Open Veterinary Journal, (2024), Vol. 14(1): 256-265
ISSN: 2226-4485 (Print)
ISSN: 2218-6050 (Online)

Original Research
DOI: *10.5455/OVJ.2024.v14.i1.22*

Submitted: 01/10/2023 Accepted: 15/12/2023 Published: 31/01/2024

The usage of the autoregressive integrated moving average model for forecasting milk production in Egypt (2022–2025)

Mohamed A. Omar, Fardos A. M. Hassan, Sara. E. Shahin and Marwa El-shahat*

Veterinary Economic and Farm Management, Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Egypt

Abstract

Background: Milk is considered one of the most important capital goods and essential sources of animal protein in the diet of the Egyptian family, as well as an effective means to improve the economic condition of farmers, considering this important view, the policymakers need accurate and advance information regarding future supply for planning on the both short and long term.

Aim: The study aims to forecast the production of milk in Egypt during the period from 2022 to 2025 using the Autoregressive Integrated Moving Average (ARIMA) model using time series data of milk production (MP) (1970–2021) obtained from the Central Agency for public mobilization and statistics (CAPMS).

Methods: Augmented Dickey-Fullar Unit Root test, Partial autocorrelation function (PACF), and Autocorrelation function (ACF) of the time series sequence were used to judge the stationarity of the data. After confirming the stationarity of the data, the appropriate ARIMA model was selected based on certain statistical parameters like significant coefficients, values of adjusted *R*-squared, Akaike information criteria (AIC), Schwarz criterion (SC), and Standard Error of Regression. After the selection of the model based on the previous parameters, the verification of the model was employed by checking the residuals of the Correlogram-*Q*-Statistics test.

Results: The most fitted model to predict the future levels of MP in Egypt was ARIMA (1, 1, and 3).

Conclusion: Using the ARIMA (1, 1, 3) model, it could be forecasted that the production of milk in Egypt would show an increasing trend from 6,152.606 thousand tons in 2022 to 6,360.829 thousand tons in 2025.

Keywords: Milk production, ARIMA, Forecasting, Egypt.

Introduction

Animal production projects in general, and dairy production in particular, are considered one of the most important pillars of providing an essential source of animal protein in the diet of Egyptian family, as well as an effective means to improve the economic, environmental, and agricultural conditions of farmers (Arafat *et al*., 2018).

Milk makes a significant contribution to the body's requirements for calcium, magnesium, selenium, riboflavin (B2), cobalamin (B12), and pantothenic acid (B5) vitamins. As a result, milk is the closest thing to ideal since it contains the majority of the elements that support growth as well as the energy-giving components of fat, protein, and carbohydrates that meet the majority of the body's nutritional needs (Mahrous, 2016).

After Asian countries (India, Pakistan, and China), Egypt was the fourth-largest producer of buffalo milk (2.6 million tonnes) and the 37th-largest producer of cow milk (3.1 million tonnes) in the world in 2011–2012 (FAO, 2011, 2012).

The amount of milk produced increased by 1% to 5,226 thousand tonnes in 2019 from 5,174 thousand tonnes in 2018 and by 6.7% to 5,578 thousand tonnes in 2020 from 2019 levels (CAPMS, 2020).

There are many Egyptian dairy products in the list of the most important Egyptian food commodities exported to world markets. Where, Egyptian export for dairy products contributed a ratio of about 7.7% of the Egyptian agricultural and food exports in the period from 2013 to 2017 (FAO, 2018).

The exports of dairy commodities in Egypt, especially cheese and its types, were around 1.1% of the value of world exports, classified 18th among the world's exporting States (CAPMS, 2016).

At the current time, Egypt has self-sufficiency in the dairy industry; however, this may not continue in the future. The people are expanding steadily, and there is an extraordinarily high demand for milk and milk products. It is important to predict how much milk will be produced in the future so that the right policy changes may be made to meet the rising demand.

***Corresponding Author:** Marwa El-shahat. Veterinary Economic and Farm Management, Department of Animal Wealth Development, Faculty of Veterinary Medicine, Zagazig University, Zagazig, Egypt. Email: *marwaalshahat17@gmail.com*

There are multiple forecasting approaches that can be used as exponential smoothing methods such as simultaneous equation regression models, vector autoregression (VAR), and autoregressive integrated moving average models (ARIMA). From these techniques, ARIMA and VAR are considered more appropriate for accurate and precise forecasting (Chaudhari and Tingre, 2013).

