Chapter

Predicting Trends, Seasonal Effects, and Future Yields in Cow's Milk through Time Series Analysis

Birhan Ambachew Taye, Alemayehu Amsalu Alen, Ashenafi Kalayu Nega and Bantie Getnet Yirsaw

Abstract

A dairy is a place that is used for handling milk and milk products. Dairy products are basically based on milk. Milk is used to prepare dairy products, such as butter, cheese, and milk powder. There is always a great demand for milk and milk products among people. This study attempted to investigate the trends in the actual yield of cow's milk production at Andassa dairy farm. We used secondary data for the study of the daily milk production of cows at Andassa dairy farm. The specific objectives of the study were—to identify whether the milk production is time-dependent or not; to predict in which season the milk production is high or low; to examine the daily trend analysis of milk production; to fit the appropriate model; and to forecast the milk production for the future. The study was conducted based on quantitative variables. So, the dependent variable is the average daily milk, and the independent variable is the time measure at which milk production is measured each day. The study used both descriptive and inferential statistics to analyze the data that were collected from the dairy farms in the sector. This study covered a total of 179 days of milk production. The results reveal that the milk yield of cows is declining, and that milk output is time-dependent, according to the time series plot, and that the model is ARIMA.

Keywords: milk production, time series analysis, forecasting

1. Introduction

A dairy is a place that is used for handling milk and milk products. Dairy products are basically based on milk. Milk is used to prepare dairy products, such as butter, cheese, and milk powder. There is always a great demand for milk and milk products among people. Most of the time, milk is used as a complete food for infants. It is used in all homes, hotels, and restaurants as well as in milk products. Most countries are expanding their production systems to increase the supply of milk and fulfill the needs of the people [1].

Unless it is produced and handled under sanitary circumstances, milk is an essential route for the spread of harmful germs to humans. As a result, sanitary milk production must be prioritized in order to give more high-quality milk to the general population. Consumers need clean, healthy, and nutritious food, which has been produced and processed in a safe, sanitary manner, and is free of pathogens [2]. As a result, premium milk production is required to meet consumer demand. Milk that is free of pathogenic bacteria and hazardous poisonous compounds, free of silt and extraneous elements, of good flavor, of normal composition, adequate in maintaining quality, and low in a bacterial count is considered to be of high quality; on the other hand, the superiority of milk has persisted deprived [3]. A number of technological innovations now permit automated daily monitoring of cow performance [4]. The use of automated milk yield recording systems for early detection of diseases requires a statistical model to forecast expected performance and actual performance. Previous research on modeling milk production in cows has focused on fitting linear or nonlinear deterministic models to daily, weekly, or monthly milk measurements. Milk production is an integral part of the Andassa agricultural farming system. Even though the area has potential for milk and dairy products, little is known about the existing dairy production system, constraints, and opportunities associated with dairying in this area. It is essential that researchers and dairy development agents understand the existing situations in order to design relevant development strategies for the specific regions [5]. Therefore, the main objective of the study was to predict trends, seasonal effects, and future yields in cow's milk through time series analysis.

2. Data and methods

This study was conducted at Andassa dairy farm, which is located in the Amhara region in northwestern Ethiopia. At Andassa dairy farm, milk production was recorded in liters per cow on a daily basis. This study employed only secondary data obtained from this dairy sector.

2.1 Statistical method

A time series is a set of ordered observations of quantitative variables taken at successive points in time. In other words, it is a set of observations recorded over time, which is usually at equal intervals. Stationary is a critical assumption in time series models. Stationary implies homogeneity in the sense that the series behaves in a similar way regardless of time, which means that its statistical properties do not change over time. Trend analysis is the characteristics of a time series that extends consistently throughout the entire period of time under consideration. In this scenario, trend analysis was used to anticipate the future amount of milk products based on the historical trend of milk production. In this case, we will look at a linear trend and estimate it using the least square estimation method, double moving average, and double exponential smoothing [6].

