International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(6): 121-126

https://www.mathsjournal.com
Received: 16-10-2024
Accepted: 22-11-2024

Muhammed Irshad M
Research Scholar, Department of Agricultural Statistics, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal, India

Kader Ali Sarkar
Assistant Professor, Department of Agricultural Statistics, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal, India

Digvijay Singh Dhakre
Assistant Professor, Department of Agricultural Statistics, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal, India

Debasis Bhattacharya
Professor, Department of Agricultural Statistics, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal, India

Corresponding Author:
Kader Ali Sarkar
Assistant Professor, Department of Agricultural Statistics, Palli Siksha Bhavana, Visva-Bharati University, Sriniketan, West Bengal, India

Forecasting of monthly cardamom price using long memory time series modelling technique

Muhammed Irshad M, Kader Ali Sarkar, Digvijay Singh Dhakre and Debasis Bhattacharya

DOI: https://doi.org/10.22271/maths.2024.v9.i6b.1906

Abstract
Cardamom is an important spice crop known for its aromatic qualities. Kerala is a major producer state of small cardamom in India. This study aimed to develop a long memory time series model for forecasting of monthly average farm wholesale price of small cardamom in Kerala. Monthly price data collected from Directorate of Economics and Statistics, Government of Kerala for the period January 2007 to December 2020. Upon analysis of data, ARFIMA (0, 0.22, 1) model was selected which in turn compared with traditional ARIMA model. The results emphasised the superior performance of ARFIMA (0, 0.22, 1) model over ARIMA model in forecasting the long memory price time series of small cardamom.

Keywords: ARFIMA, Fractional differencing, Long-memory, Cardamom price, Price forecasting

Introduction
Small cardamom, scientifically known as *Elettaria cardamomum*, is a spice from the ginger family renowned for its aromatic seeds, which enhance both sweet and savoury dishes. India ranks as the second-largest producer of small cardamom globally, following Guatemala. Small cardamom production in India is 24.5 thousand tonnes. During 2022-23, the price of small cardamom rose to Rs 1,075.45 per kg, compared to Rs 996.52 per kg in the previous year. Kerala contributing 90.6 percent of the national small cardamom production as the state produced 22.165 thousand tonnes and area under small cardamom cultivation is 40,345 hectares (Economic Review 2023, State Planning Board, Government of Kerala) [5]. The value of output from small cardamom in India is almost Rs. 2,076.14 crores and of Kerala is Rs. 1, 416.83 crores at current prices (Ministry of Statistics and Programme Implementation, Government of India, 2022-23).