The ARIMA model is considered one of the most common time series models and an incredibly precise model for short-term prediction. While some time series are made up of a variety of time-dependent random variables, the fundamental assumption of the model is that the overall variations in the time series follow specific patterns that may be roughly predicted by the associated mathematical model. By analyzing the mathematical model, it is likely to comprehend the structure and characters of time series and to make predictions with the lowest possible variance (Lihua Ma *et al*., 2018).

The study aims to forecast the future levels of milk production (MP) in Egypt to assist decision-makers in developing strategies that will enhance the development of the dairy industry as well as, assess the precision of the ARIMA model in analyzing and forecasting time series MP data.

Material and Methods

Source of data

Time series MP data (1970–2021) were obtained from CAPMS (Table 1) to forecast levels of MP in Egypt from the year 2022 to 2025 using the ARIMA model.

Model description

In econometrics, the ARIMA model is considered one of the most accurate models for time series analysis (Lihua Ma *et al*., 2018). ARIMA model designed

Table 1. Data of MP in Egypt from 1970–2021 according to CAPMS.

Year	Amount of milk produced (thousand tons)	Year	Amount of milk produced (thousand tons)
1970	1,583	1996	2,757
1971	1,635	1997	3,329
1972	1,670	1998	3,490
1973	1,705	1999	3,732
1974	1,738	2000	3,824
1975	1,770	2001	3,954
1976	1,800	2002	4,210
1977	1,828	2003	5,280
1978	1,855	2004	4,682
1979	1,881	2005	5,551
1980	1,905	2006	5,787
1981	1,927	2007	5,925
1982	1,847	2008	5,980
1983	2,015	2009	5,624
1984	2,064	2010	5,774
1985	2,087	2011	5,803
1986	2,210	2012	5,849
1987	2,131	2013	5,554
1988	2,151	2014	5,601
1989	2,178	2015	5,245
1990	2,154	2016	5,088
1991	2,483	2017	5,395
1992	2,068	2018	5,174
1993	1,746	2019	5,226
1994	2,588	2020	5,578
1995	2,693	2021	6,165

in 1970 by Box and Jenkins, was frequently used to analyze time series data to identify trends and predict future values (Chaudhari and Tingre, 2015).

The (ARIMA) models were frequently utilized in several studies. Sankar and Prabakaran, (2012) forecasted the production of milk in TamilNadu by Autoregressive (AR), moving average (MA), and (ARIMA) methods and discovered that ARIMA (1, 1, 0) was the best ARIMA model to forecast MP and predicted that the level of MP would rise from 5.96 million tonnes in 2008 to 7.15 million tonnes in 2015. Hossain and Hassan, (2013) used cubic and linear models to forecast egg, milk, and meat production in Bangladesh and found that the linear model is better for egg production while the cubic model is best for milk and meat. The analysis found that if the growth rates continue, the production of milk, meat, and eggs would be 4.55, 3.77, and 7,544.67 million tonnes, respectively, in the year 2015–2016, the chosen model was used for the following four years' forecast with a 95% confidence interval.

Mishra *et al*. (2020) concluded that ARIMA (0, 2, 1) and ARIMA (0, 1, 1) were respectively the most fitted models for predicting amounts of MP in India and Chhattisgarh. According to their research, milk output in India and Chhattisgarh is predicted to increase from 219.73 MMT in 2022 to 1.599 MMT in 2023.

The ARIMA model consists of three parts (*p*, *d*, *q*), *p* means the number of autoregressive terms (AR), *d* is the number of times the series has to be differenced before it becomes stationary (I), and q the number of moving average terms (MA) (Deshmukh and Paramasivam, 2016).

AR process of order (*p*) is represented as follows:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t;$$

The dependent variable (i.e., the variable of interest) is forecasted based on the background of future values using an AR time series model.

(*I*) means the integration demonstrating the order of a single integer, the number of differences is the order of a single integer (*d*) which is represented as follows:

$$Y'_t = Y_t - Y_{t-1}$$

where y_{t-1} and yt are lagged original series and original series, respectively.

Moving Average process of order (*q*) is represented as follows:

$$Y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t;$$

MA estimates approaching observations of the dependent variable by including historical data from the white noise process (i.e., historical forecast errors).

The common form of the ARIMA model of order (*p*, *d*, *q*) is represented as follows:

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \ldots + \theta_P Y_{t-P} \varepsilon_{t-1} - \theta_q \varepsilon_{t-2} + \varepsilon_t$$

Where,

Y_t is milk production, ε_t are independently and normally distributed with zero mean and constant varience for t = 1, 2,….., n; and $\emptyset_p$ and θ_q are also estimated.

