The Impact of Meteorological Factors on Crop Price Volatility in India: Case studies of Soybean and Brinjal

Ashok Kumar[1], **Abbinav Sankar Kailasam**[2], **Anish Rai**[3, 4], **Sudeep Shukla**[5], **Sourish Das**[3], **and Anirban Chakraborti**[1,*]

[1]School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India
[2]School of Computing and Data Science, Sai University, Chennai-603104, Tamil Nadu, India
[3]Chennai Mathematical Institute, Chennai-603103, Tamil Nadu, India
[4]Department of Physics, National Institute of Technology Sikkim-737139, Sikkim, India
[5]AI 4 Water LTD, Orpington, BR6 9QX United Kingdom
[*]anirban@jnu.ac.in

ABSTRACT

The price volatility of agricultural crops (commodities) is influenced by meteorological variables such as temperature and precipitation (and many other environmental, social and governance factors), which is a critical challenge in sustainable finance, agricultural planning, and policy-making. This paper studies the impact of meteorological variables on price volatility of agricultural crops. As case studies, we choose the two Indian states of Madhya Pradesh (for Soybean) and Odisha (for Brinjal). We employ an Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model to estimate the conditional volatility of the log returns of crop prices from 2012 to 2024. This study further explores the cross-correlations between volatility and the meteorological variables. Further, a Granger-causality test is carried out to analyze the causal effect of meteorological variables on the price volatility. Finally, the Seasonal Auto-Regressive Integrated Moving Average with Exogenous Regressors (SARIMAX) and Long Short-Term Memory (LSTM) models are implemented as simple machine learning models of price volatility with meteorological factors as exogenous variables. We believe that this will illustrate the usefulness of simple machine learning models in agricultural finance, and help the farmers to make informed decisions by considering climate patterns and making beneficial decisions with regard to crop rotation or allocations. In general, implementing meteorological factors to assess agricultural performance could help to understand and reduce price volatility and possibly lead to economic stability.

1 Introduction

A system, natural or man-made, consisting of many components that interact nonlinearly among themselves, is termed a complex system[1,2]. Complex systems like climate, hydrology, financial markets, etc., exhibit extreme events that have severe impacts on different sectors[3–5]. Understanding various complex systems is of utmost importance, as these systems and their interactions influence natural, economic, and social environments. Climate is one such complex system that has significant impacts on various fields such as agriculture, human health, and disaster management[6]. Even a slight change in the climatic behavior can have significant impacts; hence, it is essential to analyze the dynamics and influence of climate on various sectors.

Agriculture remains one of the sectors most vulnerable to climate variability, with farmers heavily reliant on meteorological variables like temperature and precipitation for crop production. However, unpredictable fluctuations in these factors pose substantial risks, disrupting planting and harvesting schedules, reducing yields, and leading to unexpected shortages or surpluses in agricultural output[7–10]. These climate-induced disruptions can have profound consequences for food security and income stability, particularly in regions where agriculture is a primary source of livelihood, such as developing countries[9,11,12].

The fluctuations in crop supply directly influence market prices, leading to increased price volatility that can destabilize local economies and threaten the financial well-being of smallholder farmers[9,10,13]. Extreme precipitation makes economic inequality worse in nations that rely heavily on agriculture. When precipitation rises by 1.5 standard deviations, low-income groups are disproportionately affected—35 times more so than in nations with less reliance on agriculture. Given this pattern, it is probable that wealth inequality will increase in agricultural areas, especially in Africa, calling for immediate attention to how climate change affects vulnerable groups[14]. Furthermore, the movement of nitrogen from land to aquatic systems is being disrupted by climate change, as rising temperatures in many parts of the United States may offset the benefits of greater precipitation on nitrogen runoff. The intricate relationship between environmental sustainability, economic inequality, and climate change is made clear by these revelations[15]. Long-term climate trends have reduced wheat and barley yields across Europe by 2.5% and 3.8%, with Mediterranean areas facing the most adverse effects. This suggests climate factors explain around 10% of yield stagnation, with policy and agricultural changes likely accounting for the rest[16]. By the end of the century, if global temperature rises above the ideal thresholds for crops (29°C for corn, 30°C for soybeans, and 32°C for cotton), the U.S., which produces 41% of the world's maize and 38% of its soybeans, may experience production declines of up to 82%[17]. The development of heat-tolerant crops is vital since warming of 2°C and 4°C would likely increase the variability of maize yields, threatening food security and the stability of the grain trade, particularly for the 800 million people living in extreme poverty[18]. Similarly, India's agriculture sector must adapt its cropping patterns for sustainable growth, particularly after COVID-19 intensified reliance on agriculture for livelihoods. In West Bengal, crop diversification, irrigation, fertilizer, and market access are driving transitions toward non-foodgrains, with climate factors like humidity and temperature playing a role. Long-term sustainability will require enhanced infrastructure, agricultural education, and decentralized support to address local dynamics effectively[19].

Price volatility in agricultural markets has been extensively studied in relation to market forces, but recent studies have increasingly emphasized the role of meteorological variables. Volatility models such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) have been used to analyze time series data, accounting for asymmetry in shocks affecting price returns. Research indicates that weather conditions, especially extreme weather events, significantly impact agricultural productivity and pricing. These conditions lead to fluctuations in supply, thereby affecting prices. For instance, studies utilizing GARCH-MIDAS frameworks have demonstrated that temperature anomalies and precipitation patterns are pivotal in understanding food price volatility. These models reveal that climate change can induce significant volatility in food prices, emphasizing the need for integrating meteorological variables into volatility modeling[9,10,13]. Forecasting price volatility through meteorological data offers a strategic approach to managing the risks posed by climate variability. By leveraging predictive models that incorporate meteorological variables, policymakers and financial institutions can develop more effective crop insurance schemes and financial tools to protect farmers from the adverse impacts of climate fluctuations[8,12,20]. Empirical results indicate that hybrid ARIMAX-LSTM models outperform traditional methods, achieving lower mean absolute percentage error (MAPE) values, thereby enhancing the accuracy of price forecasts and risk mitigation strategies[10,13].

