



Small-holder Dairy Farming for Economic Emancipation, Forecasting of Small-scale Milk Production at the Nharira-Lancashire Dairy Scheme

Washaya Soul¹, Washaya Dorine Dorcas¹

10.18805/ajdfr.DRF-414

ABSTRACT

Background: Zimbabwe has been experiencing a tremendous shrinkage in the dairy industry for the past two decades; hence, the future is uncertain. The single-variable time series analysis is a useful tool for forecasting supply and demand, particularly for small and medium businesses. As a result, models based on historical data patterns can easily proffer plausible forecasts that are critical for improved planning. The study aimed to investigate the trends of actual milk production and forecast the volume of milk at the Nharira-Lancashire dairy scheme.

Methods: The present study forecasts milk supply for up to 2 years using historical data from 1995-2020. The ARIMA time-series (p, d, q) model was applied to predict monthly milk yield. Twenty-five years of data on milk yield collected from the milk collection centre (MCC) were used for modelling. The presence of a trend was checked through time series plot, stationarity through autocorrelation (ACF) and partial auto-correlation functions (PACF).

Result: The ARIMA (1, 1, 1) model was found to be the best-fitted model for the prediction of monthly milk yield. The results showed that milk yield data is seasonal and follows a particular trend. The milk production is critically low during the months of August to September, while 2009 had the lowest volumes of milk. The forecasts showed that milk yields would increase by 27.5% in 2022.

Key words: ARIMA, Milk output, Seasonality, Trend analysis.

INTRODUCTION

Zimbabwe is witnessing a tremendous decline in the dairy industry (LIMAC, 2018; Washaya and Chifamba, 2018). Milk production has decreased from 262 million litres per year in 1995 to 54.3 million litres in 2018 (LIMAC, 2018). Milk is viewed as nature's only complete (Deshmukh and Paramasivam, 2016) food and, therefore, has immense importance in the day-to-day lives of all individuals and households (Mishra *et al.*, 2020). It provides all the essential nutrients that are required for the growth and development of the body. In principle, milk is produced and consumed in basically all countries and ranks, both in quantity and value, among the top five agricultural commodities (FAO, 2013). Globally, whole fresh cow milk represents 82.7%, followed by buffaloes (13.3%), goats (2.3%), sheep (1.3%) and camels (0.4%) (FAO, 2013). Cows are by far the most common dairy animals, with farmers in developing countries usually keeping them in herds of 2 or 3 (Washaya and Chifamba 2018; Katsande *et al.*, 2013). It is interesting to note that approximately 150 million households worldwide are engaged in milk production (FAO 2013). In most developing countries, milk is produced by smallholder farmers, where it contributes to household livelihoods (Tadesse, 2018), food security and nutrition (Maleko *et al.*, 2018; Pandit *et al.*, 2022). Milk provides relatively quick returns for small-scale producers and is an important source of cash income. In these production systems, dairy animals are a regular and consistent source of food and cash. Because of this, milk and milk products are either consumed or sold every day, which is not the

¹Great Zimbabwe University, Gary Magadzire School of Agriculture, Department of Livestock, Wildlife and Fisheries, P.O. Box 1235, Masvingo.

Corresponding Author: Washaya Soul, Great Zimbabwe University, Gary Magadzire School of Agriculture, Department of Livestock, Wildlife and Fisheries, P.O. Box 1235, Masvingo. Email: jabulaniwashaya@yahoo.com

How to cite this article: Soul, W. and Dorcas, W.D. (2026). Small-holder Dairy Farming for Economic Emancipation, Forecasting of Small-scale Milk Production at the Nharira-Lancashire Dairy Scheme. *Asian Journal of Dairy and Food Research*. **45(4)**: 700-706. doi: 10.18805/ajdfr.DRF-414.