The standard ARIMA model technique consists of four phases, including:

I- Identification of model
II- Parameter estimation (OLS)
III-Diagnostic checking
IV- Forecasting

I-Identification of model

The stationarity of the sequence must be checked before using the ARIMA model because nonstationary sequences cannot be predicted. The time series sequence's scatter plot, line graph, partial autocorrelation function (PACF), and autocorrelation function (ACF) graphs are used for determining the sequence's stationarity. The unit root of augmented Dickey-Fuller (ADF) is frequently used to examine the difference, trend, and seasonal variation and recognize the stationarity (Lihua Ma *et al*., 2018).

If the time series is stationary, there is no need to take a single integer (D), and on this occasion, we can deal with the ARMA model (*p*, *q*). However, if the time series is not stationary, the proper conversions should be applied, such as logarithm and difference. To determine the values of the autocorrelation order p and the moving average order q of the ARIMA model, the sequence's autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) are estimated. To find the most precise model, assume different values of p and q to obtain multiple ARIMA models.

II-Parameter estimation

Using the least squares method (OLS) on EViews software, after determining the appropriate orders of *p*, *d*, *q*, we obtain multiple ARIMA models and select the most fitted model among them based on certain parameters such as significant coefficients, Akaike information criteria (AIC), adjusted *R*-squared, SC and Standard Error of regression (S.E. of regression). The most appropriate model for forecasting should have the most significant coefficients, the greatest adjusted R-squared value, and the lowest AIC, SC, and regression S.E. values.

III-Diagnostic checking

The selected model from Step II is analyzed and adjusted to test its accuracy by performing a Correlogram-Q-Statistics test on the residual. If the autocorrelations and partial autocorrelations of residuals at all lag nearly zero and all *Q*-Statistics are insignificant at a significance level of 0.05 this means that there is no serial correlation in the residuals and subsequent validity of this model for predicting future time series values.

IV-Forecasting

Select the Forecast menu of the fitted model in the EViews software's Equation window. In the dialog box, Dynamic or Static can be selected then click OK to obtain the predicted value (Lihua Ma *et al*., 2018). The primary metric of an ARIMA model's realization depends on how well it forecasts both inside and outside of the sample period (Omar *et al*., 2022).

Results

Model identification

Stationarity checking

The MP data series during 1970–2021 is plotted in Figure 1.

The probability of ADF statistic = 0.9363 is greater than the threshold of 0.05, 0.01, and 0.1 significance levels (Table 2)

Table 3 shows that the probability of ADF statistic of logarithm of milk production (LMP) is still higher than 0.05 for that, the null hypothesis cannot be rejected and the LMP series still has a unit rate.

Identification

In Figure 2, with the EViews software, a correlogram test for LMP at the first-order difference is performed to get the graphs of the ACF and PACF of D (LMP) to calculate p and q values (ARMA terms), and as