This paper focuses on estimating the price volatilities of two critical agricultural commodities in India — Soybean in Madhya Pradesh, and Brinjal in Odisha using the EGARCH model. We selected these crops as they represent two different agriculture markets; Soybean being an export oriented oil seed and Brinjal bring a vegetable which is used for domestic use. Additionally, Soybean is a kharif crop, which is strictly sown during the onset of monsoon (June/July) and harvested during (October)[21,22]. On the other hand, Brinjal is cultivated all year round without seasonal cycles[23,24]. Further, we implement SARIMAX and LSTM models to forecast price volatility using the meteorological variables like maximum temperature and precipitation. This approach will help anticipate risk and offer insights into how tailored financial models can mitigate the socio-economic challenges posed by climate-driven agricultural volatility. The findings of the work will provide the following revelation: firstly, the influence of meteorological variables on the commodity prices, and secondly, which model proves a better fit for such analysis.

2 Data

2.1 Description

The Government of India's Directorate of Marketing and Inspection operates the AGMARKNET web portal, which plays a crucial role in disseminating agricultural market information, including arrivals and prices of various agricultural commodities across India. The portal collects data from local agricultural markets through a specially designed application called "Agmark," enabling farmers, traders, and researchers to access real-time market trends and insights. The website for accessing the corresponding data is: AGMARKNET. By providing reliable and transparent information, AGMARKNET aims to support informed decision-making among stakeholders in the agricultural sector, ultimately fostering improved market practices and facilitating effective price discovery. We have taken monthly data from eighteen districts in Madhya Pradesh and six districts in Odisha, covering the period from Jan-2012 to Oct-2024. The district were considered based on the data availability for this period.

The monthly data for precipitation and maximum temperature, measured at 2 meters above ground level from Jan-2012, to Oct-2024, corresponding to the crop price data for Brinjal in Odisha and Soybean in Madhya Pradesh, was sourced from NASA's POWER Project (Prediction Of Worldwide Energy Resources) and the website for the corresponding the data is NASAPOWER. This data service provides global meteorological and solar information to support applications in renewable energy, sustainable agriculture, and climate-related projects. Powered by NASA's Earth science satellite observations and reanalysis products, including MERRA-2 (Modern-Era Retrospective analysis for Research and Applications) and CERES (Clouds and the Earth's Radiant Energy System), the POWER project offers a comprehensive range of data for various sectors.

The project primarily focuses on delivering solar irradiance data to optimize solar energy systems, alongside climate variables like temperature, precipitation, humidity, and wind speed, which are critical for crop modeling and weather monitoring in agriculture. It provides global data at high spatial resolution, with options for daily to annual time steps. This data can be accessed via an easy-to-use web interface or programmatically through an API, enabling customized dataset retrieval for specific locations and timeframes.

2.2 Preprocessing, exploratory analyses and visualization

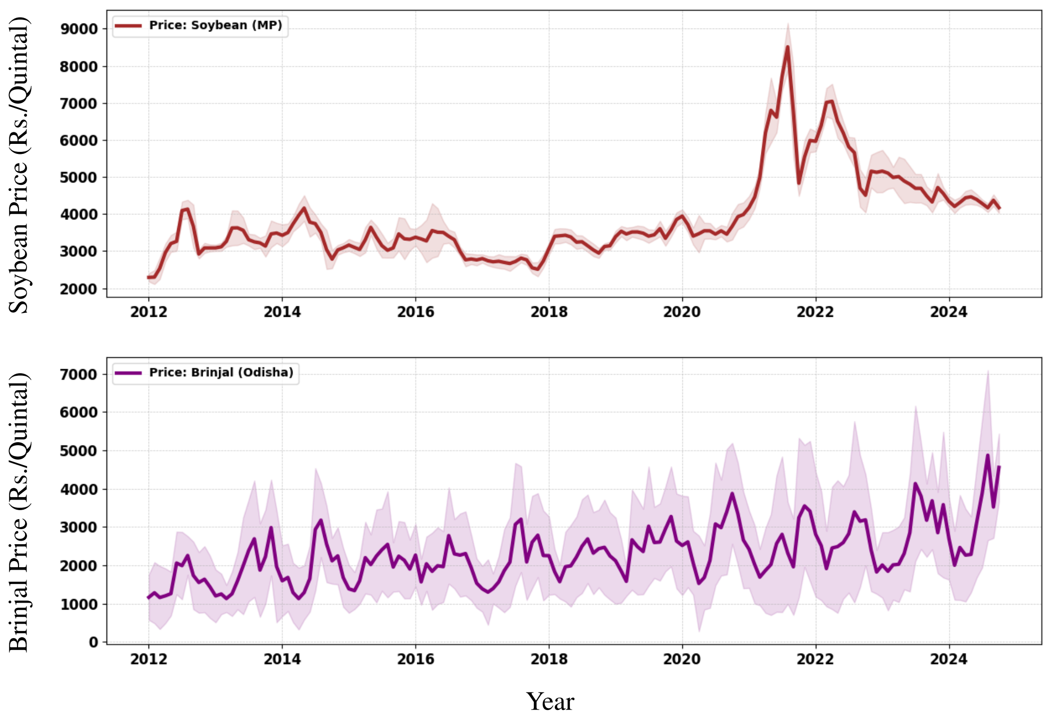

Figure 1. Monthly Price Analysis. The plots represent the monthly price time series for Soybean and Brinjal (in Rs./Quintal) over the period 2012 - 2024, respectively. The plots show soybean price trends in Madhya Pradesh (MP) and brinjal in Odisha, highlighting seasonal peaks and volatility patterns.

In order to carry out state-level analysis, we have taken the daily average of the commodity prices (soybean in Madhya Pradesh and brinjal in Odisha) from different districts. Further, we have calculated the monthly average for each state which is shown in Fig. 1. The figure represents the monthly price (Rs./Quintal) of soybean in Madhya Pradesh (MP) and brinjal in Odisha from 2012 to 2014. The band (shown by a lighter color) is the one standard deviation price band from the price. Subsequently, we have calculated the log-return of the commodity price which is defined as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1}).$$

where r_t is the log return at time t, P_t is the price of the commodity at time t, and P_{t-1} is the price at time $t-1$.

Figure 2 represents the log return of monthly prices for Soybean and Brinjal, respectively. Log return is used to measure the percentage change in prices, offering a clearer view of relative price movements over time, smoothing out seasonal variations. These plots reflect the price dynamics and volatility, aiding in understanding risk and return in the agricultural markets.