2.2 Autocorrelation function and partial autocorrelation function

The two moments of any random variable, namely its mean and variance, are well known, and since we established in the introduction that a time series is a realization of a stochastic process, this holds true for any time series. In Box–Jenkins model, the partial autocorrelation plot or partial correlogram is also often employed for model identification [7].

3. ARMA model

The ARMA model is a mixed model in which the series is partly autoregressive and partly moving average. As a result, we get a very generic time series model, as shown below.

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2+\dots\dots\dots} \Phi_p Y_{t-p} + e_t + e_t – \Theta_1 e_{t-1} – \Theta_2 e_{t-2\dots\dots\dots} – \Theta_q e_{t-q}$$

We say that $\{Y_t\}$ is a mixed autoregressive moving average process of order p and q, respectively. We abbreviate the name to ARMA (p, q).

4. ARIMA model

ARIMA models are the most general class of models for forecasting a time series that can be stationary via transformations, such as differencing and lagging. Determine the order(s) of difference required to stabilize the series before determining the best ARIMA model for it.

4.1 Box: Jenkins modeling method

The Box–Jenkins model is one of the classes of models to choose from the systematic approach to identify the correct model form. There are statistical tests for the validity of the model. Using this test, it is possible to identify the best appropriate Box–Jenkins model for fitting the data on the milk production at Andassa dairy farm.

5. Result and discussion

From the average daily milk production table, the minimum and maximum records of milk production at Andassa dairy farm are **715.0** and 2295.0, respectively, and the mean for the average daily milk production is 1416.9 (**Table 1**).

Test of randomness by using the different sign test.

Test of hypothesis.

H0: Data daily recorded in the average milk production is random.

H1: ¬H0.

Test statistics.

$$Z_{cal} = \frac{w - E(w)}{\sqrt{Var(w)}}$$

where $\text{E (w)} = \frac{1}{2}(\text{N} - 1)$

$$\text{var (w)} = \frac{\mathbf{1}}{\mathbf{12}}(\text{N} + 1)$$

From this, we have the number of points increase W = 81

$$\text{E (w)} = \frac{\mathbf{1}}{\mathbf{2}}(\text{N} - 1) = \frac{\mathbf{1}}{\mathbf{2}}(179 - 1) = 89$$

$$\text{var (w)} = \frac{\mathbf{1}}{\mathbf{12}}(\text{N} + 1) = \frac{\mathbf{1}}{\mathbf{12}}(179 + 1) = 15$$

Variable	N	N*	Mean	SE Mean	St DEV	Minimum	Q1	Median	Q3	Maximum
Average	179	0	1416.9	28.4	380.0	715.0	1180.0	1370.0	1780.0	2295.0

Table 1.
Average daily milk production.

$$\sqrt{Var(w)} = \sqrt{15} = 3.87$$

Therefore: $Z_{cal} = \frac{81-89}{3.87} = -2.0672.$

Note: Level of significance used in this case is α = 0.05.

Test rule:

We reject the null hypothesis since $|Z_{cal}| = 2.0672 > Z_{\frac{\alpha}{2}} = 1.96$. Thus, at five percent (5%) level of significance, the data on the average daily milk production is not random. This means that there is a systematic pattern. Meaning, the dissemination of the data is not balanced and it must be transformed into random form by using different methods.

6. Stationary time series

This can be judged by looking at its time plot. The time plot appears similar at different points along the time axis, and the time series plot is as shown in **Figure 1**. The average daily milk production of a cow (Y) is decreasing over time (t) and is not stationary; we must change the time series plot into a stationary form using the differencing method.

6.1 Stationary through differencing

After differencing the data in lag two, the data becomes a stationary series, as illustrated in **Figure 2**. This plot shows that the series varies around the mean. This indicates that it does not deviate or drift from the mean, and the time plots appear to be similar in various places. As a result, the time series is stationary, suggesting that the fluctuation in the average daily milk production of cows is not far apart from one another.