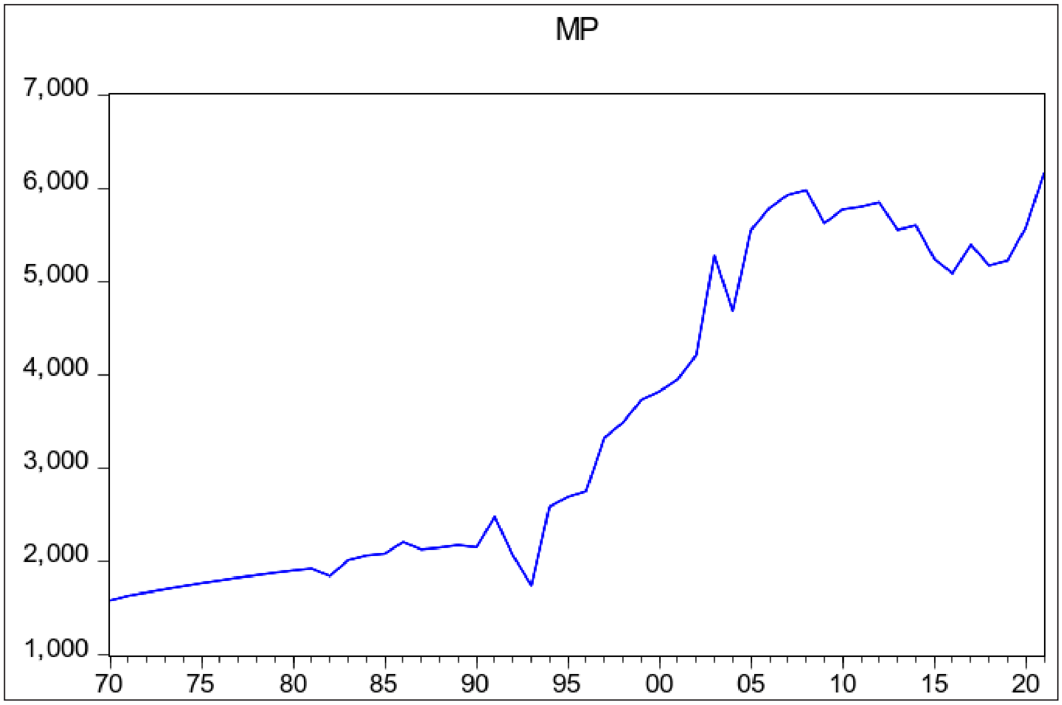

Fig. 1. The MP data from 1970 to 2021.

Table 2. ADF test on MP.

		t-Statistic	Prob.*
ADF test statistic		−0.162970	0.9363
Test critical values:	1% level	−3.565430	
	5% level	−2.919952	
	10% level	−2.597905	

Table 3. ADF test on LMP.

		t-Statistic	Prob.*
ADF test statistic		−0.624859	0.8556
Test critical values:	1% level	−3.565430	
	5% level	−2.919952	
	10% level	−2.597905	

shown bands of partial correlation and Autocorrelation dropped to be nearly zero after lag 3. Therefore, *p* and *q* values can be supposed from 0 to 3.

The result of the ADF test for DLMP is lower than 0.05 (Table 4).

Estimation

Table 5 demonstrates the consequences of the ARIMA test for multiple p and q parameters and d equal one.

Diagnostic checking

The model's validation is concerned with examining the residuals of the selected model from the estimation step to certify the adequacy of the model for forecasting, this checking occurred through the Correlogram-*Q*-Statistics test and as seen in Figure 3, all *Q*-Statistics insignificant at a significance level of 0.05, and all bands of autocorrelation and partial correlation lay within confidence level which indicates that there is no serial correlation between residuals and validity of ARIMA (1, 1, 3) model.

Forecasting

The fitted ARIMA model is used to perform two types of forecasts: Forecasting inside the sample period and Forecasting outside the sample period.

Discussion

The outcome of the Augmented Dickey-Fuller Unit Root test (ADF) on MP showed that the MP Sequence

