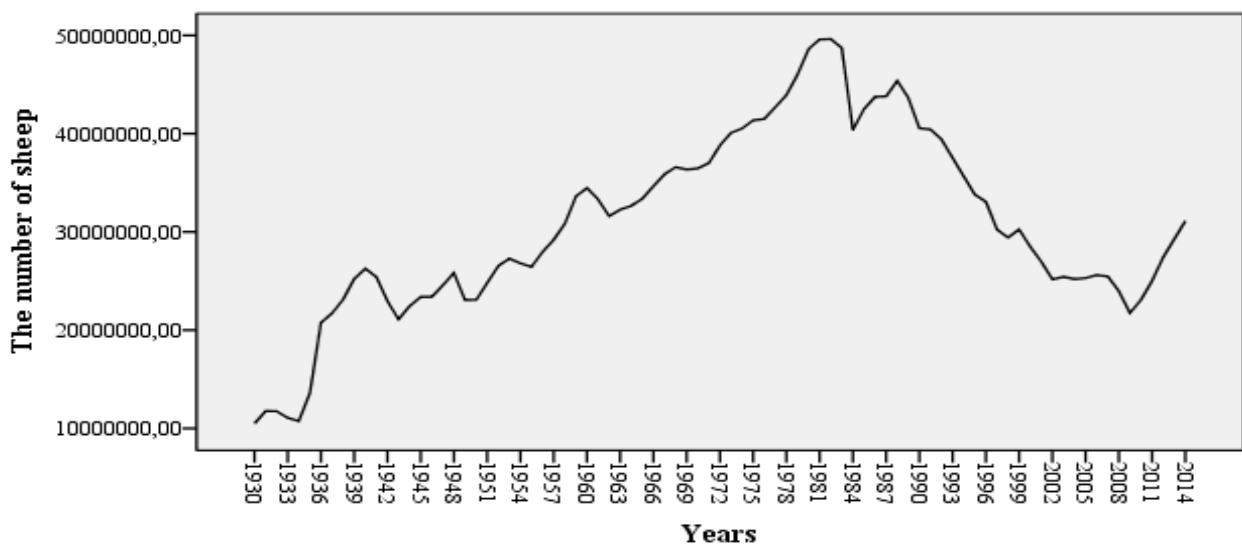
Fig. 1   Sheep number between the years 1930 and 2014 in Turkey.

Akaike Information criteria is defined as follows[26]:

$$AIC = nln\hat{\sigma}_e^2 + 2M \qquad (8)$$

Here, M = p+q+1 and it is the number of parameters in the model. Bayesian information criteria (BIC) is defined as follows [27].

$$BIC = nln\hat{\sigma}_e^2 + Mln(n) \qquad (9)$$

4. Findings and Discussion

The data on the number of sheep and goats for the period between 1930 and 2014 are given respectively in Figure 1 and Figure 2. It is seen in Figure 1 and Figure 2 that the series have a certain trend. For the purpose of seeing the existence of trend, autocorrelation (ACF) and partial autocorrelation (PACF) graphs of the number of sheep in Figure 3 and the number of goats in Figure 4 are given. When these graphs examined It was seen that a trend existed in the series, that is, the series were not stationary. When ACF graphs examined it was seen that many relationship coefficients exceeded the upper confidence limit above zero origin in both Figure 3 and Figure 4. In such cases, it was noted that the series had trends in both ACF and PACF graphs. The difference calculation process is applied to the series

so as to purify the series from trend [18]. The ACF ve PACF graphs of the series, of which first difference were calculated, are given in terms of (number of sheep) in Figure 5 and in terms of (number of goats) in Figure 6.As seen in Figure 5, The coefficient belonging to only first delay in the first difference series of the number of sheep in the defined period exceeded security border. For this reason, the series became stationary. Similarly, As seen in Figure 6, The coefficient belonging to only first delay in the first difference series of the number of goats in the defined period exceeded security border and the series became stationary.