6.2 Stationary trend and difference

The trend analysis plot reveals that it is not stationary, implying that it will require differencing to become stationary. The trend analysis becomes stationary after differencing the data by lag two, as illustrated in **Figure 3**. The amount of average daily milk output is declining, as shown by the fitted trend in **Figure 4**. For

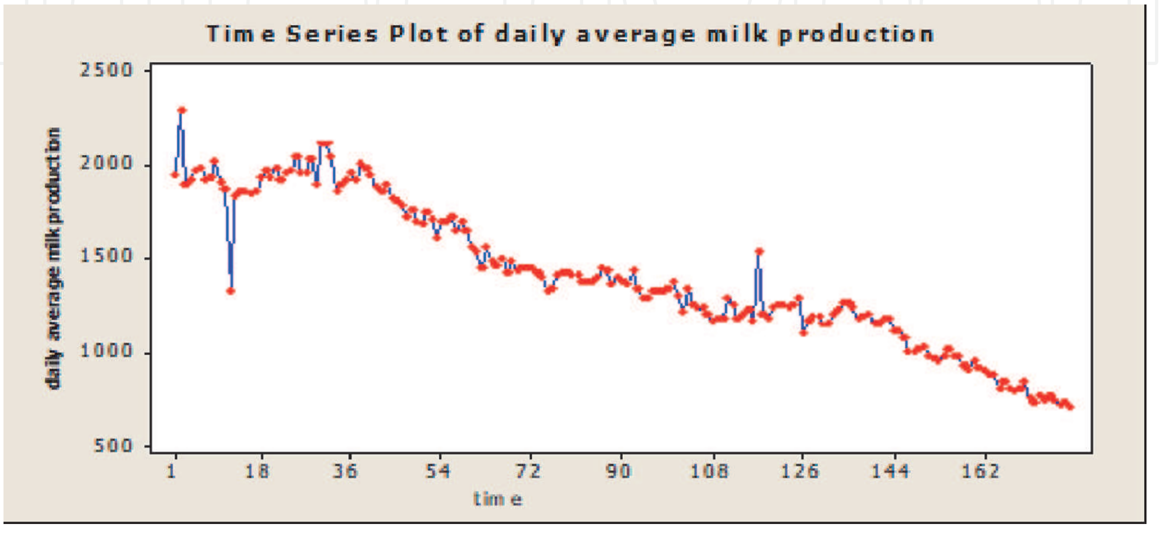

Figure 1.
Average volume of milk production (litter).

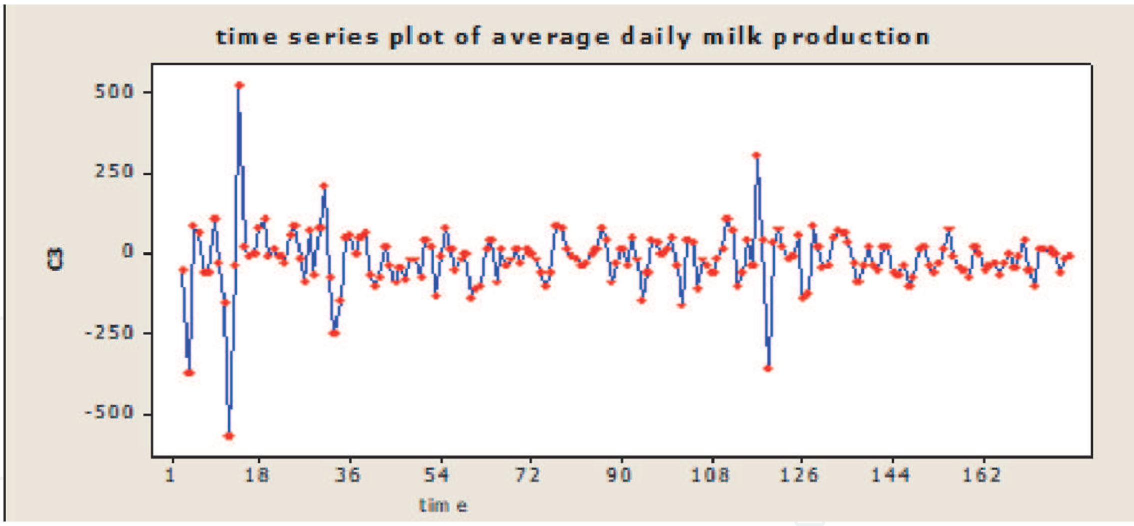

Figure 2.
Stationary time series plot.

