Analysis Of Price of Commodity Goods(Banana)

Heet Thakkar
202201431
DAIICT
Gandhinagar, India
202201431@daiict.ac.in

Uday Parmar
202203031
DAIICT
Gandhinagar, India
202203031@daiict.ac.in

Ayush Gandhi
202201057
DAIICT
Gandhinagar, India
202201057@daiict.ac.in

I. Introduction

Commodity goods such as fruits, vegetables, grains, and other essential agricultural products are the backbone of any economy. Their prices directly affect not just the cost of living for consumers, but also the income and stability of farmers, traders, and businesses across the supply chain.In this project, we delve into the time series analysis of commodity prices, focusing specifically on the price trends of bananas.Through this project, we aim to decompose and analyze the historical price data of bananas to find the hidden patterns and use appropriate forecasting models to estimate future prices. By doing so, we demonstrate how time series analysis can offer meaningful insights into real-world economic and social systems, especially those that revolve around essential commodity goods.

II. The DataSet

A. Understanding The Data

Banana ($/Kg)	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
JAN	0.40	0.39	0.42	0.34	0.34	0.38	0.33	0.42	0.45	0.43
FEB	0.37	0.44	0.38	0.38	0.42	0.44	0.39	0.46	0.49	0.49
MAR	0.42	0.46	0.42	0.41	0.41	0.45	0.50	0.45	0.53	0.70
APR	0.40	0.43	0.48	0.48	0.41	0.48	0.58	0.35	0.48	0.72
MAY	0.44	0.45	0.44	0.55	0.43	0.46	0.39	0.45	0.56	0.66
JUN	0.34	0.39	0.40	0.50	0.46	0.38	0.35	0.38	0.60	0.53
JUL	0.35	0.33	0.33	0.50	0.37	0.32	0.33	0.44	0.42	0.45
AUG	0.34	0.31	0.30	0.46	0.34	0.42	0.32	0.34	0.36	0.51
SEP	0.35	0.44	0.35	0.42	0.38	0.36	0.42	0.35	0.53	0.52
OCT	0.35	0.39	0.29	0.41	0.31	0.28	0.37	0.27	0.42	0.55
NOV	0.38	0.40	0.33	0.34	0.27	0.28	0.29	0.41	0.41	0.51
DEC	0.37	0.42	0.34	0.34	0.30	0.28	0.31	0.39	0.50	0.50

Fig. 1. Data set of banana prices over the years.

We have selected the data set of price of commodity goods particularly bananas.The motivation for selecting this particular data set is that The prices of commodity goods are influenced by a complex interplay of factors—ranging from seasonal production cycles and climatic variations to transportation costs, storage limitations, market demand, and government policies. Even small disruptions in the supply chain, such as natural disasters or geopolitical conflicts, can cause noticeable price fluctuations. This makes the task of tracking and analyzing commodity prices not only valuable but also necessary.Bananas, in particular, are among the most widely consumed fruits globally. Their affordability and availability make them a staple in many households. However, due to their perishable nature and sensitivity to climatic conditions, their prices tend to fluctuate significantly over time. This makes banana prices an ideal subject for time series analysis, which allows us to identify underlying patterns such as trends, seasonal effects, and irregular variations.Analyzing and monitoring the prices of such commodities is important for multiple reasons. For policymakers, it helps in designing effective agricultural support and subsidy systems. For farmers and producers, it provides insights into when to sell their produce to maximize returns.

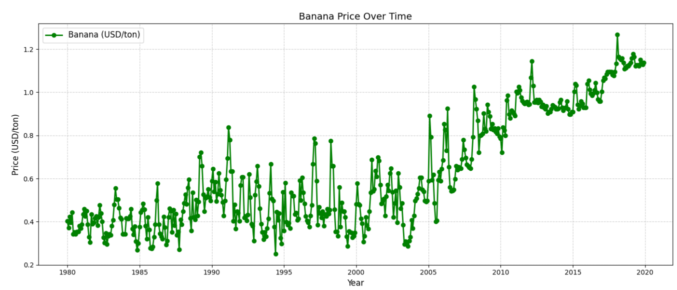

Fig. 2. Banana prices over the years.