Extended Unit root test was applied in order to determine in more accurate way that the series were stationary. Unit root test results are given in Table 1.When Table 1 examined it was seen that zero hypothesis was accepted indicating that unit root existed in the variables of number of sheep (KOS) and number of goat (KES) and these variables weren't stationary at the values of level. ADF test was applied again in order to determine that KOS and KES variables were stationary in their first differences. According to the results given in Table 1, it was seen that these variables were purified from unit root in the first difference calculation and became stationary I(1) P<0.01)
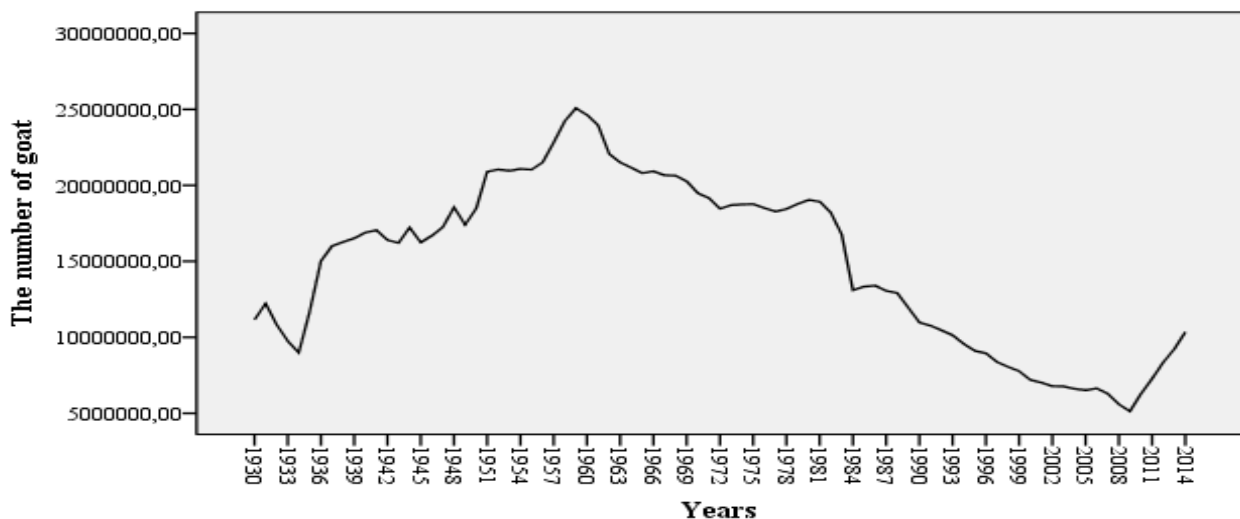

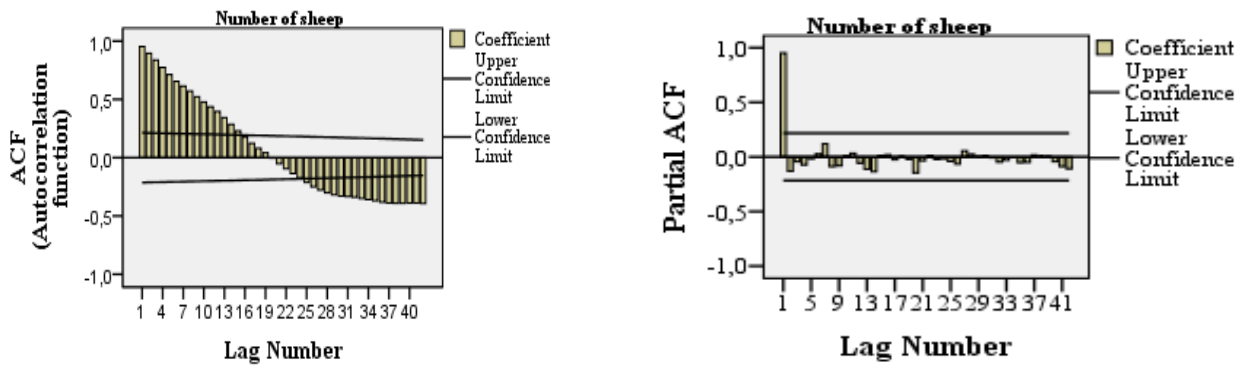Fig. 2   Goat number between the years 1930 and 2014 in Turkey