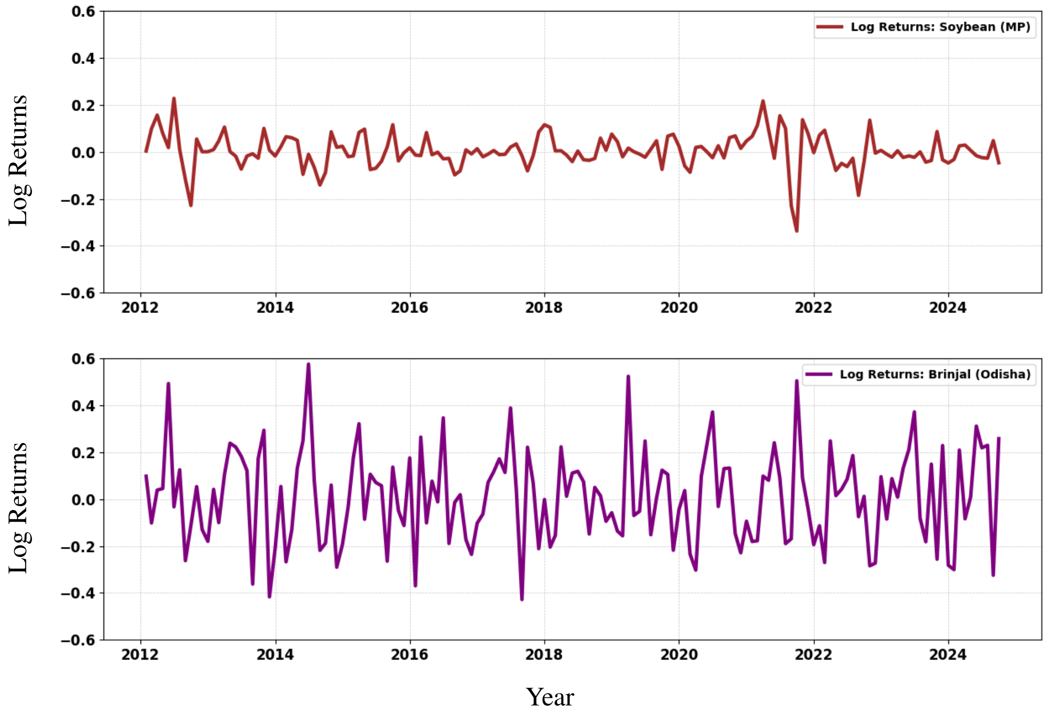

Figure 2. Log Return Analysis. The plots represent the monthly log returns of soybean and brinjal prices respectively. The log returns showcase potential market shocks in the past decade.

Figure 3 displays monthly precipitation and maximum temperature data for Madhya Pradesh and Odisha. All four plots exhibit pronounced annual seasonal patterns. The precipitation plots (in blue) show significant spikes corresponding to India's monsoon season, when rainfall is highest. The maximum temperature plots (in red) demonstrate clear yearly cycles, with peak temperatures exceeding 40°C.

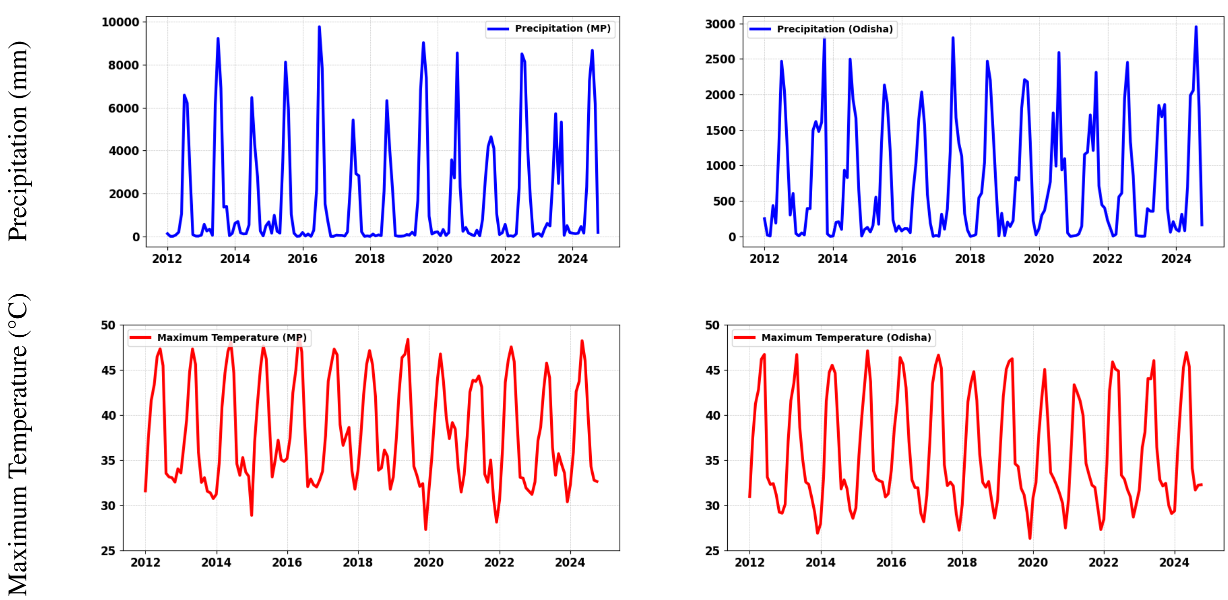

Figure 3. Meteorological Variables: The plots represent the monthly precipitation and maximum temperature from 2012 to 2024 for Madhya Pradesh (MP) and Odisha.

Table 1. Statistical Summary for Soybean and Brinjal Prices and its Log Returns

Parameter	Soybean Prices		Brinjal Prices	
	Price	Log Returns	Price	Log Returns
Observations	154	153	154	153
Mean	3875.51	0.0039	2302.55	0.0089
Standard Deviation	1136.18	0.0737	716.79	0.1992
Minimum	2289.26	-0.3373	1128.12	-0.4284
Maximum	8511.76	0.2273	4872.30	0.5752
Skewness	-	-0.6102	-	0.2302
Kurtosis	-	3.7131	-	-0.1736