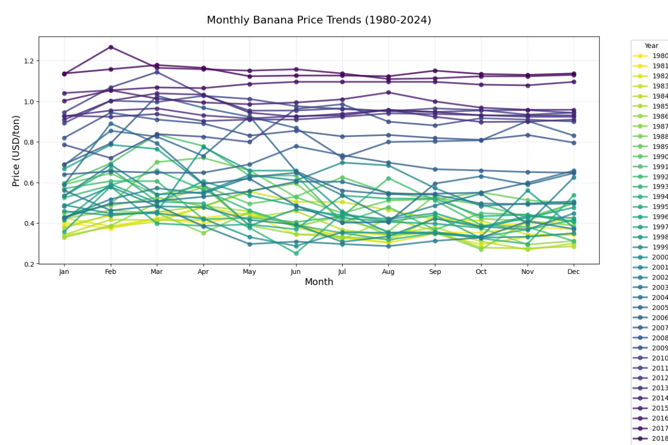

Fig. 3. Monthly Banana prices over the years.

B. Test for stationarity

Before applying any forecasting model, it is critical to check whether the time series is stationary—that is, whether its statistical properties like mean and variance remain constant over time.To assess stationarity, we used the Augmented Dickey-Fuller (ADF) test. When we applied the ADF test to our original banana price time series, we obtained a p-value of 0.94, indicating that the series was not stationary. A high p-value suggests the presence of a trend or seasonality that violates the stationarity assumption.Also the rolling mean is not flat and stable these observations suggests that the original data is not stationary.

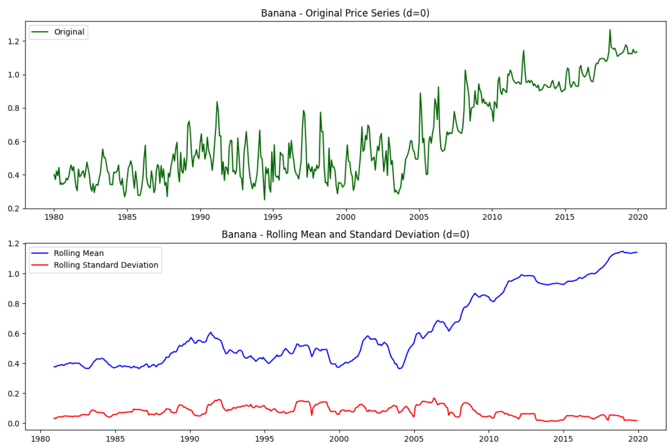

Fig. 4. Stationarity test for d=0

C. Decomposition of data

Decomposition is the first and most essential step in time series analysis. Any time series is generally composed of three key components: trend, which shows the long-term movement or direction; seasonality, which reflects repeating patterns at regular intervals (such as monthly or yearly); and residual component(includes random noise). Identifying and separating these components allows us to better understand the structure of the data and prepares it for more accurate modeling and forecasting.

So to make it stationary we started differencing the original series and applying ADF testing to find the p value and plot rolling mean and rolling std deviation. For d=1 we got p=1.6469369301394105e-25 and the plot of rolling mean became flat suggesting that the series was of order 1 i.e. d=1.

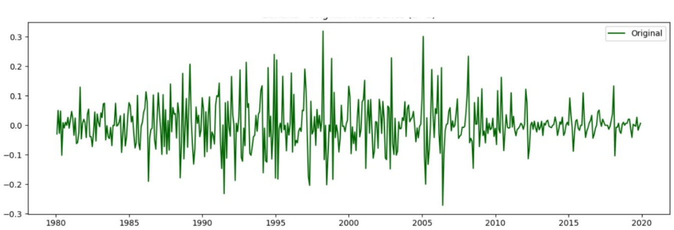

Fig. 5. The figure is the plot of time series data after applying method of differencing of order 1 which gives us the detrended data.