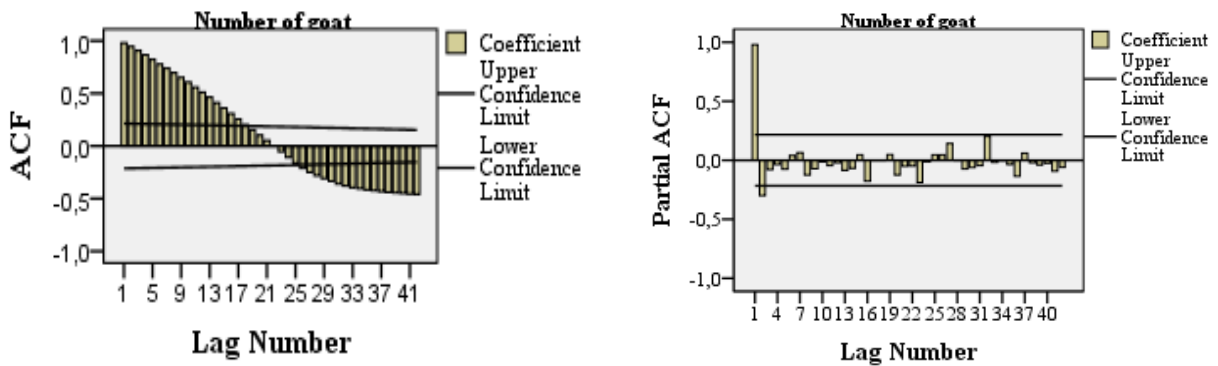**Fig. 3** Autocorrelation (ACF) and partial autocorrelation function (PACF) graphics in terms of number of sheep.

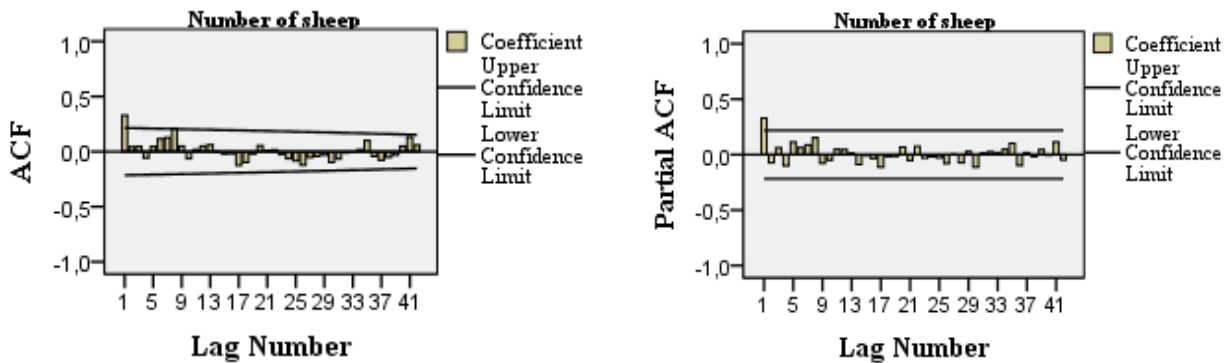**Fig. 4** Autocorrelation (ACF) and partial autocorrelation function (PACF) graphics in terms of number of goats.