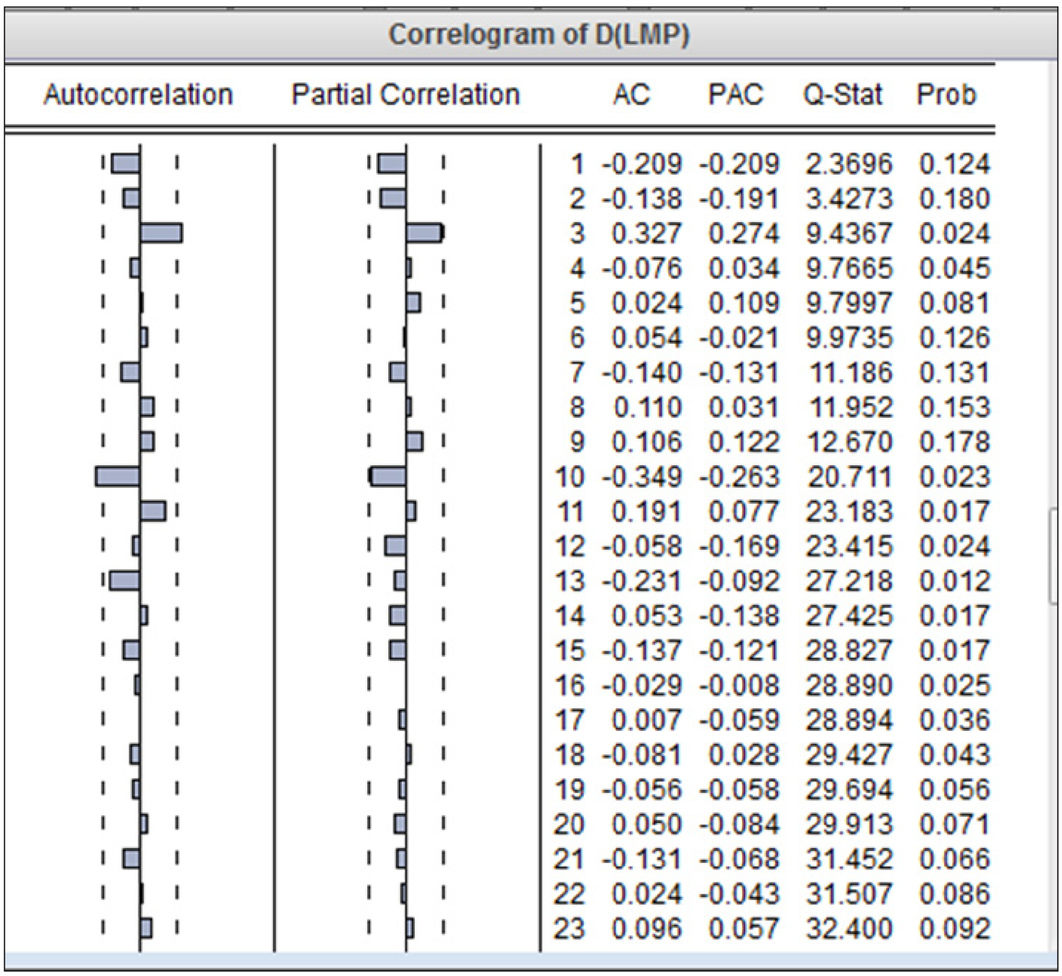

Fig. 2. Graphs of ACF and PACF of the D (LMP) series.

Table 4. ADF test on D (LMP).

		t-Statistic	Prob.*
ADF test statistic		−8.533243	0.0000
Test critical values:	1% level	−3.568308	
	5% level	−2.921175	
	10% level	−2.598551	

Table 5. Results of the ARIMA (*p, d, q*) test.

(*p*, *d*, *q*)	Adjusted *R*-squared	Akaike info criterion (AIC)	SC	Standard Error of regression (SE of regression)
(0,1,1)	0.014957	−1.949727	−1.836091	0.088666
(0,1,2)	0.922832	−0.956087	−0.805991	0.137661
(0,1,3)	0.938467	−1.156563	−1.006467	0.122926
(1,1,0)	0.962565	−1.703596	−1.553500	0.095880
(1,1,1)	0.965125	−1.736689	−1.549070	0.092544
(1,1,2)	0.962109	−1.671953	−1.484333	0.096462
(1,1,3)*	**0.967414**	**−1.813862**	**−1.626242**	**0.089455**
(2,1,0)	0.948939	−1.361786	−1.211690	0.111979
(2,1,2)	0.956171	−1.468623	−1.281003	0.103746
(2,1,3)	0.955639	−1.481322	−1.293702	0.104374