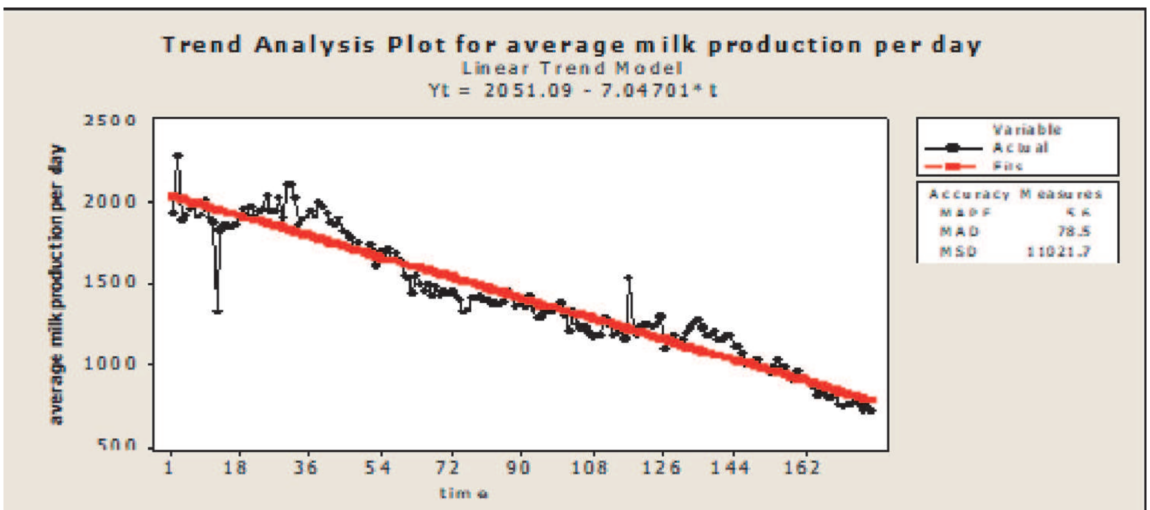

Figure 3.
Trend analysis.

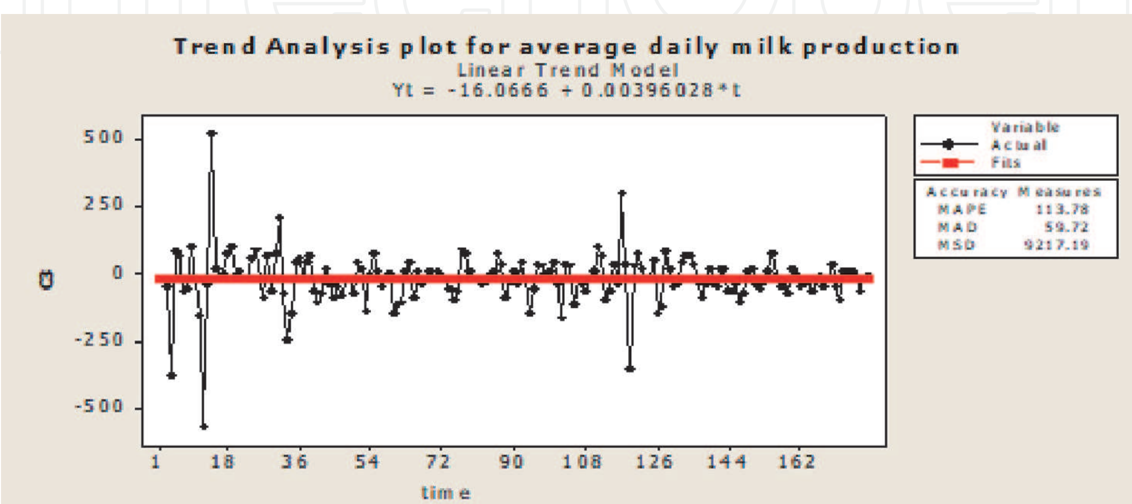

Figure 4.
Stationary trend analysis.