Fig. 5 ACF and PACF graphics of first difference series of the number of sheep.
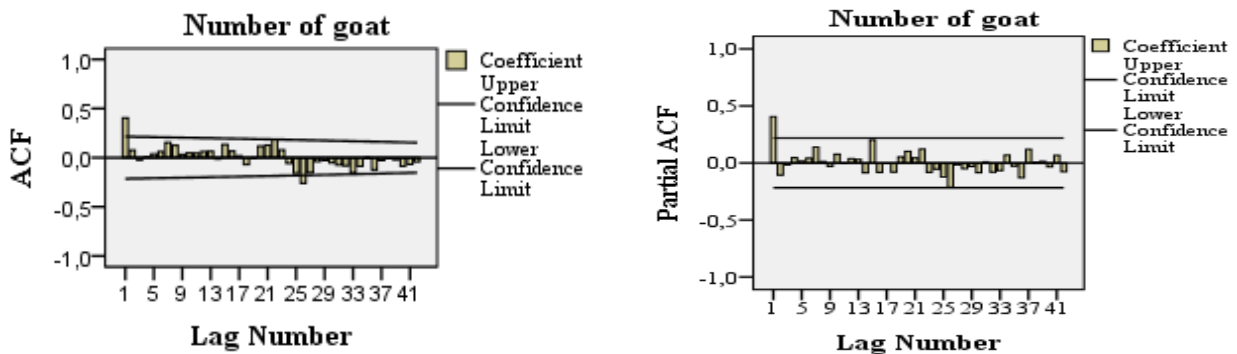

Fig. 6 ACF and PACF graphics of first difference series of the number of goats.

Table 1　ADF Unit Root Tests Of The Series.

variables	ADF Values		Results
	Level	First difference	
KOS	-2.127	-6.353**	I(1)
KES	-1.114	-5.802**	I(1)

Table 2　Estimation Of Parameters.

The number of sheep

Parameter	Coefficient	Standart error	t-statistics	p
Difference	1			
MA (1)	0.372	0.103	3.608	0.001

Number goat

Parameter	Coefficient	Standart error	t-statistics	p
Difference	1			
MA (1)	0.422	0.102	4.152	0.000

(*) sign means that first difference calculation of the series at 5% significance level isn't unit rooted, and (**) sign means that first difference calculation of the series at 1% significance level isn't unit rooted. MacKinnon's critical values at significance levels of 1%, 5% and 10% are respectively –3.511, -2.897 ve –2.586. Here; KOS represents number of sheep, and KES represents the number of goats.

After the series having become stationary, it was proceeded to the model determination phase. It was seen in Table 5 that autocorrelation value for first delay in the first difference series of the number of sheep in ACF and PACF graphs was out of limits and autocorrelation value was within limits for other delay values. Since the value after first delay in ACF graph is zero (0), the number of sheep is first level integrated moving average model, that is, it is ARIMA(0,1,1). Similarly, it was seen that that autocorrelation value for first delay in the first difference series of the number of goats in ACF and PACF graphs was out of limits and autocorrelation value was within limits for other delay values. Since the value after first delay in ACF graph inclined faster to zero (0) than the value in PACF graph, the number of goats is first level integrated moving average model, that is, it is ARIMA(0,1,1).Significance test results of parameter estimates for obtained models is given in Table 2.

When significance of the parameters belonging to both models (ARIMA(0,1,1)) of the number of sheep and the number of goats were tested, it was determined that parameter estimates were determined as statistically significant (P<0.01).While The model of the number of sheep was obtained as $X_t = -1.372e_{t-1} + 0.372e_{t-2} + e_t$ The model of the number of goats was obtained as $X_t = -1.422e_{t-1} + 0.422e_{t-2} + e_t$. According to the Box-Ljung test performed to determine if these models were convenient, it was seen that the model of the number of sheep was convenient since Q=9.115 ve P=0.937>0.05 and the model of the number of goats was convenient since Q=10.796 ve P=0.867>0.05. In addition, alternative models were tested through assigning various values to the model parameters from AIC and BIC statistics and the most convenient model was determined according to the results in Table 3. The most convenient model in terms of the number of sheep and the number of goat was determined as ARIMA(1,1,0) since AIC and BIC values are the smallest values. This model obtained was different from the models of the the number of sheep and the number of goats obtained by Cenan ve Gürcan (2011) for the period between 1936 ad 2005 [9].