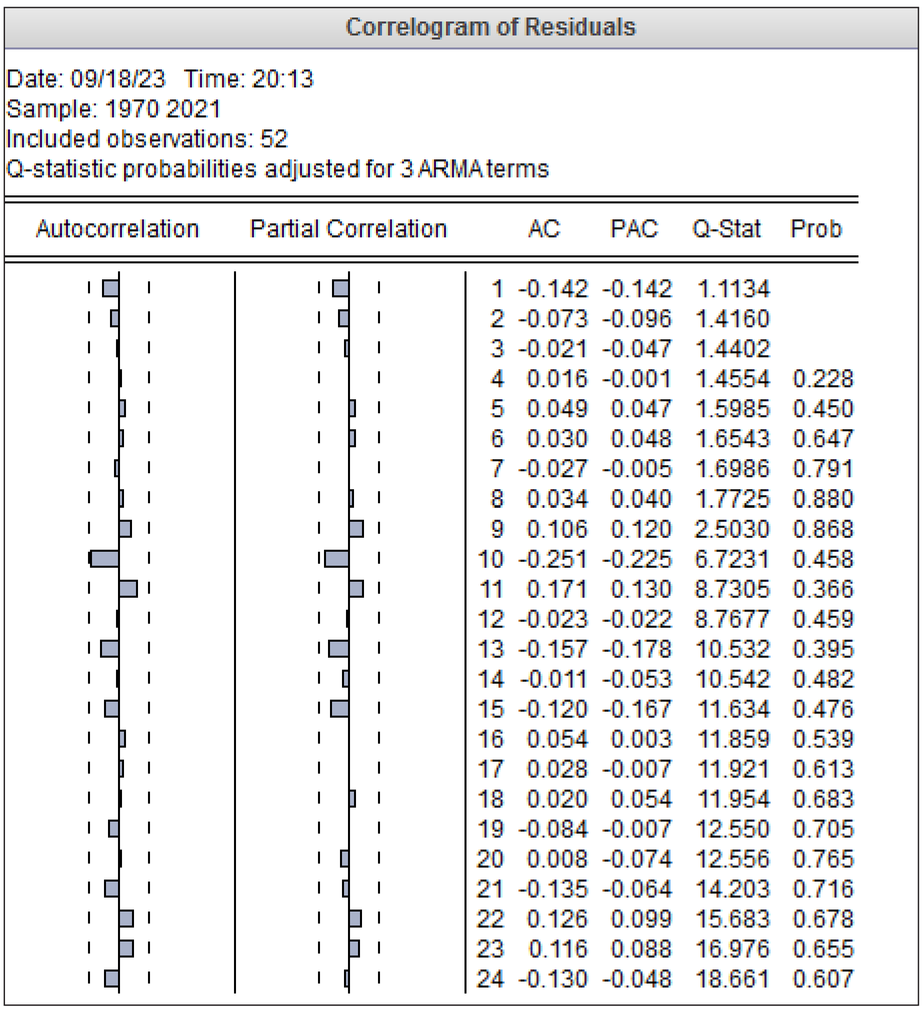

Fig. 3. Graphs of the ACF and PACF of the residual series.

Table 6. The ARIMA (1, 1, 3) model's estimation result.

Dependent Variable: LMP
Method: ARMA Maximum Likelihood (OPG–BHHH)
Date 18/09/23 Time: 14:59
Sample: 1970 2021
Included observations: 52
Convergence achieved after 40 iterations
Coefficient covariance computed using the outer product of gradients

Variable	**Coefficient**	**Std. Error**	***t*-Statistic**	**Prob.**
C	8.056412	0.553249	14.56199	0.0000
D	−0.032975	0.132797	−0.248313	0.8050
AR(1)	0.985211	0.053886	18.28317	0.0000
SIGMASQ	0.007233	0.001111	6.512216	0.0000
R-squared	0.969970	Mean dependent var		8.043565
Adjusted R-squared	0.967414	S.S. dependent var		−1.813862
S.E. of regression	0.089455	AIC		−1.813862
Sum squared resid	0.376102	SC		−1.626242
Log likelihood	52.16041	Hannan−Quinn criterion		−1.741933
F-statistic	379.5251	Durbin−Watson stat		2.109495
Prob. (*F*-statistic)	0.000000			
Inverted AR Roots	.99			
Inverted MA Roots	.37+.64i	.37−.64i	−.74	

is nonstationary and the result of the ADF test proved that, the probability of ADF statistic = 0.9363 is greater than the threshold of 0.05, 0.01, and 0.1 significance levels (Table 2), so the original sequence is exponential and to eliminate nonstationary pattern of sequence we take natural logarithm of the MP data to obtain LMP sequence and repeat ADF test on LMP sequence.

Furthermore, the first-order difference is performed on LMP to obtain the D (LMP) sequence which is exposed to ADF and the result of the ADF test for DLMP is lower than 0.05 (Table 4) so the D (LMP) sequence becomes stationary after the logarithmic change and the first-order difference.

Notably, the most appropriate model to forecast the MP series was ARIMA (1, 1, 3) because this model had the greatest adjusted *R*-squared value and the lowest AIC, SC, and S.E. of regression values, and all statistical parameters of the model were statistically significant at a significance level of 5% as shown in (Table 6). In contrast, the findings obtained by Sankar and Prabkaran (2012), Chaudhari and Tingre (2013), and Lihua Ma *et al*. (2018) where all of them found that the most fitted model was ARIMA (1, 1, 0).

In addition, Figure 4 reports the inverse roots of AR and MA characteristics, and as shown all AR and MA roots have a modulus of less than one and are found within the unit circle. This concerned another proof of the suitability and stability of the selected model.

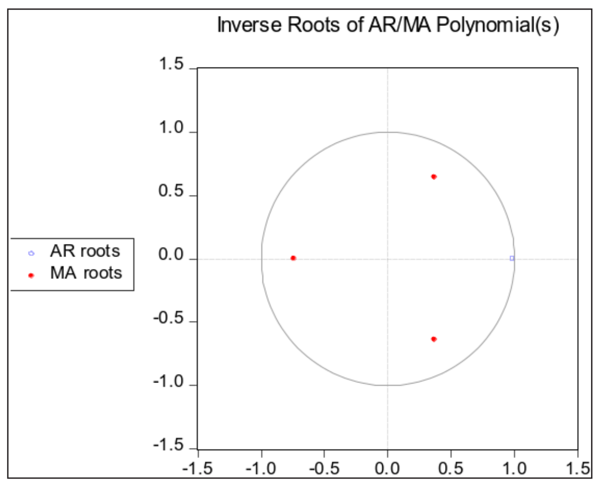

Fig. 4. The estimated inverse roots of AR and MA are stable (stationary).

Finally, an ARIMA (1, 1, 3) model was used to fit the DLMP data, and the results are presented in Figure 5. The lower blue line represents the residuals, The red line represents the actual data and the green line represents the fitted data.