Time-series analysis focuses on sequential data collected over time. A key characteristic of time-series data is the dependence between consecutive observations. This interdependence is captured efficiently using various dynamic and stochastic models. Time-series analysis plays a vital role in everyday life, primarily for forecasting. It involves estimating future values of key variables by analysing the past data. Price forecasting in agriculture is essential for decision-making by farmers, agribusinesses, and governments. This study is to develop a model for forecasting of monthly price of small cardamom in Kerala. Box *et al*. (1970) introduced a method in time-series forecasting by developing the Autoregressive Integrated Moving Average (ARIMA) model The general notation for an ARIMA model is ARIMA (*p*, *d*, *q*), where '*p*', '*d*', and '*q*' are the autoregressive, differencing and moving average parameters. which are widely used due to their adaptability, flexibility, and broad applicability. Irshad et. al (2024) [10] explored ARIMA modelling in forecasting of weekly clove prices of Kerala and attained reasonable accuracy. However, ARIMA model is based on some assumptions also; it cannot model the situations involving long-range dependency over time. The Autoregressive Integrated Moving Average (ARFIMA) model, first introduced by Granger and Joyeux (1980) [7] and later refined by Hosking (1981) [9], known as a major advancement in time-series modelling by allowing fractional differencing. This parametric tool was especially developed to analyse long-memory processes and with the advantage of fractional differencing parameter (*d*). The Hurst exponent (*H*), a statistical measure explained by Booth *et al*. (1982) [3] in one of their studies, is a valuable statistical tool for measuring long-term memory in time-series data. Geweke and Porter-Hudak (1983) [6] developed the GPH estimator, which is an important tool for estimating the long-memory parameter, '*d*' in ARFIMA models. This novel strategy employs a simple linear regression of the logarithmic periodogram against a deterministic regressor. The ability of GPH estimator to estimate the slope parameter using Ordinary Least Squares (OLS) is an important feature as per their study. Previous scholars have done practical comparisons of *H* and *d* estimators, as their finite sample characteristics might differ greatly from their asymptotic qualities. Taqqu *et al*. (1995) [17] performed empirical research on nine estimators utilizing a single series of 10,000 data points, testing five different combinations of *H* and *d* values and repeating the procedure fifty times. Bhardwaj and Swanson (2006) [2] studied the fractional I (*d*) processes and presented evidence that ARFIMA models, estimated using standard estimation procedures sometimes providing better out-of-sample predictions than AR, MA, ARMA, GARCH, and related models. In a study, the long-term memory of Hong Kong Hang Sheng index using MRS analysis and establishes an ARFIMA model using fractional differencing. (Xiu and Jin, 2007) [18]. Karia *et al*. (2013) [11] examined non-stationarity in the prices of crude palm oil, which necessitates first-order differencing. But doing so may result in over differencing, which would remove significant trend components. Hence, he suggested ARFIMA model with fractional differencing. Octaviyani *et al*. (2019) [14] modelled wind speed for airplane operations by using the ARFIMA model. They considered two estimation methods for the differencing parameter: parametric (Exact Maximum Likelihood) and semiparametric (Geweke and Porter-Hudak, Smooth GPH, Local Whittle, and Rescale Range). The study finds that the GPH method provides the best estimation over other estimators. Mitra and Paul (2021) [13] applied the ARFIMA model to daily wholesale rice price data in India to capture long memory in the dataset. Though the data was stationary, it showed significant long memory and hence chosen ARFIMA for their modelling purpose. The ARFIMA model was evaluated using MAPE, RMSE, and MAE, proved with superior forecasting performance compared to the ARIMA model.

Materials and Methods

Data description

To conduct the study, monthly farm whole sale price data for the state of Kerala is collected from the Directorate of Economics and Statistics, Government of Kerala for the period January 2007 to December 2020 and 90 percent data is used for model building and remaining 10 percent data is kept for model validation.

ARIMA model

ARIMA model is a widely used time series forecasting method which combines autoregression, differencing, and moving averages. ARIMA is designed to capture different patterns and trends in time series data, making it useful for predicting future values. The core of the ARIMA model involves the synthesis of Autoregressive (AR) and Moving Average (MA) polynomials, creating a complex polynomial representation as illustrated in equation. The ARIMA (*p*, *d*, *q*) model is then employed across all data points in the time series data

$$y_t = \mu + \sum_{i=1}^{p}(\sigma_i y_{t-1}) + \sum_{j=1}^{q}(\theta_j \varepsilon_{t-1}) + \varepsilon_t, \quad (1)$$

where μ represents the mean value of the time series data, p denotes the number of autoregressive lags, σ: signifies the autoregressive coefficients (AR), q: stands for the number of lags in the moving average process., θ represents the moving average coefficients (MA), ε denotes the white noise in the time series data.

Tests for stationarity

A time series is a stochastic process characterized by a random variable and a time variable, and it can be either stationary or nonstationary. A series is considered stationary if its mean and variance remain constant over time. Various statistical tests are used to check stationarity, including the ADF test (Dickey and Fuller, 1979) [4], Phillips-Perron test etc. Null hypothesis for this test is η =0 i.e the data is non stationary against the one-sided alternative hypothesis η <0 i.e. the data set is stationary. Phillips and Perron (1988) [15] have provided an alternative test for checking stationarity of a time series which can handle the serial correlation and heteroskedastic error variance. After this test if a series turns out to be a non-stationary then by differencing the series method the series can be transformed as stationary series.