Submitted: 13-06-2024 **Accepted:** 26-10-2024 **Online:** 25-12-2024

case with crops or meat. It is estimated that world milk production is projected to increase by 177 million metric tonnes by 2025, at an average growth rate of 1.8% per year in the next 5 years (FAO, 2013). In the same vein, per capita consumption of dairy products is also projected to increase by 0.8% to 1.7% per year in developing countries and by 0.5% to 1.1% in developed economies (FAO, 2013). However, the current trends in Zimbabwe indicate otherwise. It is therefore essential to know the future production at certain localities that contribute to the growth and development of the dairy sector. There is always a great demand for milk and milk products among people (Taye *et al.*, 2020) and most developing countries are failing to fulfil this need (Shankar *et al.*, 2023) hence, deliberate efforts to improve and expand production systems and

techniques are germane. To achieve this objective, nations should consider milk monitoring as an operations management tool used for both individual cows and herd-level management. As suggested by Haloun *et al.* (2016) the economic profitability of dairy farms is dependent on milk yield. Several studies and much focus have been centred on individual cow data to predict future trends as well as the main factor which influences milk production (Ayadi *et al.*, 2014; Taiwo and Adewumi, 2017). Nevertheless, dairy production is influenced by a plethora of challenges such that a general knowledge of a particular problem may be inadequate to effect an appropriate managerial decision, hence a timeline may be convenient to capture correctly trend, seasonal and cyclic components. Previous research on modelling milk production in cows focused on fitting linear or nonlinear deterministic models to daily, weekly, or monthly milk measurements from lactations either partial or complete lactation data sets (Taye *et al.*, 2020). This study was designed to examine the trend of milk production at the Nharira-Lancashire dairy scheme as an example, to fit the appropriate model and forecast the milk production for the future, based on the previously recorded data from 1995 to 2020.

MATERIALS AND METHODS

Study site

This study was conducted at Nharira Lancashire Milk Collection Centre (MCC) which is located in Agro-ecological Region III, 172 km southeast of Harare in the Chikomba District of Mashonaland East Province. It lies 1450 m above sea level on latitude 19°25' S and longitude 30°35' E. The area receives between 600 to 800 mm of rainfall per annum and the temperature ranges from 5°C to 35°C. Ground frost is common during the cold winter months (May to July). Granite-derived sandy soils dominate this area which is highly susceptible to leaching. Sandy clays are also predominant and are characterized by water logging during peak rainfall months. Tree Bush Savanna is the predominant vegetation type comprising of *Brachystegia*, *Julbernardia* and *Acacia* species.

Data and methods

Milk yields were recorded in liters per cow daily and averaged on a monthly and yearly basis over time, so this study employed only secondary data. The time series analysis method was employed for data analysis. The basic time series models as introduced by Box and Jenkins (1976) were employed in the current study to choose the best-fitting linear model.

Time series method

Time series analysis comprises methods or processes that break down a series into components and explainable portions that allow trends to be identified and estimates and forecasts to be made (Kantz and Schreiber, 2004). Time series analysis probes data points in an attempt to

understand a principal underlying context by use of a model, then forecast future values based on known past values. In the current study autoregressive integrated moving average (ARIMA) model was used. An ARIMA model is a statistical model used to estimate the temporal dynamics of an individual times series. ARIMA models have three components: (1) an autoregressive (AR) component, (2) an integration (I) component and (3) a moving average (MA) component. In ARIMA (p, d, q) time series; p denotes the number of autoregressive terms (AR), d the number of times the series has to be differenced before it becomes stationary (I) and q the number of moving average terms (MA). ARIMA models were used because they are univariate. For ARIMA models to be useful, the main assumption is that past data patterns continue to be evident in the future (Ramasubramanian and Kannan, 2006). Therefore, these models use the pattern to predict future expected values.

Trend analysis

Because the data collected was extensible and showed a term tendency of data to grow or to decline, trend analysis was performed. The trend model used was the linear trend, in which the mean of Y_t is expected to change linearly with time where $Y_t = \beta_0 + \beta_1 t$ based on the least square estimation method and moving average.