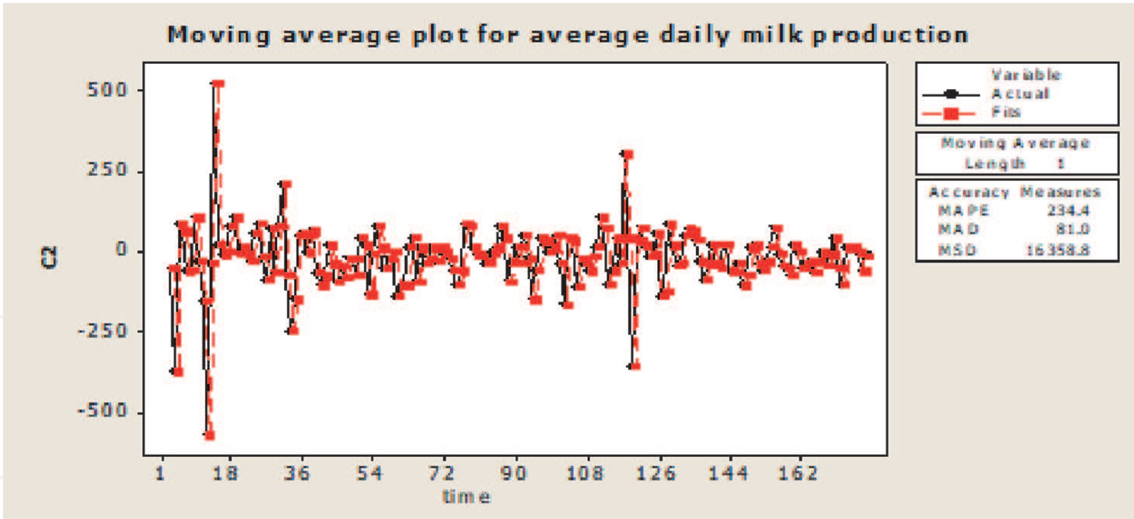

Figure 5.
Moving average plot.

179 days of data, the slope of the trend is −16.066, which represents the rate at which the amount of average milk per day is decreasing. This also implies that the average daily milk consumption decreases over time.

6.3 Moving average

Figure 5 depicts the moving average plot after the data has been transformed into stationary form by differentiating the observations.

Accuracy	Measures
MAPE	234.4
MAD	81.0
MSD	16358.8

For all those three measures, the smaller value is a better fit for the model, that is, MAD = 81.0 is a better fit for the model.

7. Daily milk production autocorrelation function

There is a lag in the numbers. As a result, we must test the AR model to ensure that it is enough. That is, the autocorrelation graph in **Figure 6** shows that the average milk output is one point outside the lower bound, which is AR [1].

7.1 Partial autocorrelation function: for average daily milk production

In Box–Jenkins models, the partial autocorrelation plot or partial correlogram is also often employed for model identification. In the same way, there is one lag number. As a result, we must test the MA model to ensure that it is adequate. That is, under the partial autocorrelations graph, the average daily milk output shows that there is a point outside the bottom boundaries, which is MA [1]. In other words, the distributions of average daily milk output are neither balanced nor equal, as observed in **Figure 7**.