Even after detrending the time series, we see in Fig.6 , that our time series still has inconsistancy in variance and thus our data is non-stationary. Our original data seems to have

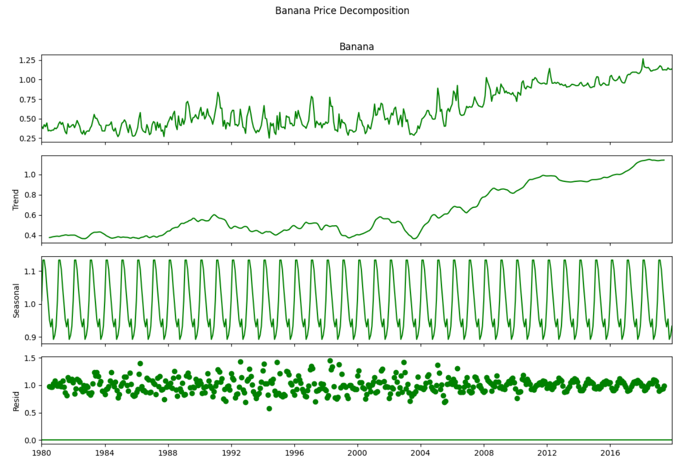

Fig. 6. Decomposition of original data

multiplicative seasonality, meaning that seasonality (variance) increases in magnitude over time.

This confirms that the data is multiplicative in nature i.e:

$$Y_t = U_t \cdot S_t \cdot X_t \tag{1}$$

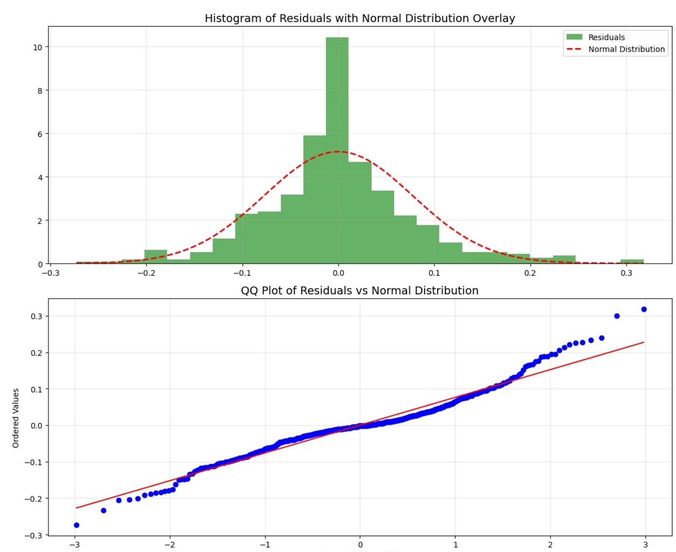

Fig. 7. Histogram and Q-Q plot

The histogram shows a symmetric bell-shaped distribution with a good match to the normal distribution overlay (red dashed line). The distribution appears slightly leptokurtic (more peaked than normal) with slightly heavier tails, but follows the normal curve closely overall.

The QQ plot shows an excellent alignment with the theoretical normal quantiles along most of the distribution. There are minor deviations at the extreme ends (particularly in the upper right), which suggests slightly heavier tails than a perfect normal distribution.

D. Autocorrelation

To find out which model is appropriate for our time series we used Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF), which helps in determining the parameters of ARIMA and SARIMA models.The PACF plot helps in selecting the order of the AR (AutoRegressive) component. For example, if the PACF shows a sharp cut-off after lag p, we consider using an AR model of order p.The ACF plot helps in identifying the order of the MA (Moving Average) component. If the ACF cuts off after lag q, we consider using an MA model of order q.

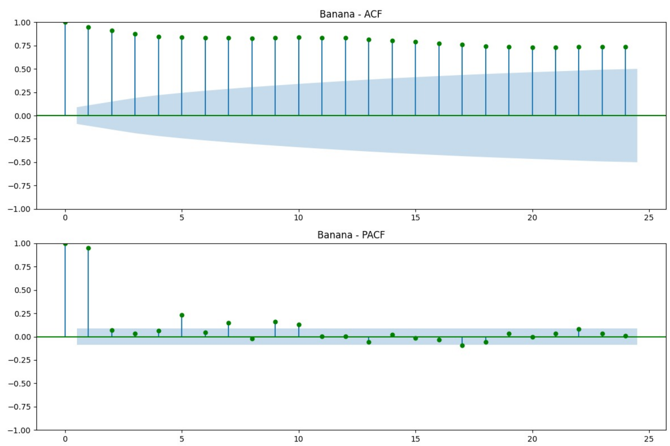

Fig. 8. ACF and PACF plots of Original data showing continues high values of autocorrelation