RESULTS AND DISCUSSION

Descriptive statistics for milk yield

The overall mean monthly and yearly milk yield of the series are presented in Fig 1 and 2 respectively. The highest milk yield was observed in January throughout the series while the lowest yield was received in August, however, results show a remarkable seasonal nature, in which milk yield decreased from summer, autumn and winter to spring. Over the year's milk production dropped and the least milk yield was recorded in 2009. There is also evidence of yearly fluctuations over the series. Autocorrelation was evident across all months.

Model identification

Milk yield data for Nharira-Lancashire forecasting was non-stationary and showed seasonality (Fig 3). Fig 3 suggests that statistical properties (mean, variance, autocorrelation) are not constant over time hence data was seasonal, thus the time series is trendy (increasing or decreasing over time). The first bar, where the lag = 0 will always equal 1.0, that is, an observation Y_t is always perfectly correlated with itself. The drop off after 0 of ACF bars, is suggestive of a MA (1) model.

Because data was not stationary first order differencing was done using AR (1) and MR (1) models to give the autocorrelation function (ACF), a partial autocorrelation function (PACF) and IACF as shown in Fig 4. A spike at lag 1 (ACF graph) for milk data means that January's outputs are correlated with December's, December's outputs are correlated with November's and so on. In this case, the

spike is negative which means the relationship is negative from one month to the next, indicating a decline in milk yield from one month to the next. However, there were positive correlations at lags 3, 4, 6, 7, 10 and 11 within the first 12 months. Nonetheless, the bars do not extend outside two standard errors, which leads us to believe that they are not significant, yet lags 1, 5 and 12 were significant. The PACF adjusts for all previous lags, in contrast to the ACF which does not adjust for other lags. So, for example, the PACF at lag 5 is the correlation that results after removing the effect of correlations due to the terms at lags 4, 3, 2 and 1. The drop in PACF bars after 1 is suggestive of

an AR (1) model thus PACF spikes indicate lags to try for a pure AR model.

Parameter estimation and validation

A diagnostic check for different ARIMA models was done in IBM Corp (SPSS version 22) (2013) based on the criteria of minimum MAPE and BIC values and the results are shown in Table 1. ARIMA model (1, 1, 1) has a lower MAPE (100) and BIC (18.07) and the highest R - square (48%) and is thus considered for further analysis as shown in Table 2. Although Models (0,1,1); (0,1,0) and (1,1,0) had lower MAPE values their BIC and R-square values were high and low

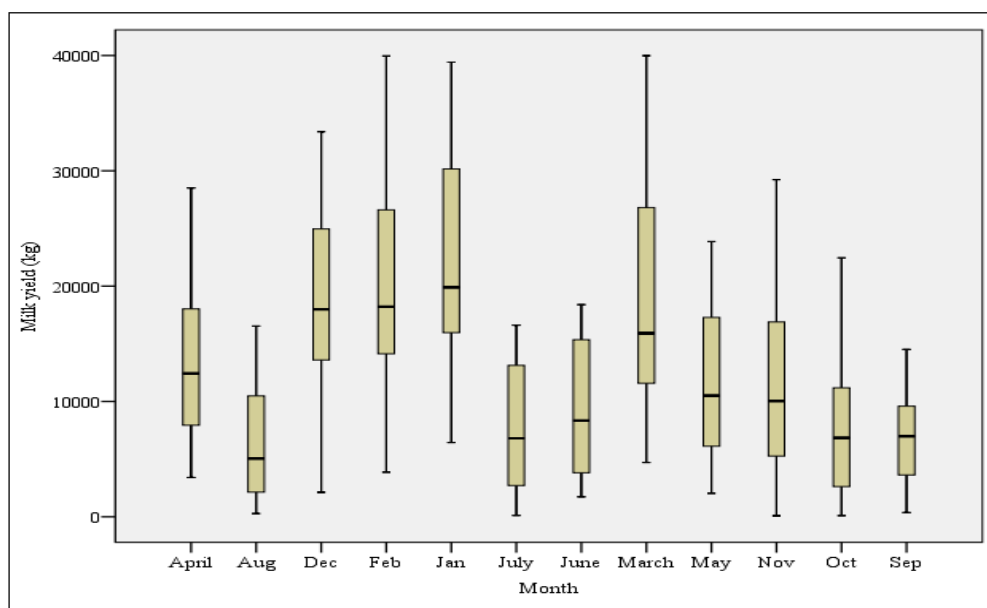


Fig 1: Descriptive analysis of the series of Nharira- Lancashire milk production per month.