Table 1 shows the statistical summary for soybean and brinjal monthly prices. The statistics is based on a consistent sample size of 154 observations for each crop. The mean price for soybean is Rs. 3875.51, indicating a significantly higher average compared to brinjal prices, which average is Rs. 2302.55. This disparity highlights a relatively greater market value of soybean. In terms of log returns, soybean exhibit a mean of 0.0039, while for brinjal it is 0.0089, suggesting both crops have shown modest increase in value over time. The standard deviation, which measures price variability, reveals that soybean prices are more volatile, with a standard deviation of 1136.18 compared to 716.79 for brinjal. For log returns, the standard deviations are 0.0737 for soybean and 0.1992 for brinjal, indicating that log return for brinjal exhibits greater variability. The price range for soybean spans from 2289.26 to 8511.76, while brinjal prices range from 1128.12 to 4872.30. This broader price range for soybean suggests a more diverse market response, however the relative change is higher for brinjal. Distribution characteristics further illuminate the behavior of returns. Soybean log returns show a negative skew of -0.6102, indicating a tendency for extreme negative returns, which may signal potential downside risks. In contrast, brinjal log returns exhibit a positive skew of 0.2302, reflecting a slight tendency for extreme positive returns, which could indicate more upside potential. The kurtosis value for soybean log returns is 3.7131, indicating a leptokurtic distribution characterized by a sharper peak and fatter tails, suggesting a higher likelihood of extreme values. Conversely, brinjal log returns have a kurtosis of -0.1736, indicating a flatter distribution with less probability of extreme returns. However, as the value is very close to zero the behavior will be similar to normal distribution.

3 Methodology and Results

3.1 EGARCH Model for Price Volatility Estimation

We begin our analysis by fitting the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model to the log-return of soybean and brinjal price data across Madhya Pradesh and Odisha respectively.

3.1.1 Methodology

The EGARCH model, an extension of the GARCH family, is particularly suited for financial time series with asymmetric volatility. It captures the so-called "leverage effect," where negative shocks to the market result in larger increases in volatility compared to positive shocks of the same magnitude.The EGARCH model is defined by the following equations for the conditional variance σ_t^2:

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \left(\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{j=1}^{o} \gamma_j \frac{\varepsilon_{t-j}}{\sigma_{t-j}} + \sum_{k=1}^{q} \beta_k \ln \sigma_{t-k}^2$$

Here,

- ω: The constant term representing the baseline level of the log conditional variance.
- α_i (for $i = 1, \dots, p$): These parameters capture the impact of the magnitude of past shocks on current volatility. The term

  $$\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}}$$

  represents the centered absolute standardized residual, where $\sqrt{\frac{2}{\pi}}$ is subtracted so that its expected value is zero under a normal distribution.

- γ_j (for $j = 1, \ldots, o$): These coefficients measure the asymmetric or leverage effect. They allow the model to differentiate between the impacts of positive and negative shocks on volatility.

- β_k (for $k = 1, \ldots, q$): These parameters model the persistence in volatility by incorporating the effect of past conditional variances, expressed in logarithmic form.

- ε_t: The residuals (or innovations) from the mean equation, representing the unexpected shocks at time t.

- σ_t: The conditional standard deviation (volatility) at time t, with σ_t^2 being the conditional variance.

- p, o, q: The orders of the model corresponding to the number of lagged terms for the absolute standardized shocks, the asymmetric shocks, and the lagged log variances, respectively.

3.1.2 Results

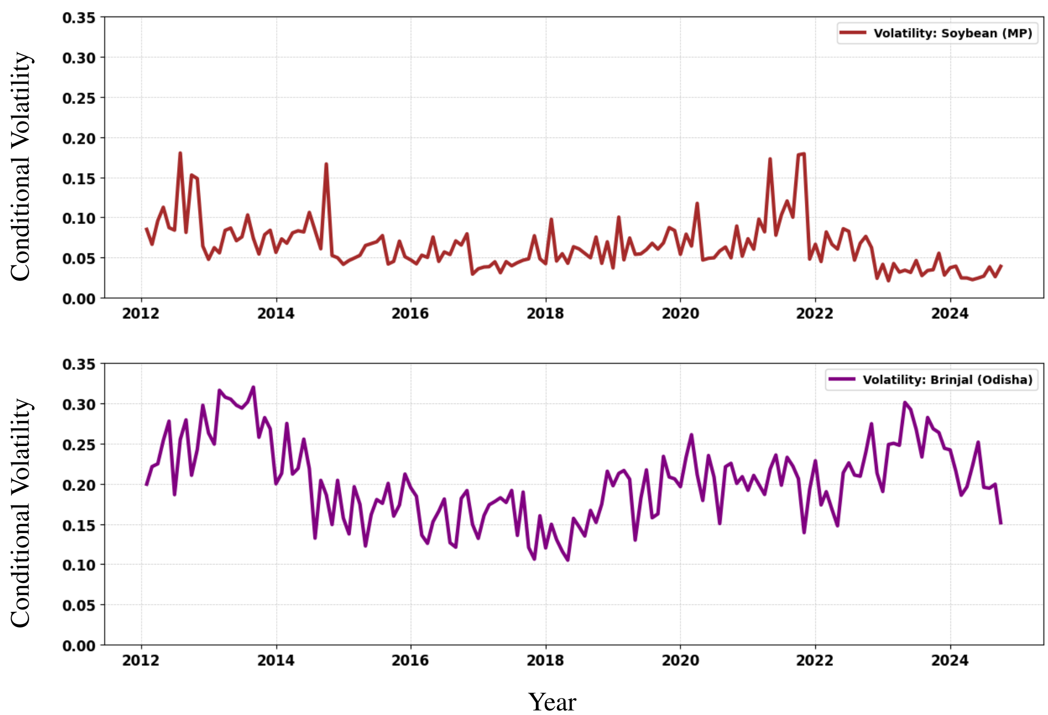

Figure 4. Conditional Volatility: The plots represent the conditional volatility of monthly log returns of Soybean and Brinjal prices from 2012 to 2024 respectively.

Figure 4 illustrates the conditional volatility of monthly log returns using the EGARCH model for Madhya Pradesh and Odisha, effectively capturing the volatility patterns in the prices of soybean and brinjal in these states. From the figure, we observe that soybean experienced a significant spike in volatility during late 2021 when the impact of COVID-19 pandemic was still fresh. As it is an export-oriented oilseed, soybean was heavily affected by the aftereffects of the lockdown. These factors impacted their supply-demand chain resulting in extreme volatility in their prices, which is evident from Fig. 2 as well. Significant volatility spike in soybean price is observed during 2013 as well which was due to weakness in rupee and the soybean import demand from Iran[25]. On the other hand, as brinjal is a vegetable which is mainly consumed in the domestic market, it does not show such drastic spike in volatility. This shows that the sensitivity towards various events are different for different crops. Table 2 contains the parameter estimates obtained from the EGARCH model fitted on crop price volatility.