Model selection criteria

Akaike Information Criterion **(**AIC) is based on information theory. It quantifies the trade-off between the goodness of fit of the model and the complexity of the model. The lower the AIC value, the better the model is considered. The AIC value is calculated using the formula:

$$\text{AIC} = -2 * \text{log-likelihood} + 2 * k, \quad (2)$$

where *k* is the number of model parameters.

BIC is a criterion for model selection among a finite set of models, and it is derived from a Bayesian perspective. Like AIC, BIC penalizes models for complexity, but the penalty for additional parameters is more severe in BIC. Similar to AIC, a lower BIC value indicates a better-fitting model The BIC value is calculated using the formula:

$$\text{BIC} = -2 * \text{log-likelihood} + k * \log(n), \quad (3)$$

where *k* is the number of model parameters and *n* is the sample size.

Long memory process

The concept of long-memory or long-range dependence in a time series correlation between observations located at distant time points. When this decay is considerably slower, resembling a hyperbolic trend, compared to the typical ARIMA process, it is indicative of long-range dependence. The Auto-correlation Function (ACF) is a valuable tool for describing the rate of this dependence decay. Long-memory time series exhibits a persistent structure within their ACF. In the case of well-known stationary processes like ARMA, their ACF typically exhibits exponential decay, denoted as $\rho_k \approx |m|^k$ where $|m|$ is less than 1. This property is the characteristic of stationary ARMA (*p*, *q*) processes. In contrast, long memory processes showcase a notably slower

decay rate in their autocorrelation function. This slower decay aligns with $\rho_k \approx Hk^{2d-1}$, where k approaches infinity, and H is a constant while d represents the long memory parameter.
The plotting of Autocorrelation Function (ACF) is the most popular method of detecting the presence of long-memory process of a time-series data. If the plot shows the decaying of ACF very slowly or hyperbolically at larger lags, indicates the possible presence of long-memory property. This process cannot be modelled using ARIMA as ARIMA process does not support fractional integration and non-stationary process.

GPH estimator: The Hurst parameter, denoted as "H", is used to describe the autocorrelation structure of a time series. A value of H between 0.5 and 1 indicates long-memory or positive autocorrelation, while H between 0 and 0.5 suggests short-memory or anti-persistent behaviour. The GPH estimator is employed to estimate the Hurst parameter by analysing the autocorrelation function of the time series at different lags. When the autocorrelation function decays slowly or hyperbolically, it can suggest a high value of H, indicating long-memory characteristics.
$d = (1 - H)$, where H is the Hurst parameter and d is the GPH estimate which is fractional parameter. Value of d ranges between 0 to 0.5 which confirm the stationary and persistent long- memory.

ARFIMA model

Long-memory time-series model, known as ARFIMA model, offer the flexibility to employ non-integer values for the differencing parameter, (Hosking, 1981) [9]. These models prove valuable in encapsulating a fundamental trait of time series with long memory: the gradual decay in dependence between two data points, contrasting with the exponential decay observed in the standard ARIMA process. In the integrated section of an ARIMA model, the differencing operator $(1 - L)$ where L represents the backshift operator, is traditionally raised to an integer power. The ARFIMA model, on the other hand, provides the advantage of allowing non-integer differencing (Granger and Joyeux, 1980) [7].

$$(1 - L)^2 = 1 - 2L + L^2 \quad (4)$$

whereas, in ARFIMA process, the power of the model is allowed to take fractional values, with the expression of the term illustrated using the following formal binomial series expansion (Helms, 1984) [8],

$$(1 - L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \cdots = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j, \quad (5)$$

where $\binom{d}{j} = \frac{d!}{j!(d-j)!}$

An ARFIMA (p, d, q) process of a time series y_t is defined as

$$\rho(L)(1 - L)^d = \theta(L)e_t, \quad (6)$$

where $\rho(L) = 1 - \rho_1 L - \cdots - \rho_1 L^p$, $\theta(L) = 1 - \theta_1 L - \cdots - \theta q L^q$ are the AR and MA operators sharing no common roots. $(1 - L)^d$ is the fractional differencing operator and e_t are assumed to be independent and identically distributed (i.i.d) with zero mean and constant variance σ^2. When $d \in (0, ½)$, the autocovariance function of this solution satisfies $\lim_{k\to\infty} \rho(k)/[ck^{1-2d}] \to 1$ where c is a constant.