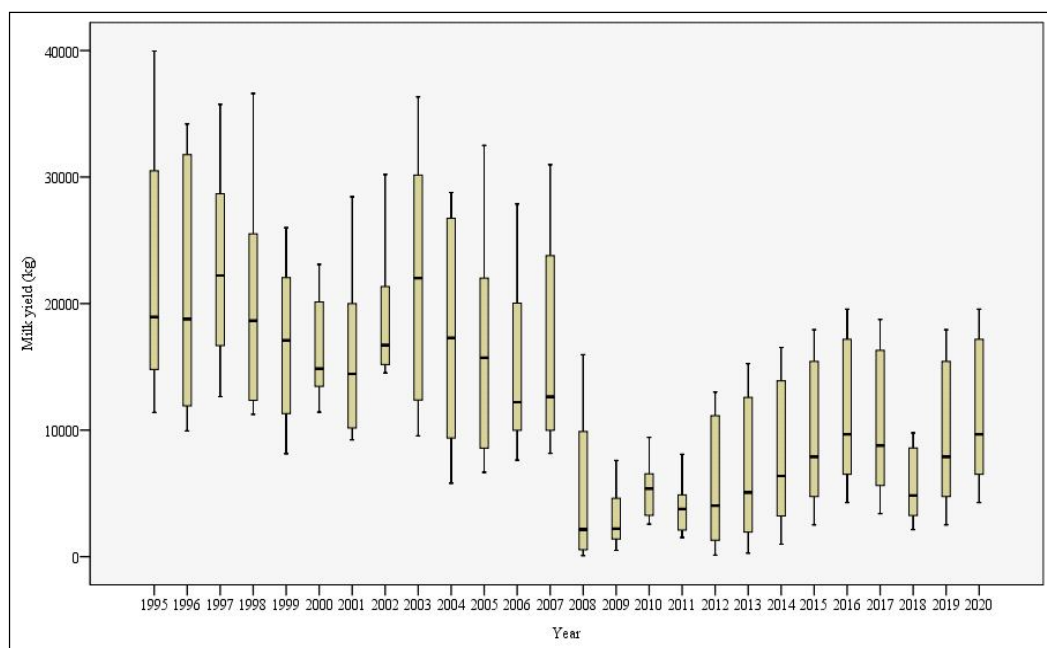


Fig 2: Descriptive analysis of the series of Nharira- Lancashire milk production per year.

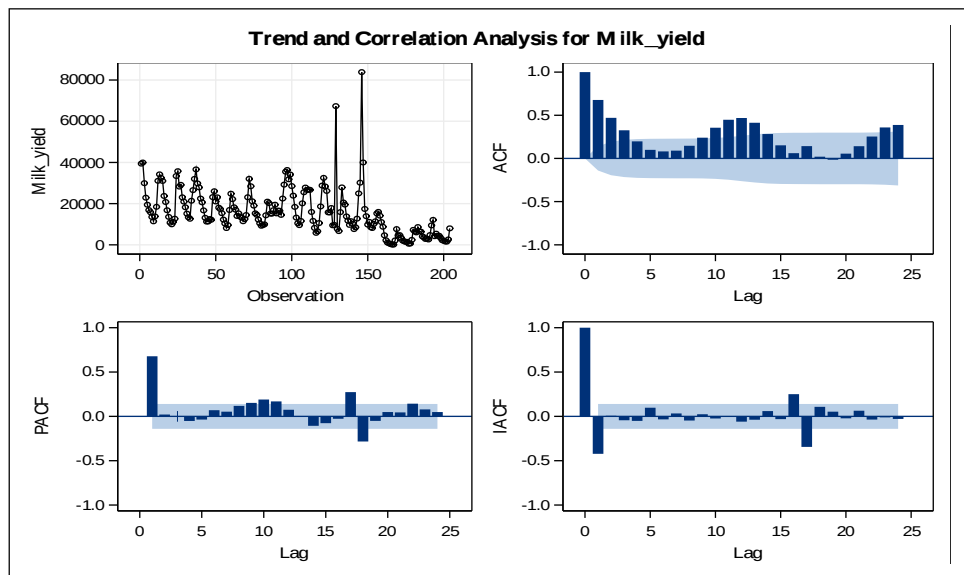


Fig 3: The ACF, PCAF and IACF for milk output in Nharira Lancashire from 1995 - 2020.