The first is used to boost confidence in the model, and the latter is used to provide accurate forecasts for use in planning and other applications (Chaudhari and Tingre, 2015).

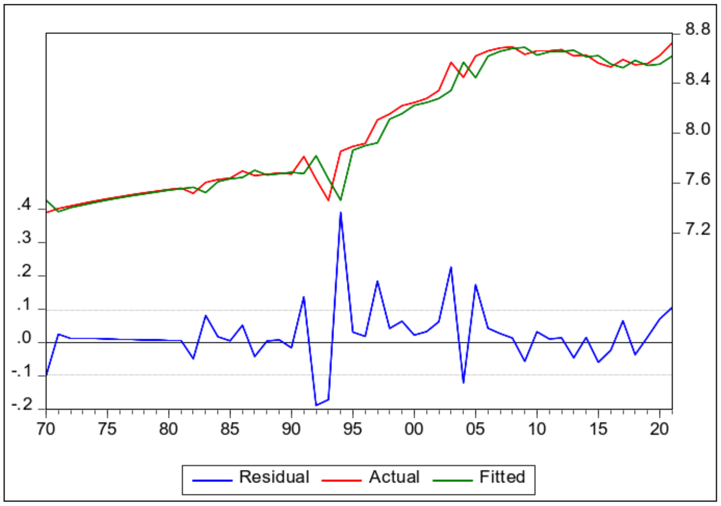

Fig. 5. Residual series, actual series, and fitted series of DLMP sequence.

Table 7. Forecasting during the sample period.

Year	Actual values	Fitted values	Year	Actual values	Fitted values
1970	1,583		1996	2,757	2,497.514
1971	1,635	1,598.482	1997	3,329	3,134.977
1972	1,670	1,646.389	1998	3,490	3,461.411
1973	1,705	1,679.714	1999	3,732	3,604.627
1974	1,738	1,737.203	2000	3,824	3,790.407
1975	1,770	1,764.444	2001	3,954	3,806.63
1976	1,800	1,796.583	2002	4,210	3,977.407
1977	1,828	1,816.509	2003	5,280	4,183.406
1978	1,855	1,846.227	2004	4,682	5,260.239
1979	1,881	1,872.071	2005	5,551	4,747.365
1980	1,905	1,901.236	2006	5,787	5,993.456
1981	1,927	1,923.896	2007	5,925	5,430.356
1982	1,847	1,945.774	2008	5,980	6,212.744
1983	2,015	1,867.383	2009	5,624	5,808.086
1984	2,064	2,027.496	2010	5,774	5,760.783
1985	2,087	2,032.358	2011	5,803	5,612.577
1986	2,210	2,164.883	2012	5,849	5,655.606
1987	2,131	2,234.147	2013	5,554	5,780.262
1988	2,151	2,168.097	2014	5,601	5,573.788
1989	2,178	2,182.027	2015	5,245	5,615.968
1990	2,154	2,148.759	2016	5,088	5,121.02
1991	2,483	2,160.997	2017	5,395	5,060.948
1992	2,068	2,480.249	2018	5,174	5,191.239
1993	1,746	2,092.124	2019	5,226	5,118.129
1994	2,588	1,877.905	2020	5,578	5,313.781
1995	2,693	2,391.874	2021	6,165	5,502.495

Table 8. Forecasting outside sample period.

Year	Forecast
2022	6,152.606
2023	6,208.373
2024	6,428.162
2025	6,360.829

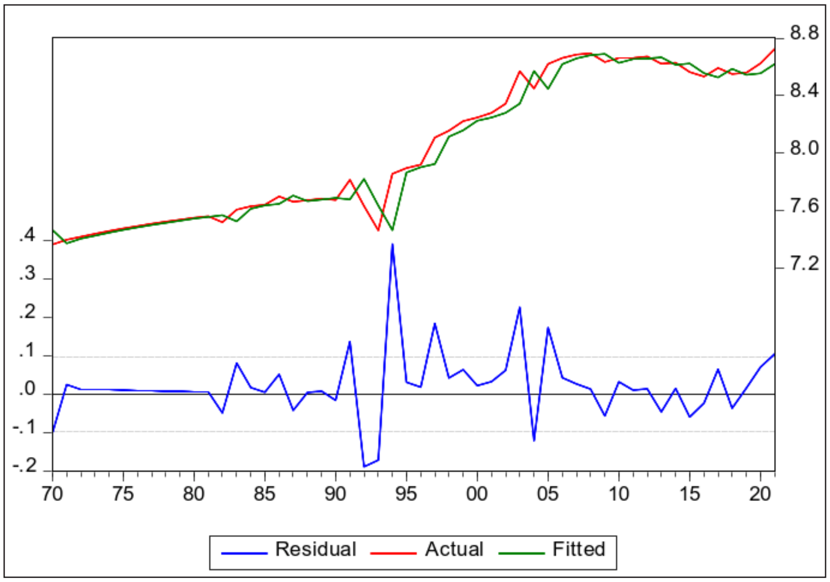

Fig. 6. The forecasting results (logarithmic values) of the MP from the year 2022 to 2025.