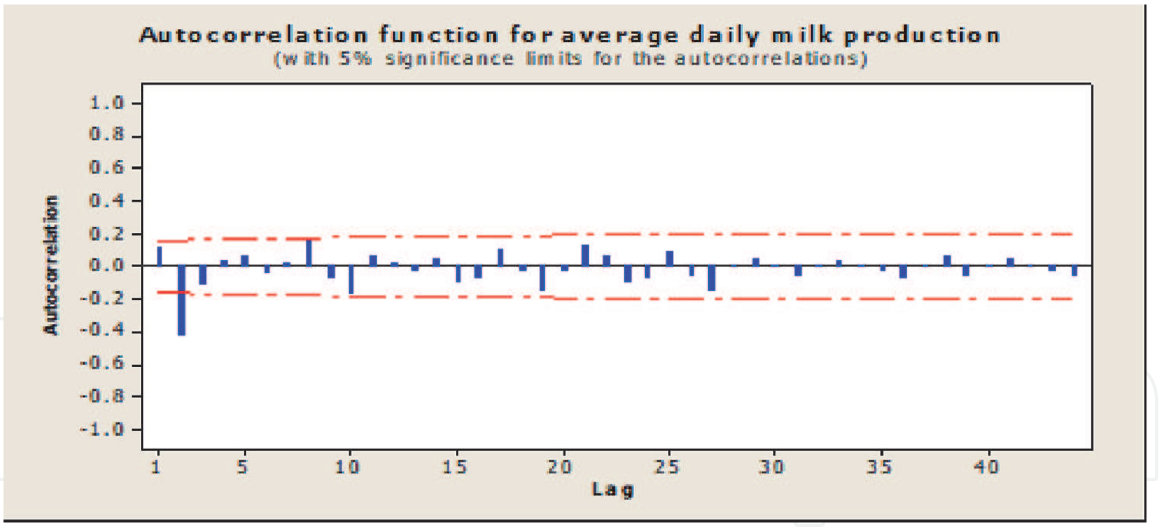

Figure 6.
Autocorrelation plot.

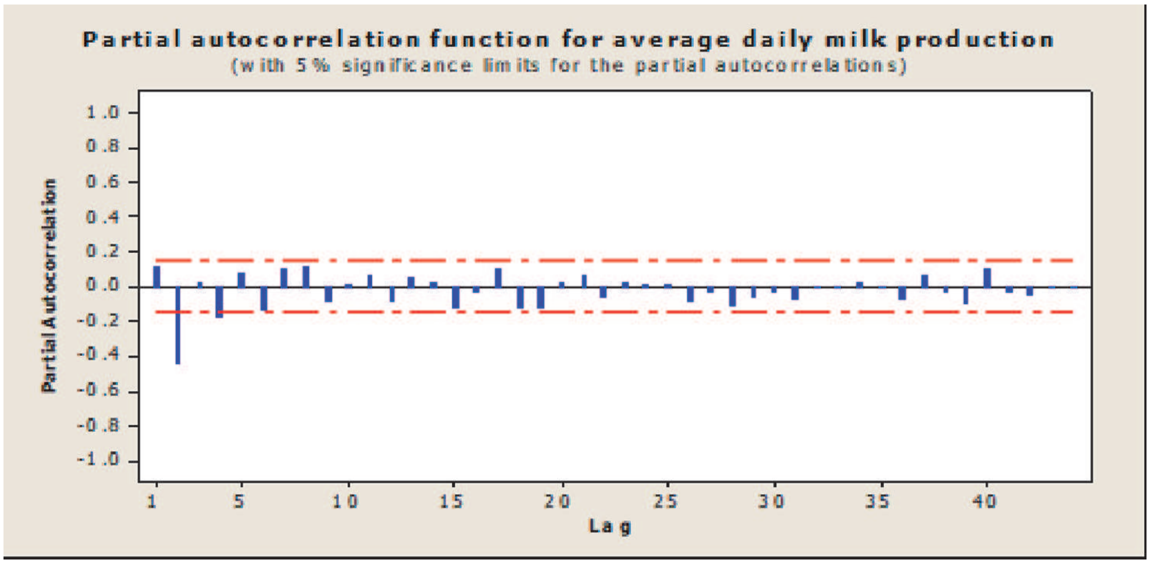

Figure 7.
Partial autocorrelation plot.

7.2 ARMA process

By combining the autoregressive and moving average processes, we obtain a very general time series model, ARMA (1,1).

7.3 ARIMA process

ARIMA (p, d, q) stands for autoregressive integrated moving average process, where d denotes the number of times the data is differenced before it is an ARMA (p, q). As a result, the ARIMA model is ARIMA (p, d, q) = ARIMA (1,2,1).