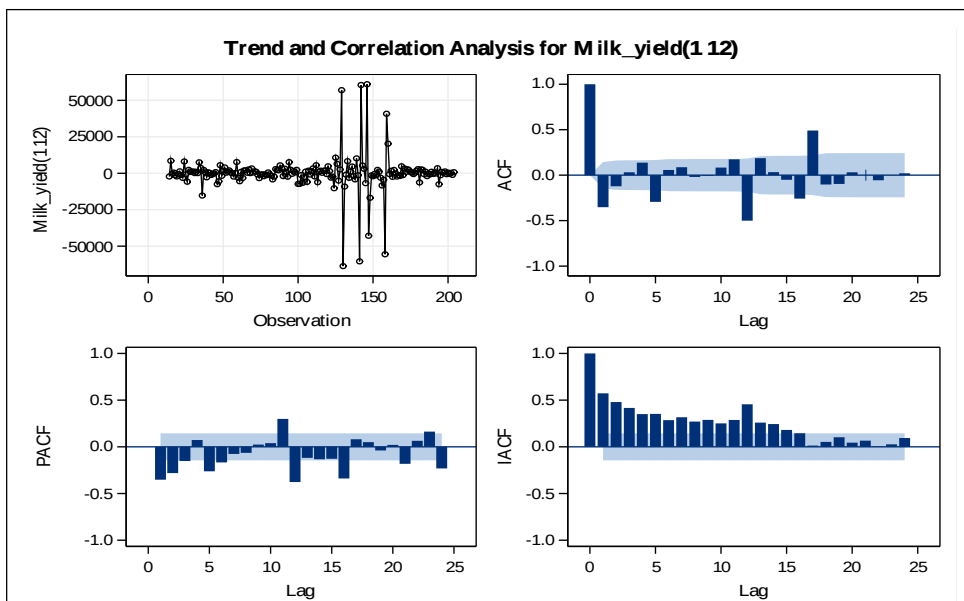


Fig 4: Trend analysis of Nharira –Lancashire milk output data from 1995 - 2020.

Table 1: Diagnostic checking for Nharira –Lancashire milk yield data.

Models	SPSS						
	RMSE	MAPE	MaxAPE	MAE	MaxAE	BIC	R square
1,1,1	8091.639	100.576	4730.660	4571.167	60454.898	18.076	0.480
1,0,0	8317.374	118.713	58054.252	4609.481	5516.900	18.104	0.460
1,0,1	8333.454	116.152	5325.847	4654.827	58085.822	18.134	0.461
0,1,1	8729.783	35.299	582.628	4682.173	57465.349	18.201	0.395
0,1,0	8921.302	31.331	772.015	4259.813	59499.172	18.219	0.365
0,0,1	9206.981	221.331	11429.794	6140.769	62979.467	18.308	0.339
1,1,0	8781.864	33.698	630.380	4570.333	57948.068	18.213	0.388

respectively. The R-square values are average which is not generally the case with continuous data and regression models, the values were not spurious. White noise for the data was checked and the results are shown in Table 3.

This is an approximate statistical test of the hypothesis that none of the autocorrelations of the series up to a given lag are significantly different from zero. In this case, the white noise hypothesis is rejected very strongly. The p-value for the test of the first six autocorrelations is printed as <0.001. The ACF graph also confirms that data is white noise because there are significant spikes in the ACF (Fig 2), thus there is a systematic variation that can be extracted from the series. The form of the estimated ARIMA (1, 1, 1) model for milk yield is shown in Table 4.