Results and Discussions

Descriptive statistics

The descriptive statistics (Table 1) for monthly small cardamom prices in Kerala showed variability and skewness in the data. The prices ranged from a minimum of Rs 267.9 per kg to a maximum of Rs.3515 per kg, with a mean value of 910.7. The standard deviation of 583.41 reflects substantial fluctuations around the mean. The Coefficient of Variation (CV) is 64.06%, showing a high degree of relative variability in the data. The skewness of 2.38 indicates a positively skewed distribution and kurtosis of 9.04 implies leptokurtic distribution, suggesting a sharp peak and heavier tails, indicating occasional extreme price movements.

Table 1: Descriptive statistics of monthly price of cardamom in Kerala

Series	Min	Max	Mean	St. dev.	CV (%)	Skewness	Kurtosis
Cardamom	267.9	3515	910.7	583.41	64.06	2.38	9.04

Time series plot

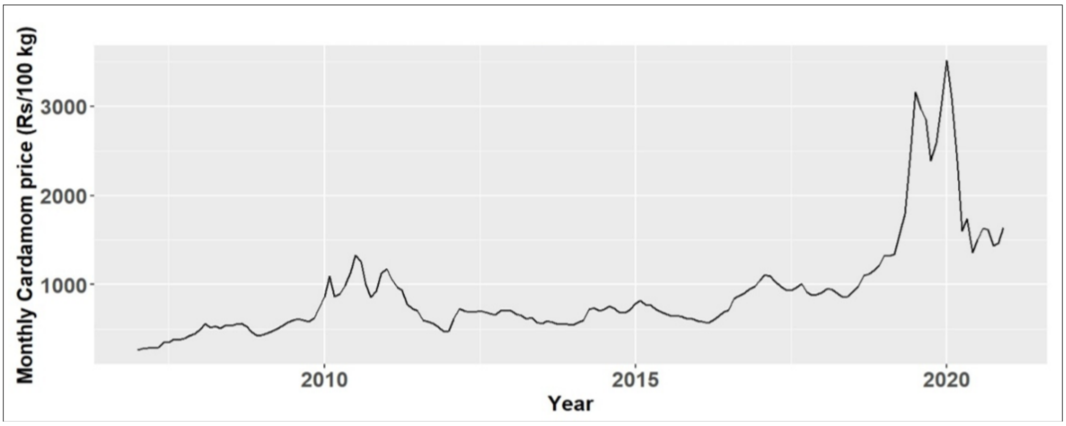

Fig 1: Time series plot of monthly small cardamom price of Kerala

Test for stationarity

The stationarity tests for small cardamom prices using the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test indicate non-stationarity in the series (Table 2).

Table 2: Tests for stationarity

ADF test		PP test	
d	***p*-value**	***Z***	***p*-value**
-2.78	0.24	-13.23	0.35

ACF and PACF plots: The ACF plot as seen in the figure shows considerable amount long memory property with significant lag up to 18 months which is the similar trend shown by the PACF plot also as in figure 2(a) and 2(b).