7.4 ARIMA model: for average daily milk production

Final estimates of parameters.

Category	B	Std. Error	T	P-value
AR [1]	−0.2066	0.0742	−2.79	0.006

Category	B	Std. Error	T	P-value
MA (1)	0.9977	0.0004	2352.08	0.000
B_0	0.2512	0.1447	1.74	0.084

Improved box-pierce χ^2 statistic				
Gap	12	24	36	48
χ^2	67.1	92.0	102.9	109.0
DF	9	21	33	45
P-value	0.000	0.000	0.000	0.000

As we have seen from the MINITAB output, the ARIMA model (1, 2, 1) equation is described as follows.

$$Y_t = -0.2066Y_{t-1} + 0.9977e_{t-1} + e_t$$

7.4.1 Testing of parameters

The final estimates are those that reduce the sum of squared errors to the point where no other estimates yield lower sums of squared errors. As shown in the MINITAB output, a model should include significant parameters. The p-value of ARIMA (1, 2, 1) is less than the significance level (=0.05). This means the parameters are significantly different from zero and have the smallest sum-squared error possible. Then it has parameters that are statistically significant. As a result, the model is adequate.

7.4.2 Forecasting

The process of obtaining the forecast point and the final model in its original form is as follows.

$$Y_t = -0.2066Y_{t-1} + 0.9977\ e_{t-1} + e_t$$

		95% Bounds	
Period	Forecast	Lower	Upper
180	26.663	0	273.768
181	54.100	0	369.879
182	82.662	0	460.574
183	111.242	0	541.606

We can use the 95% confidence interval (CI) defined above to assess the accuracy of the anticipated number. We can state that the forecasted value is accurate since the entire forecast values are found between the lower and upper intervals.

8. Conclusions

The average amount of milk produced at Andassa dairy farm is dropping. The data for 179 days reveal a high degree of variability in daily milk production compared to other days, implying that the amount of milk produced varies greatly from day to day. Because the slope of the trends over the 179 days is −16.066, the amount of milk is falling. The daily milk production graph in the autocorrelations and partial autocorrelations graphs reveals that the top and lower boundaries do not encompass the entire observation. For ARIMA (1, 2, 1), a parameter with a p-value less than the level of significance (0.05) is a parameter. This indicates that the parameter is significantly distinct from zero and has the smallest value.

Author details

Birhan Ambachew Taye*, Alemayehu Amsalu Alen, Ashenafi Kalayu Nega and Bantie Getnet Yirsaw
Department of Statistics, Woldia University, Woldia, Ethiopia

*Address all correspondence to: workneshambachew@gmail.com

References

[1] Jayne TS, Mather D, Mghenyi E. Principal challenges confronting smallholder agriculture in sub-Saharan Africa. World Development. 2010; **38**(10):1384-1398

[2] Taye BA, Alene AA, Nega AK, Yirsaw BG. Time series analysis of cow milk production at Andassa dairy farm, west Gojam zone, Amhara region, Ethiopia. Modeling Earth Systems and Environment. 2021;**7**(1):181-189

[3] Hanna N, Ahmed K, Anwar M, Petrova A, Hiatt M, Hegyi T. Effect of storage on breast milk antioxidant activity. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2004;**89**(6):F518-FF20

[4] Jacobs J, Siegford J. Invited review: The impact of automatic milking systems on dairy cow management, behavior, health, and welfare. Journal of Dairy Science. 2012;**95**(5):2227-2247

[5] Tassew A, Seifu E. Smallholder dairy production system and emergence of dairy cooperatives in Bahir Dar Zuria and Mecha Woredas, Northwestern Ethiopia. World Journal of Dairy & Food Sciences. 2009;**4**(2):185-192

[6] Kirchgässner G, Wolters J, Hassler U. Introduction to modern time series analysis. Springer Science & Business Media; 2012

[7] Cryer JD. Time series analysis. Springer; 1986