Forecasting

Forecast for milk output was done for five years but data representing 2021 and 2022 are shown in Table 5. Forecasted results continue to show the seasonality of milk yield.

$$(1-B)^{\text{milk yield } t} = 19.7029 + \frac{(1 - 0.3775B)}{(1 - 0.90467B)} \text{ at}$$

The series analysed in the current study is considered consistent and stable and did not present abnormal values and showed remarkable fluctuations and yearly peaks in which the highest milk yields occurred between December and April. Milk yield historically shows trend and seasonality (Moura *et al.*, 2020; Sanchez *et al.*, 2014). The effects of season on milk yield have been proposed (Sánchez *et al.*,

2014) thus the seasonality of milk output is inevitable to model future productions (Thakur and Gupta, 2020), the current study also confirms this observation. The main reason for this trend is to do with pasture availability and adequate grazing as mentioned by Tavirimirwa *et al.* (2019); Ngongoni *et al.* (2007) and Phiri *et al.* (2007). Agricultural production data generally is non-stationary because of seasonality and trend (Taye *et al.*, 2020). In the current study milk yield data was non-stationary and first-order differencing was applied to achieve stationarity, however, the time series data was considered not strictly stationary *i.e.* because some of the data properties remain unaffected by a shift in the time origin (Montgomery *et al.*, 2008). In such a case a second form of a time series stationarity, which is called covariance stationary, where the mean and variance are finite should be considered as prescribed by (Heymans *et al.*, 2014), this was however not considered in the current study.

Stationarity indicates that the marginal distribution of Y is the same at any point in time (Haloun *et al.*, 2016), results from the current study do not show strict stationarity. A non-stationary time series is, therefore, a time series that does not have either a constant mean, variance, or covariance over time (Savit 2018; Taye *et al.*, 2020; Heymans *et al.*, 2014; Moura *et al.*, 2020). The lack of strict stationarity, even after differencing, was suggestive of a trend because there was a long-term tendency for the data to increase or decline. When second-order differencing was employed, the volatility around the mean became limited, indicative of over-differencing; hence, we disregarded second-order

Table 2: Model statistics for Nharira-Lancashire milk yield data.

Model 1	Number of predictors	Model statistics				
		Model fit statistics		Ljung-box Q(18)		
		Stationary R ²	R-squared	Statistics	DF	Sig.
Milk output -	0	0.185	0.483	56.097	16	0.000

Table 3: Autocorrelation check for white noise.

Autocorrelation check for white noise									
To Lag	Chi-square	DF	Pr > ChiSq	Autocorrelations					
6	174.90	6	<.0001	0.678	0.470	0.326	0.198	0.100	0.082
12	312.99	12	<.0001	0.090	0.147	0.240	0.356	0.449	0.468
18	378.84	18	<.0001	0.413	0.283	0.152	0.062	0.143	0.020
24	464.09	24	<.0001	-0.015	0.056	0.142	0.254	0.359	0.388

Table 4: Estimated ARIMA (1,1,1) model for milk yield at Nharira-Lancashire dairy scheme.

Model for variable milk yield			
Estimated mean			19.70296
Period(s) of differencing			1,12
Autoregressive factors			Moving Average Factors
Factor 1:	1 - 0.37752 B**(1)		1 - 0.90467 B**(1)
Factor 2:	1 - 0.08416 B**(12)		1 - 0.99976 B**(12)

The estimated model shown in this output is.

Table 5: Forecasted Nharira – Lancashire data for milk yield.

Year	Forecast											
	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
2021	17930.10	17390.74	14248.43	9386.27	7592.21	5303.45	3671.58	2515.11	5452.37	4185.51	8210.64	16601.09
2022	19558.52	19038.79	15915.82	11073.11	9298.65	7029.53	5417.33	4280.51	7237.62	5990.41	10035.38	18445.66

differencing in the current study. The R^2 values in the current study would be considered average and not as high as would be expected for most regression models (Heymans *et al.*, 2014; Haloun *et al.*, 2016), so they were non-spurious. Spurious results are characterised as empirical results with very high R-squared estimates and very low Durbin-Watson statistics, which can be regarded as excellent results but are not of any use (Heymans *et al.*, 2014). The monthly and yearly correlation in milk yield in the current study was expected, however, the autocorrelation was not, hence, we checked for white noise, of which the hypothesis was strongly rejected. The only plausible explanation for this was that the series showed seasonal variation.