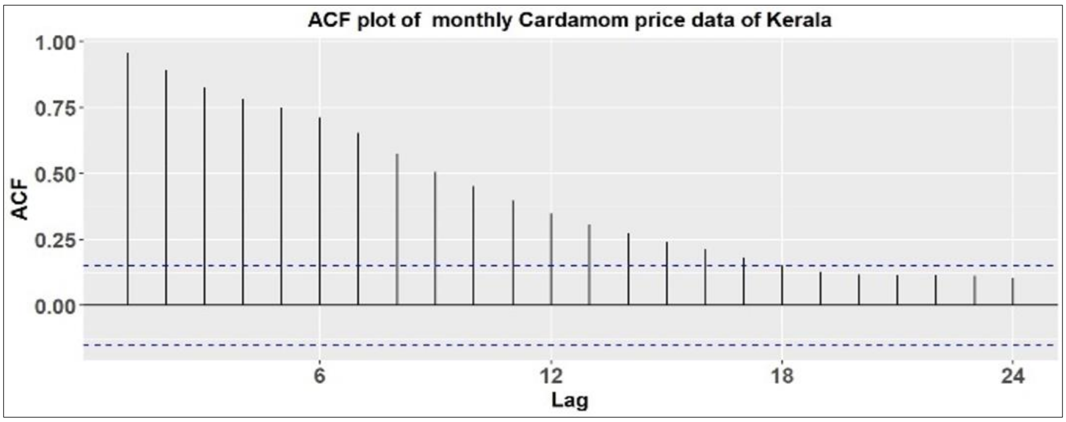

Fig 2(a): ACF plot of raw data

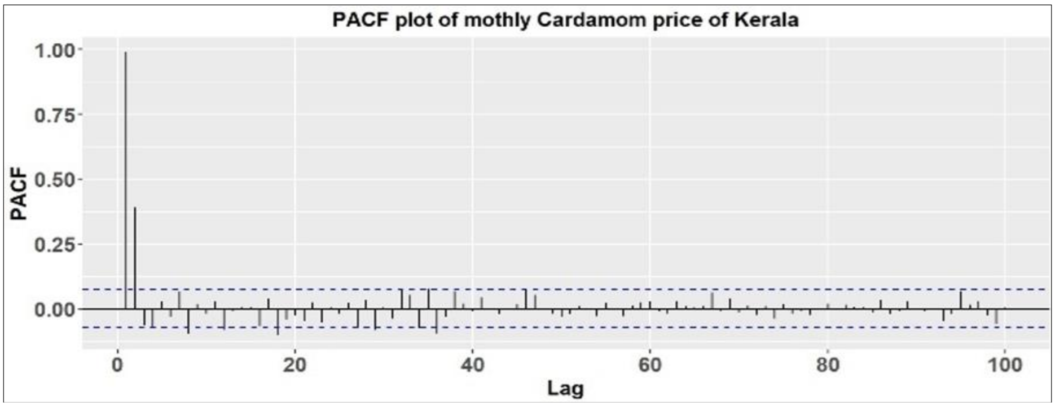

Fig 2(b): PACF plot of raw data

ARIMA model fitting

Based on the AIC and BIC values, ARIMA (3, 1, 2) model is selected and the orresponding AIC as well as BIC values are given in the Table 3.

Table 3: AIC and BIC values of ARIMA models of monthly cardamom price

Sl. No	Model	AIC	BIC
1	ARIMA (3, 1, 2)	1738.83	1756.85

The estimates of ARIMA (3, 1, 2) model are given in the Table 4. The coefficient estimation were followede the method of Maximum Likelihood Estimation (MLE). The *p*-values for all the coefficients come out to be less than or equal to. 0.05 which implies the model parameters are valid and statistically significant.

Table 4: ARIMA (3, 1, 2) coefficients of monthly cardamom price of Kerala

Parameter	Estimate	Standard error	*p*-value
AR (1)	0.5	0.2	0.05
AR (2)	-0.46	0.15	<0.01
AR (3)	-0.27	0.11	0.01
MA (1)	-0.77	0.21	<0.001
MA (2)	0.45	0.19	0.05

GPH estimator

The long-memory parameters of small cardamom, estimated using the GPH and Sperio tests are given in Table 5, indicate the presence of long memory in the price series. Both tests yield fractional differencing values close to 0.2 (GPH: 0.22, Sperio: 0.2) with *p*-values less than 0.01, strongly rejecting the null hypothesis of no long memory.

Table 5: Long memory parameters of the monthly cardamom price

GPH test		Sperio test	
d	***p*-value**	***d***	***p*-value**
0.22	<0.01	0.2	<0.01

ARFIMA model fitting

Based on the minimum values of AIC and BIC, ARFIMA (0,0.22,1) model is being selected as in the Table 6. The coefficient of moving average parameter is estimated to be 0.71 with standard error 0.21 which is highly significant also is given in Table 7.