Table 2. Parameter Estimates for Madhya Pradesh and Odisha

Parameter	Madhya Pradesh		Odisha	
	Estimate	Std. Err	Estimate	Std. Err
μ (Mean)	0.00063	0.009570	0.010211	1.21e-10
ω (Constant)	-0.663758	0.206315	-0.433305	1.90e-11
α_1 (ARCH Coefficient)	0.625088	0.234577	-0.569581	2.57e-10
α_2 (ARCH Coefficient)	-0.012992	0.102649	-	-
α_3 (ARCH Coefficient)	-0.862315	0.320954	-	-
γ_1 (Leverage Effect)	-0.004637	0.210729	0.012321	1.14e-11
γ_2 (Leverage Effect)	-0.071931	0.222575	-	-
γ_3 (Leverage Effect)	0.490231	0.173930	-	-
β_1 (GARCH Coefficient)	0.091624	0.215075	0.398459	1.01e-12
β_2 (GARCH Coefficient)	0.801353	0.189453	0.038338	5.31e-12
β_3 (GARCH Coefficient)	-	-	0.430995	1.33e-11

3.2 Cross-Correlation and Granger-causality

3.2.1 Methodology

To understand the dynamic relationship between conditional volatility and meteorological variables (maximum temperature and precipitation) we calculate the cross-correlation between them at different lags. Cross-correlation analysis is a statistical tool that measures the correlation between two time series as a function of the time-lag applied to one of them. This technique allows us to quantify the extent to which past climate conditions influence current volatility, which is critical for forecasting future price movements based on meteorological trends.

Given two time series: the first representing the conditional volatility obtained from EGARCH, denoted as X_t, and the second representing the meteorological variables (maximum temperature or precipitation), denoted as Y_t, the cross-correlation function (CCF) is computed as:

$$\text{CCF}(k) = \frac{\sum_{t=k+1}^{T}(X_t - \bar{X})(Y_{t-k} - \bar{Y})}{\sqrt{\sum_{t=k+1}^{T}(X_t - \bar{X})^2 \sum_{t=k+1}^{T}(Y_{t-k} - \bar{Y})^2}}$$

where k is the lag applied to the second time series Y_t, $\bar{X}$ and $\bar{Y}$ are the mean values of the time series X_t and Y_t, respectively, and T is the length of the time series. The cross-correlation function measures the correlation of volatility on meteorological variables at different lag values. CCF indicates the relationship between two variables, however, it does not comment anything of their cause-effect relation or whether one variable holds the information to forecast another. To understand this relationship, the Granger-causality test is required.

Granger-causality test is a statistical test to identify the cause and effect of one time series on another[26]. We estimate the Granger-causality between volatility and past meteorological variables to analyze the cause of the meteorological variables on the volatility. In order to estimate the cause of time series Y on time series X, a regression is performed. X is modeled from it's own lagged values and the lagged values of Y. Further, we evaluate the significance of coefficient linked with Y to check if they are useful for forecasting X.

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \ldots + \alpha_k X_{t-k} + \beta_1 Y_{t-1} + \ldots + \beta_k Y_{t-k} + \varepsilon_t \tag{1}$$

where k is the no. of lags, α and β are the coefficients of the lagged values of X and Y, respectively, and ε is the prediction error.

Granger-causality employs hypothesis testing to evaluate the significance of the coefficients by computing the p value[26]. The null hypothesis states that X does not cause Y indicating that the lagged values does not hold any information to improve the forecasting. If the p value is greater than 0.05 we accept the null hypothesis. However, we reject the null hypothesis if the p value is less than 0.05.

3.2.2 Results

The cross-correlation analysis measures how volatility is related to different lagged values of meteorological variables. Figure 5 shows the cross-correlation between volatility with precipitation and temperature, respectively. We have calculated the cross-correlation for different lags from -60 to +60. The vertical dashed line signifies the correlation at lag 0. From the correlation plot between volatility and precipitation, we observe a noticeable rise in correlation at lag 3 and 4. This suggests that there is a positive relationship between volatility and precipitation at lag 3 and 4. Similarly, a relatively high correlation is observed between the volatility and maximum temperature at lag 4. This shows that around lag 3/4 the correlation between volatility and meteorological variables are relatively high. This high correlation is repeating at a lag of 12 which shows the annual cycle as each lag signifies one month period. It would be interesting to know the cause-effect relationship of these variables. It is interesting to observe that there is no such seasonal correlation between volatility and meteorological variables in the case of brinjal in Odisha, as seen in fig. 6. From cross correlation, no conclusive relation is obtained for the case of brinjal.

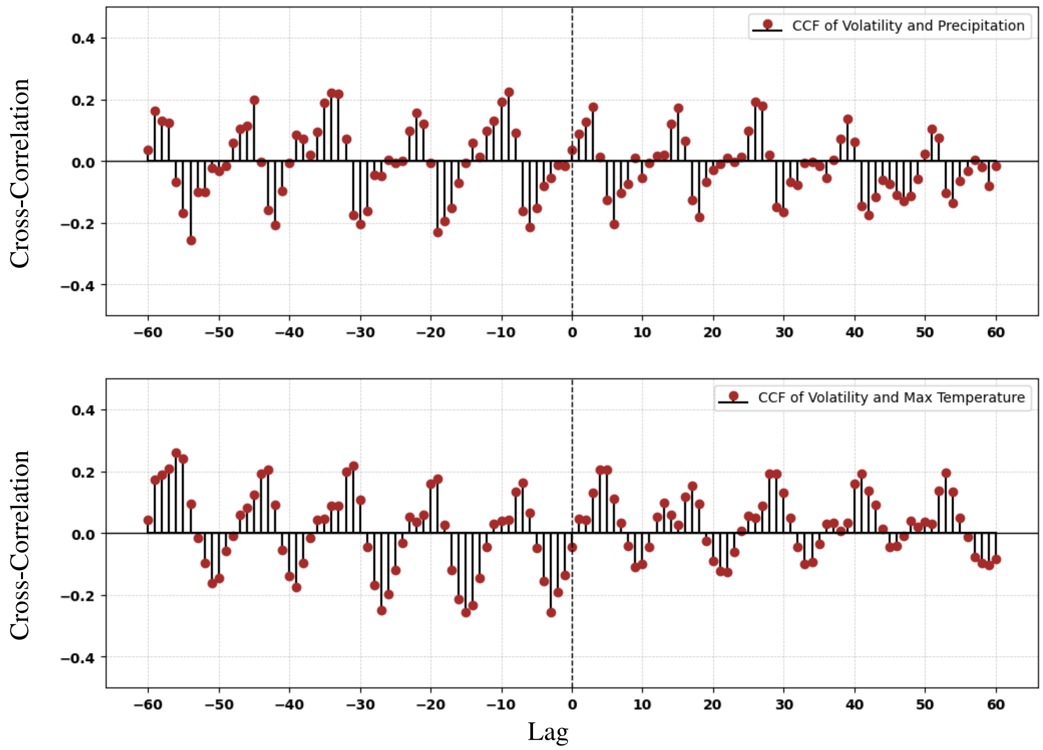

Figure 5. **Cross-Correlation Analysis**. Cross-correlation analysis with 60 lags for Madhya Pradesh (MP): First is for cross-correlation between Volatility and Precipitation, and second is for cross-correlation between Volatility and Maximum Temperature with varying time lags.