CONCLUSION

Milk yield follows seasonality, which entails that it's governed by grazing. There is a greater potential within the study area for improvement, particularly during the dry season, when milk outputs drop significantly. The observed fluctuations could be due to climatic factors and weather conditions as well. Apparently, these fluctuations adversely affected the monthly milk yield and the contribution of small-scale dairy farmers to the national milk output. Different orders of auto-regression and the moving average process of monthly milk yield were selected by estimating the ARIMA models at different p,d, q values and ARIMA (1, 1, 1) was found to be the best-fitted model. It was demonstrated that the application of ARIMA models in the prediction of milk production can be achieved and relied on for future decision-making in the dairy sector.

Conflict of interest

The authors declare no conflicts.

REFERENCES

- Ayadi, M., Matar, A.M., Aljumaah, R.S., Alshaikh, M.A. and Abouheif, M. (2014). Factors affecting milk yield, composition and udder health of Najdi ewes. *International Journal of Animal and Veterinary Advances*. **6(1)**: 28-33. <https://doi.org/10.19026/ijava.6.5613>.
- Bakhtavoryan, R., Capps, O., Salin, V., Dallakyan, S. (2018). The use of time-series analysis in examining food safety issues: The case of the peanut butter recall. *Journal of Food Distribution Research*. **49(2)**: 57-80.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition, San Francisco: Holden Day.
- Deshmukh, S.S. and Paramasivam, R. (2016). Forecasting of milk production in India with ARIMA and VAR time series models. *Asian Journal of Dairy and Food Research*. **35(1)**: 17-22. <https://doi.org/10.18805/ajdfr.v35i1.9246>.
- FAO. (2013). Goals, Sustainable Development. 'The Global Dairy Sector': Facts' www.fao.org (accessed 16 December 2020).
- Haloun, T., Marek, J., Rajmon, R. and Kr's'ikova' M. (2016). Analysis in a time series of milk-yield production' <https://dk.up ce.cz/bitstream/handle/10195/67228> (accessed 16 December 2020).