Table 3 Model selection criteria.

ARIMA	The number of sheep		The number of goat	
	AIC	BIC	AIC	BIC
(1,1,0)	31.639	31.698	30.212	30.271
(1,1,1)	31.652	31.739	30.207	30.288
(1,1,2)	31.659	31.776	30.199	30.316
(0,1,1)	31.619	31.677	30.205	30.262
(0,1,2)	31.642	31.729	30.218	30.305
(2,1,0)	31.670	31.758	30.176	30.264
(2,1,1)	31.670	31.788	30.199	31.116
(2,1,2)	31.661	31.808	30.149	30.296

The error graph of the series of number of sheep for the most convenient model conforming the data is given in Table 7. Since the errors (excesses) belonging to the number of sheep took place with in the security limits in general in Table 7, it is White Noise series. Since the errors (excesses) belonging to the number of goats took place with in the security limits in general in Table 8, it is White Noise series

The predicted values obtained through ARIMA (0,1,1) model for the series of number of sheep and goats are given in Table 4.It has been estimated according to the prediction done for the period between 2015 and 2020 that the number of sheep would be 31 824 930 in 2015 and increased up to 33 101 441 in 2020.It has been also estimated that the number of goats would be 10 734 984 in 2015 and
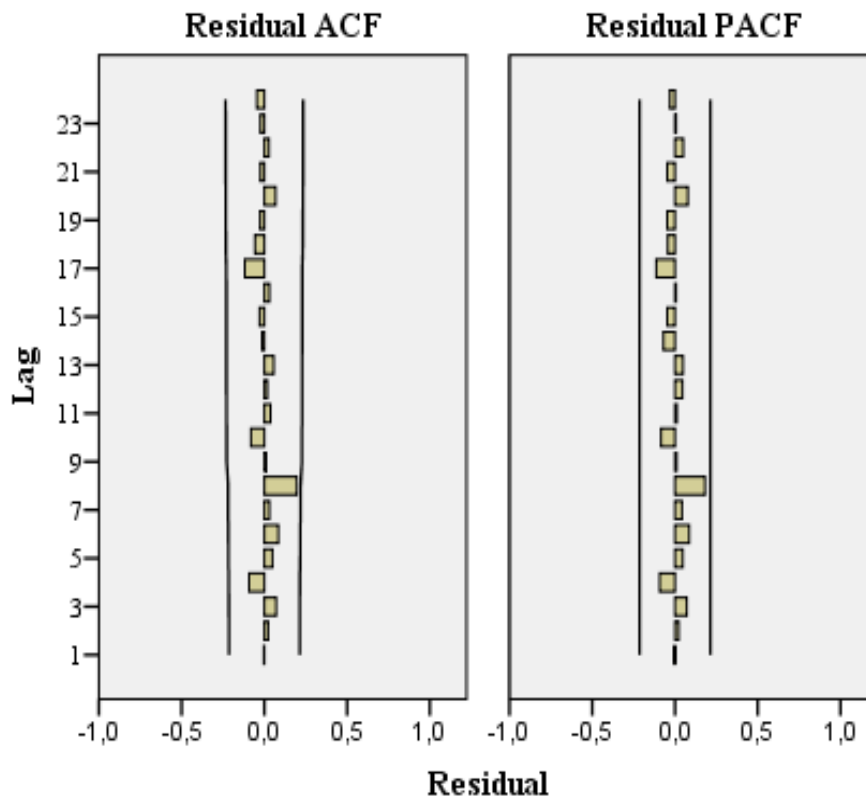