After the estimation of CCF, it is important to analyze the cause-effect relation between volatility and the meteorological variables. In order to check for a causal relation between volatility and meteorological variables, we conducted a Granger-causality test for both price time series. We first checked the stationarity using Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron tests for volatility, maximum temperature, and precipitation for both cases. We found that both time series are stationary. Hence, we estimated the Granger causality to test whether maximum temperature and precipitation have an effect on the volatility. We found that for Madhya Pradesh, there exists a causal relation between the volatility and meteorological variables at lag 3. The lag of 3 months is consistent with the sowing and harvest of the soybean crop. It is sown during the onset of the monsoon in the month of June/July and harvested during October. However, in the case of Odisha, we see no causal relation between the volatility and meteorological variables. This may be due to the fact that brinjal is grown all year round and is not specific to certain conditions. However, the meteorological variables play an important role in the yield and price of brinjal. So, we use the meteorological variables as inputs to forecast the volatility using ARIMAX and LSTM.

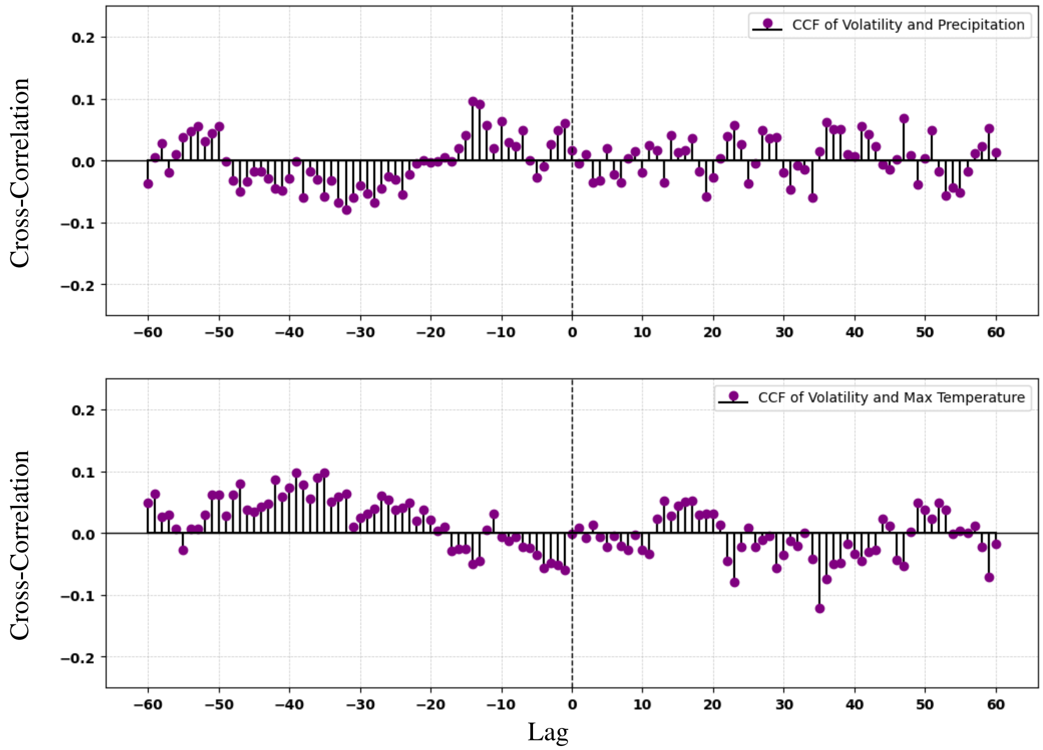

Figure 6. Cross-Correlation Analysis. Cross-correlation analysis with 60 lags for Odisha: the first plot is for cross-correlation between Volatility and Precipitation, and the second is for cross-correlation between Volatility and Maximum Temperature, each with varying time lags.

3.3 Predictive Modeling (SARIMAX and LSTM)

3.3.1 Methodology

The third stage of our analysis focuses on developing predictive models that forecast price volatility by incorporating meteorological inputs. To achieve this, we employ two distinct modeling approaches: the **SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous variables)** model and the **LSTM (Long Short-Term Memory)** neural network model. Each method offers unique advantages in capturing the temporal dynamics and complex relationships between climatic variables and agricultural price volatility.

SARIMAX Model The SARIMAX model expands upon the traditional ARIMA framework by incorporating both seasonal components and external variables, making it particularly well-suited for agricultural price analysis. The general form of the SARIMAX model can be written as:

$$\Phi_P(L^s)\phi_p(L)(1-L)^d(1-L^s)^D\sigma_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t + \mathbf{x}_t\beta$$

where:

- $\Phi_P(L^s)$: Seasonal autoregressive operator of order P, defined as $\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \cdots - \Phi_P L^{Ps}$.
- $\phi_p(L)$: Non-seasonal autoregressive operator of order p, defined as $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$.
- d: Order of non-seasonal differencing.
- D: Order of seasonal differencing.
- s: Seasonal period (e.g., $s = 12$ for monthly data with yearly seasonality).
- $\Theta_Q(L^s)$: Seasonal moving average operator of order Q.
- $\theta_q(L)$: Non-seasonal moving average operator of order q.

- ε_t: Error term.
- $\mathbf{x}_t$: Vector of exogenous regressors (e.g., maximum temperature *tasmax* and precipitation *pr*), where `tasmax` is the maximum atmospheric temperature, and `pr` is precipitation.
- β: Coefficient vector for exogenous regressors.