Table 6: ARFIMA model selection

Model	AIC	BIC
ARFIMA (0, 0.22, 1)	1799.09	1808.14

Table 7: ARFIMA (0, 0.22, 1) coefficients of monthly cardamom price

Parameter	Estimate	Standard error	*p*-value
MA (1)	0.71	0.21	<0.001

Model validation

The performance comparison between ARIMA (3, 1, 2) and ARFIMA (0, 0.22, 1) models for small cardamom price is presented in Table 8. On the training data, ARIMA shows slightly better accuracy with lower RMSE and MAPE. However, on the testing data, ARFIMA significantly outperforms ARIMA with a much lower RMSE and MAPE than ARIMA. This indicates that ARFIMA demonstrates better predictive performance, due to its ability to capture long-memory dynamics. It is also worth notable that ARFIMA could manage to perform this much better with considerably a smaller number of parameters as compared to the ARIMA model. The very high amount of error in the test data of ARIMA possibly due to its inability to capture the sudden peaks and fluctuation seen in the actual test data.

Table 8: Comparison of ARIMA (3, 1, 2) and ARFIMA (0, 0.22, 1) models

Sl. No	Model	Training Data		Testing Data	
		RMSE	MAPE	RMSE	MAPE
1	ARIMA (3, 1, 2)	78.8	5.89	6610.63	338.28
2	ARFIMA (0, 0.22, 1)	87.27	7.57	539.78	20.89

Fitted plot

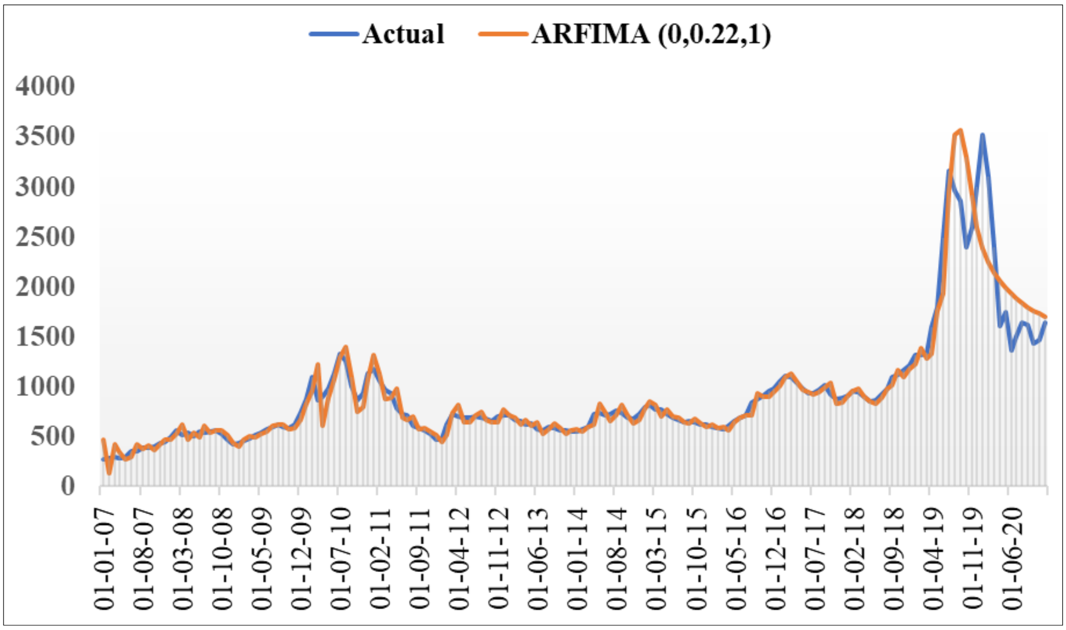

Fig 3: Actual vs Fitted plot of monthly small cardamom price of Kerala

Actual vs fitted plot of ARFIMA (0, 0.22, 1) is given in figure 3. The figure shows good model fitting especially in the training data. IN the forecasted part, even if there were unexpected peaks and downs in the actual data, ARFIMA somehow managed to address those fluctuations up to certain extent.