- Heymans, A., Van Heerden, C. and Vuuren, G.V. (2014). The prominence of stationarity in time series forecasting. *Studies in Economics and Econometrics*. **38**: 1-16.
- IBM Corp. Released (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- Kantz, H. and Schreiber, T. (2004) Nonlinear time series analysis. Cambridge University Press, Cambridge.
- Katsande, S., Matope, G., Ndengu, M. and Pfukenyi, D.M. (2013). Prevalence of mastitis in dairy cows from smallholder farms in Zimbabwe. *Onderstepoort Journal of Veterinary Research*. **80(1)**: 1-7. <https://doi.org/10.4102/ojvr.v80i1.523>.
- LIMAC (2018). Livestock and Meat Advisory Council, Annual Report www.livestockzimbabwe.com (accessed 16 December 2019).
- Maleko, D., Msalya, G., Mwilawa, A., Pasape, L. and Mtei, K. (2018). Smallholder dairy cattle feeding technologies and practices in Tanzania: Failures, Successes, Challenges and Prospects for Sustainability. *International Journal of Agricultural Sustainability*. **16(2)**: 201-13. <https://doi.org/10.1080/14735903.2018.1440474>.
- Mishra, P., Fatih, C., Niranjan, H.K., Tiwari, S., Devi, M. and Dubey, A. (2020). Modelling and forecasting of milk production in Chhattisgarh and India. *Indian Journal of Animal Research*. **54(7)**: 912-917.
- Moura, L., Saúde, S., Gabriel, G.T. and Balestrassi, P.P. (2020). Forecasting of buffalo milk in a Brazilian dairy using the Arima model. *Buffalo Bulletin*. **39(2)**: 201-213.
- Montgomery, D.C., Jennings, C.L. and Kulahci, M. (2008). Introduction to Time Series Analysis and Forecasting. Hoboken, N.J. Wiley.
- Ngongoni, N.T., Mwale, M., Mapiye, C., Moyo, M.T., Hamudikuwanda, H. and Titterton, M. (2007). Evaluation of cereal-legume intercropped forages for smallholder dairy production in Zimbabwe. *Livestock Research for Rural Development*. **19**: 145-148.
- Pandit, P., Bakshi, B., Paul, M. and Pooja, B.S. (2022). ARIMA-genetic algorithm approach for forecasting milk production in India. *Asian Journal of Dairy and Food Research*. **43(4)**: 784-789. doi: 10.18805/ajdfr.DR-1782.
- Phiri, M.S., Ngongoni, N.T., Maasdorp, B.V., Titterton, M., Mupangwa, J.F. and Sebata, A. (2007). Ensiling characteristics and feeding value of silage made from browse tree legume-maize mixtures. *Tropical and Subtropical Agroecosystems*. **7(3)**: 149-156.
- Ramasubramanian, P. and Kannan, A. (2006). A Genetic-algorithm based neural network short-term forecasting framework for database intrusion prediction system. *Soft Computing*. **10(8)**: 699-714. <https://doi.org/10.1007/s00500-005-0513-9>.
- Sánchez, L., Cabanas, G., Abad, Y. and Torres, T. (2014). Use of ARIMA models for predicting milk production. Case study in UBPC "Maniabo", Las Tunas. *Cuban Journal of Agricultural Science*. **48(3)**: 213-218.
- Savit, R. and Manuca, R. (2018). Stationarity and nonstationarity in time series analysis'. *Physica D*. **99**: 134-161. [https://doi.org/10.1016/S0167-2789\(96\)00139-X](https://doi.org/10.1016/S0167-2789(96)00139-X).
- Shankar, S.V., Ajaykumar, R., Ananthakrishnan, S., Aravinthkumar, A., Harishankar, K., Sakthiselvi, T. and Navinkumar, C. (2023). Modeling and forecasting of milk production in the Western Zone of Tamil Nadu. *Asian Journal of Dairy and Food Research*. **42(3)**: 427-432. doi: 10.18805/ajdfr.DR-2103
- Tadesse, G. (2018). Impact of Climate change on smallholder dairy production and coping mechanism in sub-Saharan Africa - Review. *Agricultural Research and Technology: Open Access Journal*. **16(4)**: 41-54. <https://doi.org/10.19080/artoaj.2018.16.556000>.
- Taiwo, I.S. and Adewumi, O.O. (2017). Genetic and non-genetic factors affecting yield and milk composition in goats. *Advances in Dairy Research*. **05(02)**: 2-5. <https://doi.org/10.4172/2329-888x.1000175>.
- Tavirimirwa, B., Manzungu, E., Washaya, S., Ncube, S., Ncube, S., Mudzengi, C. and Mwembe, R. (2019). Efforts to improve zimbabwe communal grazing areas: A review. *African Journal of Range and Forage Science*. **36(2)**: 73-83. <https://doi.org/10.2989/10220119.2019.1602566>.
- Taye, B.A., Alene, A.A., Nega, A.K. and Yirsaw, B.G. (2020). Time series analysis of cow milk production at Andassa Dairy Farm, West Gojam Zone, Amhara Region, Ethiopia. *Modelling Earth Systems and Environment*. <https://doi.org/10.1007/s40808-020-00946-z>.
- Thakur, R. and Gupta, R.K. (2020). Prediction of milk yield using ARIMA. *Journal of Entomology and Zoology Studies*. **8(3)**: 1119-23.
- Washaya, S. and Chifamba, E. (2018). Smallholder dairy farming: A solution to low milk production in Zimbabwe. *Journal of Dairy and Veterinary Sciences*. **8(2)**: 8-10. <https://doi.org/10.19080/jdvs.2018.08.555735>.