The SARIMAX model captures the linear relationships between the dependent variable (price volatility) and both its own past values and external meteorological variables while also accounting for seasonal patterns. By incorporating these exogenous variables, the model can predict future volatility based on past patterns in the data and concurrent climate conditions. By incorporating these exogenous variables, the model can predict future volatility based on past patterns in the data and concurrent climate conditions. The model fitting involves selecting appropriate values for p, q, d, P, Q and D where d and D represent the differencing and seasonal differencing order to make the time series stationary when needed. To determine the optimal model parameters, we optimized the **Akaike Information Criterion (AIC)**, which helped balance model complexity with predictive accuracy.

LSTM Model To capture non-linear and complex dynamics between climatic variables and price volatility, LSTM model is used. The LSTM model is a type of recurrent neural network (RNN) that is designed to handle long-term dependencies in sequential data, making it well-suited for time series forecasting in agricultural markets, where the effects of weather patterns on prices may exhibit delayed interactions.

The model consists of a series of memory cells, each containing three main components: the input gate, the forget gate, and the output gate. These gates control the flow of information into and out of the memory cells, allowing the network to "remember" important information for long periods of time and "forget" irrelevant information. The mathematical formulation of an LSTM unit is as follows:

The Forget Gate in an LSTM model decides what information should be discarded from the memory cell, computed as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f),$$

where f_t is the forget gate vector, σ is the sigmoid activation function, W_f are the weight matrices, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias. The Input Gate determines what new information should be added to the memory, calculated as

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{and} \quad \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

where i_t is the input gate vector and $\tilde{C}_t$ is the candidate cell state. The cell state is updated as

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t,$$

combining the previous state with the new information. Finally, the Output Gate produces the final output based on the updated cell state, calculated as

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{and} \quad h_t = o_t \cdot \tanh(C_t)$$

where o_t is the output gate vector, and h_t is the hidden state (the final output of the LSTM unit at time t).The LSTM model is trained on sequences of past price volatility and meteorological data (maximum temperature and precipitation). Its architecture enables it to learn both short-term and long-term dependencies, making it highly effective in capturing the delayed and intricate effects of climate variability on agricultural price volatility.

3.3.2 Results

This section contains the forecast results of the conditional volatility using the SARIMAX and LSTM models. We use the meteorological variables (i.e., precipitation and maximum temperature) as independent variables. Each model is tested on two crops from different states: soybean from Madhya Pradesh, and brinjal from Odisha. We use mean absolute percentage error (MAPE) to estimate the predictive accuracy of the models.

Figure 7 represents the conditional volatility plots for soybean and brinjal, respectively. The solid lines, brown for soybean and purple for brinjal, show the conditional volatility obtained from the EGARCH model. The dashed lines, orange for soybean and green for brinjal, are the volatility predicted from the SARIMAX model during the testing period. We obtained a MAPE of 0.55 and 0.19 for soybean and brinjal, respectively. The result shows that forecasting error is higher for the price volatility of soybean. This may be because the volatility during the testing period (Sep 2019 - Oct 2024) was more dominated by the external factors such as COVID-19 than the meteorological factors. As earlier discussed as soybean is an export-oriented crop it is impacted by global disruptions. On the other hand brinjal is not affected by the disruptions hence, we obtain a lower MAPE.

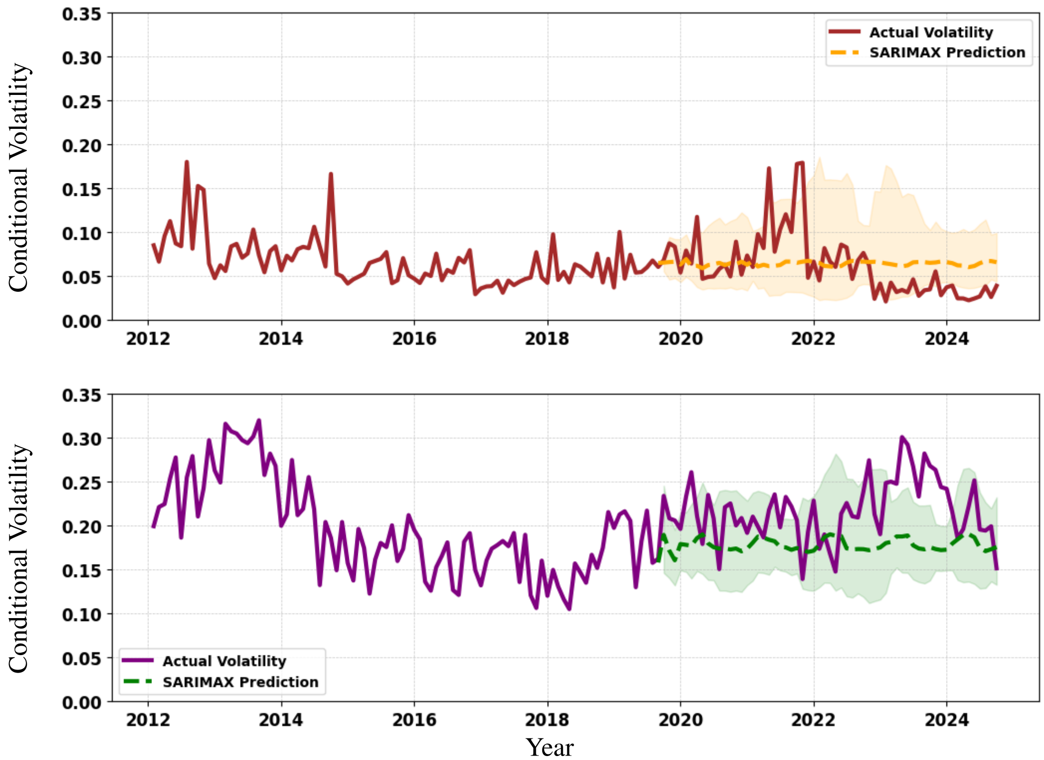

Figure 7. Comparison of SARIMAX Forecasting Models. The top plot illustrates the conditional volatility of log returns for Soybean prices in Madhya Pradesh (MP), while the bottom plot shows the conditional volatility of log returns for Brinjal prices in Odisha. In the first plot, the orange line represents the SARIMAX model's prediction for Soybean price conditional volatility during the testing period from September 2019 to October 2019. Similarly, in the second plot, the green line represents the SARIMAX prediction for Brinjal price conditional volatility over the same testing period.