Conclusions

This study aimed to develop forecasting model for monthly price of small cardamom for Kerala, a major small cardamom producing state in India. Since the data suggests long memory property as evidenced by ACF plot and GPH estimator, ARFIMA (0, 0.22, 1) model is chosen to model the data. Best ARFIMA model is selected is based on the lowest AIC and BIC values. The study also compared the model accuracy with ARIMA (3, 1, 2) model. The comparison showed almost similar performance on the training dataset, while ARFIMA (0, 0.22, 1) model outperformed considerably on test dataset. Hence the study proposes ARFIMA (0, 0.22, 1) for forecasting monthly price data of small cardamom for Kerala. The proposed model can be used by stakeholders of small cardamom production, processing and marketing for their planning and budgetary allocations. This model will also help the governments to anticipate the future prices and make polices in favour of farmers involved in the small cardamom production.

References

1. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716-723. https://doi.org/10.1109/TAC.1974.1100705
2. Bhardwaj G, Swanson NR. An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series. Journal of Econometrics. 2006;131(1):539-578. https://doi.org/10.1016/j.jeconom.2005.01.016
3. Booth GG, Kaen FR, Koveos PE. R/S analysis of foreign exchange rates under two international monetary regimes. Journal of Monetary Economics. 1982;10(3):407-415.
4. Dickey D, Fuller W. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. 1979;74:427-431.
5. Economic Review. State Planning Board, Government of Kerala; c2023. Available at https://spb.kerala.gov.in/economic-review/ER2023/
6. Geweke J, Porter-Hudak S. The estimation and application of long-memory time series models. Journal of Time Series Analysis. 1983;4:221-238.
7. Granger CWJ, Joyeux R. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis. 1980;4:221-238.
8. Helms BP, Kaen FR, Roseman RE. Memory in commodity futures contacts. The Journal of Futures Markets. 1984;10:559-567.
9. Hosking JRM. Fractional differencing. Biometrika. 1981;68:559-567.
10. Irshad MM, Sarkar KA, Dhakre DS, Bhattacharya D. Comparative analysis of statistical model and machine learning algorithms in forecasting black pepper price of Kerala. Biological Forum – An International Journal. 2024;16(8):63–68.
11. Karia AA, Bujang I, Ahmad I. Fractionally integrated ARMA for crude palm oil prices prediction: Case of potentially over difference. Journal of Applied Statistics.

2013;40(12):2735-2748. https://doi.org/10.1080/02664763.2013.825706

12. Ljung GM, Box GEP. On a measure of a lack of fit in time series models. Biometrika. 1978;65(2):297-303.
13. Mitra D, Paul RK. Forecasting of price of rice in India using long-memory time-series model. National Academy Science Letters. 2021;44(4):289-293. https://doi.org/10.1007/s40009-020-01002-1
14. Octaviyani DI, Wijaya MY, Fitriyati N. Estimation parameter ddd in autoregressive fractionally integrated moving average model in predicting wind speed. InPrime: Indonesian Journal of Pure and Applied Mathematics. 2019, 1(2). https://doi.org/10.15408/inprime.v1i2.13676
15. Phillips PCB, Perron P. Testing for unit roots in time series regression. Biometrika. 1988;75:335-346.
16. Statistical Report on Value of Output from Agriculture and Allied Sectors. Ministry of Statistics and Programme Implementation, Government of India; c2022-23. Available at https://www.mospi.gov.in/download
17. Taqqu MS, Teverovsky V, Willinger W. Estimators for long-range dependence: An empirical study. Fractals. 1995;3(4):785-798.
18. Xiu J, Jin Y. Empirical study of ARFIMA model based on fractional differencing. Physica A: Statistical Mechanics and Its Applications. 2007;377:138-154. https://doi.org/10.1016/j.physa.2006.11.030