Forecasting inside sample period: this type of forecast is obtained simply through plotting the actual series and fitted series of MP sequence as shown in Table 7.

Forecasting outside sample period: based on the fitted ARIMA (1, 1, 3) model, the production of milk in Egypt was forecasted for the years 2022–2025, and the results are shown in Table 8 and Figure 6.

Conclusion

The ARIMA (1, 1, 3) was found to be the most reliable and appropriate model to forecast future levels of MP in Egypt in the period from the year 2022 to 2025, and the results of forecasting indicated that MP would show an increasing trend from 6,152.606 thousand tons In 2022 to 6,360.829 thousand tons in 2025.

Acknowledgments

The authors are grateful to the CAPMS in Egypt for providing the data that helped to complete this study.

Authors contributions

This research was conceptualized by Mohamed A. Omar. All authors participated in the study. All authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

This research did not receive any specific funding.

Data availability

The manuscript contains the data necessary to understand the conclusions of the study.

References

Arafat, M.A., Salim. S.E.A. and El-Sayed, M. 2018. Economics of dairy production in arid regions. Arab Universities J. Agri. Sci., 26(2B), 1275–1288.

CAPMS. 2016. Central Agency for public mobilization and statistics: Bulletin of industrial statistics, for fiscal year 2015/2016, Issue No. (107), Ref. No. 71-01111-2016. Available via https://www.capmas.gov.eg/Pages/Publications.aspx?page_id=5104&Year=23565

CAPMS. 2020. Central Agency for public mobilization and statistics, statistical year book: Annual Bulletin of Statistics Livestock, Ref. No.71-22111-2020. Available via https://www.capmas.gov.eg/Pages/Publications.aspx?page_id=5104&Year=23466.

Chaudhari, D.J. and Tingre, A.S. 2013. Forecasting of MP in India: an application of ARIMA model. Indian J. Dairy Sci., 66, 72–78.

Chaudhari, D.J. and Tingre, A.S. 2015. Forecasting eggs production in India. Indian J. Anim. Res. 49(3), 367–372.

Deshmukh, S.S. and Paramasivam, R. 2016. Forecasting of milk production in India with ARIMA and VAR time series models. Asian J. Dairy Food Res., 35(1), 17–22.

FAO. 2011. Food and Agriculture Organization of the United Nations: FAOSTAT Statistics Database. Available via https://www.fao.org/3/i3107e/i3107e04.pdf.

FAO. 2012. Food and Agriculture Organization of the United Nations: FAOSTAT Statistics Database. Available via https://www.fao.org/3/i2490e/i2490e00.htm.

FAO. 2018. Food and Agriculture Organization of the United Nations: Report of the Organization for economic cooperation and development (OECD) in cooperation with the food and Agriculture Organization. Available via https://doi.org/10.1787/agr_outlook-2018-en.

Hossain, J. and Hassan, F. 2013. Forecasting of milk, meat and egg production in Bangladesh. Res. J. Anim. Vet. Fish. Sci. 1, 7–13.

Lihua, M., Chao, H., Rongchao, L. and Yanben, H. 2018. ARIMA model forecast based on EViews software. IOP Conference Series: Earth and Environmental Science, IOP Publishing. 208, 1755–1315.

Mahrous, H.M. 2016. Milk production and its role in food security in Egypt. Inter. J. Pharm. Tech. Res. 9(12), 329–340.

Mishra, P., Faith, C., Vani, G.K., Tiwari, S., Ramesh, D. and Dubey, A. 2020. Time series investigation of milk production in major states of India using ARIMA modeling. J. Anim. Res. 10(1), 77–84.

Omar, M.A., Shahin, S.E. and Roshdy, M. 2022. Forecasting Inflation in Egypt (2019–2022) by using auto regressive integrated moving average (ARIMA) Models. J. Adv. Vet. Res. 12, 670–676.

Sankar, T.J. and Prabakaran, R. 2012. Forecasting milk production in Tamil Nadu. Inter. Multidiscip. Res. J. 2, 10–15.