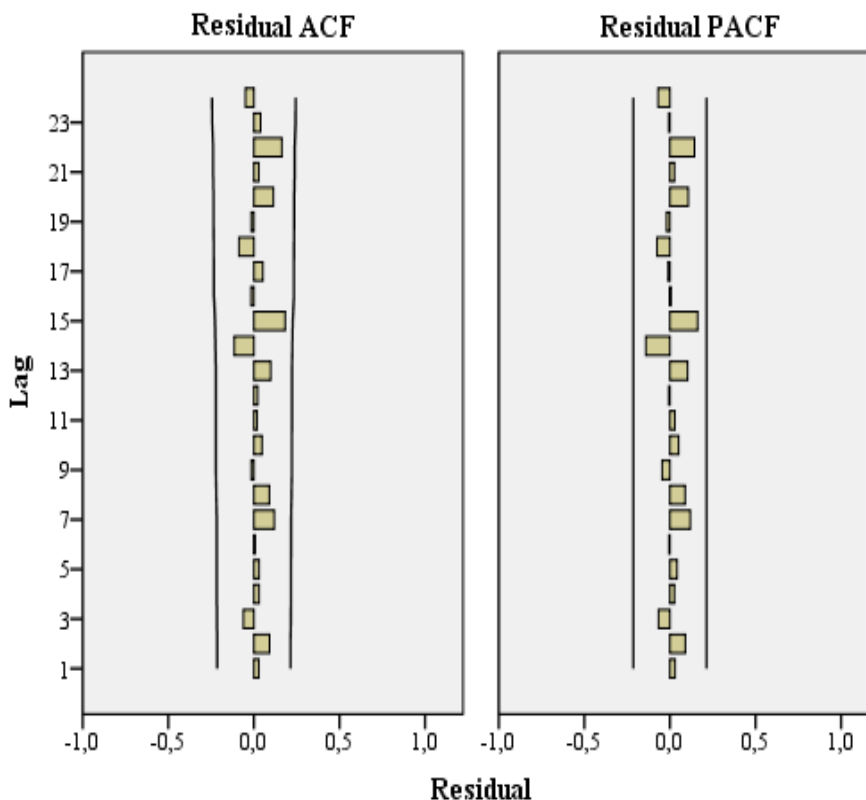**Fig. 7**   **ACF and PACF graphs of error terms of number of sheep.**

Fig. 8   ACF and PACF graphs of error terms of number of goats

Table 4   The predictions of the number of sheep and goats

Years	2015	2016	2017	2018	2019	2020
KOS	31 824 930	32 080 233	32 335 535	32 590 837	32 846 139	33 101 441
KES	10 734 984	10 737 909	10 740 834	10 743 760	10 746 685	10 749 610

KOS: The number of sheep, KES: The number of goats

increased up to 10 749 610 in 2020. The rise in the number of small cattle is an important advantage. So as to keep this advantage, small cattle raising should be continued, animal products that will contribute to the economy of country should be developed.

5. Conclusion

Time series method, ARIMA was used for the estimations for future on the number of sheep and the number of goats for 85 years covering the period between 1930 and 2014 in Turkey. The series aren't stationary but they became stationary after their first differences were calculated. In the analysis performed, the series of both the number of sheep and the number of goats were modeled in the form of ARIMA(0,1,1).

The convenience of models was determined by Box-Ljung test. It was seen that the errors of the series were white noise. It has been estimated through predictions obtained by these models that the number of sheep and goats would increase between the years 2015 and 2020. However, it has been expected that the number of sheep would be higher than the number of goats.

References

[1] Yule, G. U. (1927). On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wölfer's Sunspot Numbers, Phil. Trans., A226, 267.

[2] Slutsky, E. (1927). The Summation of Random Causes As The Source of Cyclic Processes, Problems of Economic Conditions, 3, 1, 1927; (English trans. in

Econometrica, 5, 105,

[3] Wold, H. O. (1954). A Study in The Analysis of Stationary Time Series. Almquist and Wicksell, Uppsala.

[4] TÜİK, (2015). Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları. Türkiye İstatistik Kurumu Haber Bülteni, Sayı: 18616, 28 Ocak 2015, Ankara.