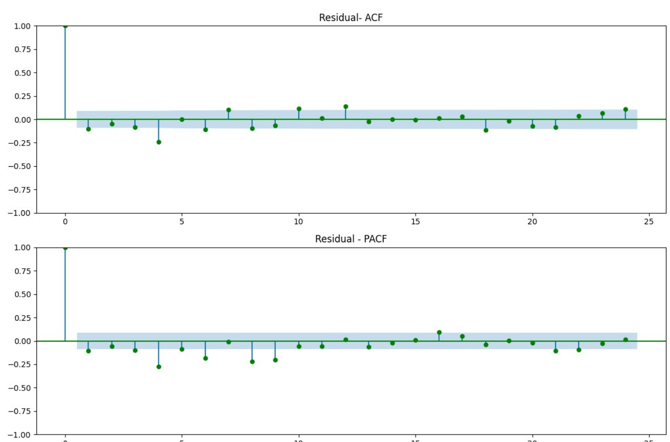

Fig. 9. ACF and PACF plots of data with d=1 suggests that the autocorrelation values are within the significance bands

The ACF plot of the original banana price series reveals strong persistence with exceptionally high autocorrelation values with minimal decay which confirms the need of differencing as suggested by outputs of ADF test. The ACF of the series after differencing shows that most of the autocorrelation values lie in the confidence interval and the absence of any pronounced pattern in these residual plots confirms that the multiplicative decomposition effectively extracted the primary components of trend and seasonality. The residuals centering around 1.0 (visible in your decomposition plot) rather than 0 further validates the appropriateness of your multiplicative model specification over an additive approach.

III. Model Selection

In our analysis of monthly banana prices, we evaluated several time series forecasting models—including AR(1), MA(1), ARMA, ARIMA, and the Random Walk model—to identify the most accurate and reliable method for modeling and predicting future prices. After comparing the models based on their performance metrics and visual diagnostics, the ARIMA(4,1,4) model was selected as the best fit for our dataset, as it yielded the lowest Root Mean Square Error (RMSE) and captured the underlying dynamics of the series more effectively than simpler models. The selection of ARIMA(4,1,4) was informed by the patterns observed in the ACF and PACF plots of the differenced series. The ACF showed significant autocorrelations at lag 4, and the PACF also displayed notable spikes at early lags, suggesting that both autoregressive and moving average components played an important role in the structure of the data. Additionally, the Augmented Dickey-Fuller (ADF) test confirmed that the original price series was non-stationary (with a p-value ¿ 0.05), which warranted the use of first-order differencing (d = 1) to stabilize the series. Overall, the ARIMA(4,1,4) model provided a strong and interpretable forecasting framework, making it a suitable choice for estimating banana prices with improved accuracy and reliability over time. This can be represented mathematically as:

$$\begin{aligned}\left(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3-\phi_4 B^4\right) X_t \\ =\left(1+\theta_1 B+\theta_2 B^2+\theta_3 B^3+\theta_4 B^4\right) \varepsilon_t \quad (2)\end{aligned}$$

Using Python libraries to fit the model, we obtained the parameters as:

AR coefficients: ϕ_1=0.0998 , ϕ_2=0.8537 , ϕ_3=-0.1346 , ϕ_4=-0.4186 .

MA coefficients as: θ_1= -0.3805 , θ_2=-0.9994 , θ_3=0.2603 , θ_4=0.3245.

IV. Conclusion

In this project, we analyzed monthly banana prices using time series techniques. After decomposing the data and testing for stationarity, we applied first-order differencing and used ACF and PACF plots to guide model selection. Among the models tested, ARIMA(4,1,4) provided the most accurate forecasts, effectively capturing both short-term changes and long-term trends. Given the seasonal nature of the data, we also considered the potential of SARIMA for future improvement. Overall, this analysis shows how time series models can be powerful tools for understanding and forecasting commodity prices like bananas.

Links

Click to view Google collab file