[5] TÜİK, (2014). İstatistik Göstergeler 1923-2013. Türkiye İstatistik Kurumu, Yayın no: 4361, Ankara.

[6] TÜİK, Hayvancılık istatistikleri. http://tuikapp.tuik.gov.tr/hayvancilikapp/hayvancilik.zul, 2014, (Erişim Tarihi: 13.03.2015).

[7] FAO, (2013). Food and Agriculture Organization of the United States. http://faostat3.fao.org/download/Q/QA/E (Erişim tarihi:14.04.2015).

[8] Aral, S., & Cevger, Y. (2000). Türkiye'de Cumhuriyetten günümüze izlenen hayvancılık politikaları. Türkiye -2000 Hayvancılık Kongresi, 38-56, Ankara.

[9] Cenan, N., & Gürcan, İ. S. (2011). Türkiye çiftlik hayvan sayılarının ileriye yönelik projeksiyonu: ARIMA Modellemesi. Veteriner Hekim Derneği Dergisi 82 (1), 35-42.

[10] Okumuş, M. (2012). Tarım Politikaları ve Zaman Serileri Analizi: Türkiye'de Pamuk Fiyatlarına Bir Uygulama. Adnan Menderes Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Aydın.

[11] Özer, O. O., & İlkdoğan, U. (2013). Box-Jenkins Modeli Yardımıyla Dünya Pamuk Fiyatının Tahmini. Tekirdağ Ziraat Fakültesi Dergisi, 10(2), 13-20.

[12] Brockwell P. J., & Davis R. A. (2006). Time Series: Theory and Methods. Springer, New York, 580 p.

[13] Brocklebank, J. C., & Dickey, D. A. (2003). SAS for Forecasting Time Series, SAS Institue Inc., Cary, NC, USA, 398 p.

[14] Janacek, G., Swift, L. (1993). Time Series forecasting, simulation, applications. Ellis Horwood in Mathematics and its Applications, England.

[15] Cryer, J. D. (1986). Time Series Analysis. PWS Publishers, USA, 286 p.

[16] Hamilton, J. D. (1994). Time Series Analysis. Princeton University Pres, 41 William Sr., Princeton, New Jersey, 799 p.

[17] Box, G. E. P., Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control, Holden Day, San Francisco.

[18] Kadılar, C. (2009). SPSS Uygulamalı Zaman Serileri Analizine Giriş. Bizim Büro Yayınevi, Ankara, 300 p.

[19] Akdi, Y. (2010). Zaman Serileri Analizi (Birim Kökler ve Kointegrasyon). Gazi Kitabevi, Ankara, 396 s.

[20] Işığıçok, E. (1993). Değişkenler Arasındaki İlişkilerin Araştırılmasında Nedensellik Testleri ve Bir Uygulama Denemesi. Doktora Tezi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Bursa.

[21] Sevüktekin, M., & Nargeleçekenler, M. (2010). Ekonometrik Zaman Serileri Analizi Eviews Uygulamalı. Nobel Yayın No: 770, 592 p.

[22] Shumway R. H., & Stoffer D. S. (2006). Time Series Analysis and its Applications with R Examples. Springer, New York, 578 p.

[23] Wei, W. W. S. (2006). Time Series Analysis. Addison Wesley Publishing Company, New York, 614 p.

[24] Günay, S., Eğrioğlu, E., & Aladağ, Ç. H. (2007). Tek Değişkenli Zaman Serileri Analizine Giriş. Hacettepe Üniversitesi Yayınları, Ankara, 230 s.

[25] Bowerman B. L., & O'Connell R. T. (1993). Forecasting and Time Series: An Applied Approach. Duxbury Press.

[26] Franses, P. H. (1998). Time Series Models for Business and Economic Forecasting. Cambridge University Press.

[27] Cooray T. M. J. A. (2008). Applied Time Series. Analysis and Forecasting. Narosa Publishing House Pvt. Ltd, New Delhi, 280 p.