Similarily, in figure 8 the solid lines, brown for soybean and purple for brinjal, show the conditional volatility obtained from the EGARCH model. The dashed lines, orange for soybean and green for brinjal, are the volatility predicted from the LSTM model during the testing period. We obtained a MAPE of 0.45 and 0.21 for soybean and brinjal, respectively. Due to similar reasons as stated above, the forecasting error may be higher for the price volatility of soybean.

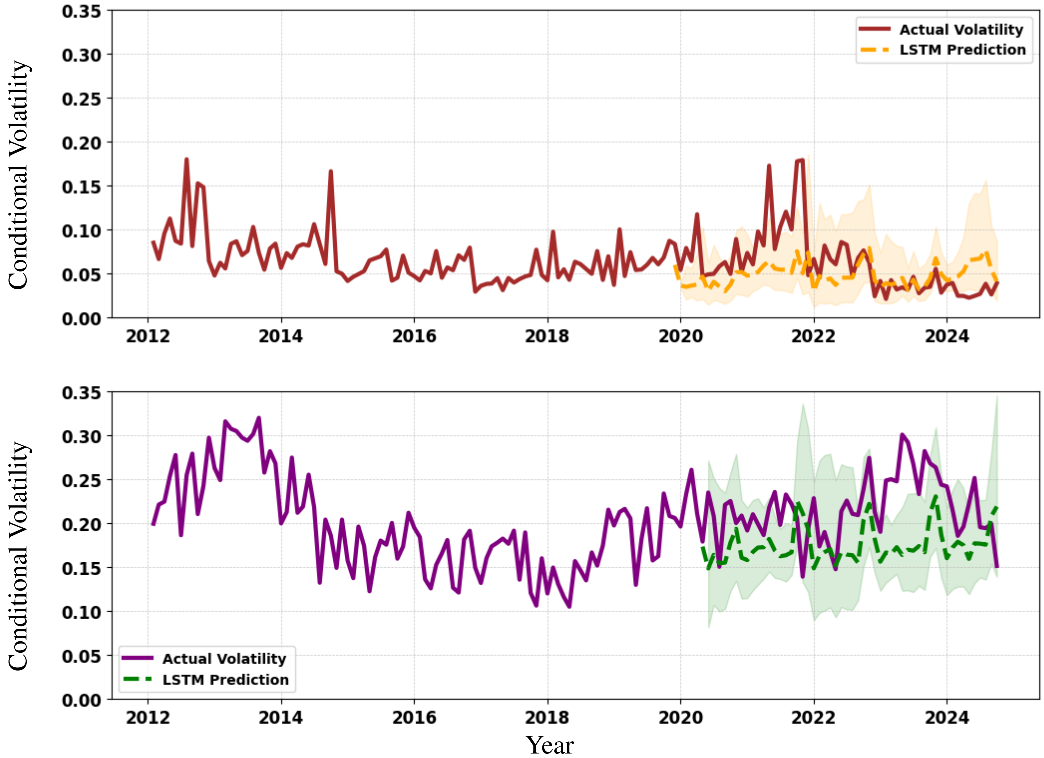

Figure 8. LSTM Forecasting Models Comparison.The top plot illustrates the conditional volatility of log returns for Soybean prices in Madhya Pradesh (MP), while the bottom plot shows the conditional volatility of log returns for Brinjal prices in Odisha. In the first plot, the orange line represents the LSTM model's prediction for Soybean price conditional volatility during the testing period from December 2019 to October 2019. Similarly, in the second plot, the green line represents the LSTM prediction for Brinjal price conditional volatility over the same testing period.

State	Crop	Model	MAPE
Madhya Pradesh	Soybean	SARIMAX	0.55
		LSTM	0.45
Odisha	Brinjal	SARIMAX	0.19
		LSTM	0.21

Table 3. RMSE and MAPE Comparison for Different Models

4 Discussions

We obtain various implications from the results obtained from our work. The analysis of EGARCH shows that the two crops exhibit significantly different volatility patterns due to the nature of the product, their market demand, and their sensitivity towards external factors. We found that soybean is more sensitive, as it is an export-oriented product that is susceptible to global disruptions. On the other hand, brinjal, being a vegetable, is mostly consumed in the domestic market, and hence, it is not volatile despite global disruptions. This is also supported by the statistical analysis, which showed that for soybean, the skewness and kurtosis indicate a higher chance of extreme movements than brinjal. From these results, we could say that policy for export-specific crops should be framed keeping into consideration their sensitivity towards external factors.

The CCF results indicated that the volatility of soybean is more correlated to the meteorological variables at lags 3 and 4, indicating that these variables impact the price volatility. The lag of 3 and 4 is also consistent with the sowing period (June/July) and harvesting period (October) of the soybean. This is further justified by the Granger-causality test, which shows a causal relationship between the meteorological variables and volatility at lag 3. This clearly shows that soybean is quite sensitive to weather conditions. Brinjal, on the other hand, shows no significant relation with the meteorological variables. Brinjal is grown throughout the year and does not have strict seasonal cycles; hence, the low correlation and absence of causal relation are meaningful. However, meteorological variables play a crucial role in the production and pricing of brinjal.

The application of models such as SARIMAX and LSTM will help to forecast future price volatility based on meteorological variables. Also, it will help us choose the better model for such analysis. The supremacy of one model is hard to find from our analysis as for soybean LSTM performs better, and for brinjal, LSTM performs better based on the MAPE. However, we clearly see that the overall MAPE for brinjal is lower than for soybean. This is consistent with the results obtained from volatility, which indicated that soybean is more volatile around 2020 due to the COVID-19 effect. This clearly shows that apart from the meteorological variables, which are important for forecasting, other factors also affect the price volatility and should be included in the model to improve the accuracy.

This work will have more implications, particularly in areas that are more affected by climate change. The forecasts will help farmers optimize their sowing and harvesting time. They can make rational decisions by considering how the weather patterns will impact prices. They can also make a beneficial decision regarding crop rotation and allocations. In general terms, implementing meteorological factors to assess agricultural performance could help to understand and reduce price volatility that may result in economic stability.

5 Summary

Our study highlights the impact of meteorological variables on the price volatility of soybean and brinjal. Price volatility is estimated using the EGARCH model which shows that soybean in comparision to brinjal shows more sensitivity towards global disruptions like COVID-19. The export-oriented nature of soybean is the main reason for such sensitivity. On the other hand, as brinjal is a vegetable mainly focused on domestic market is less sensitive to external disruptions. This result also aligns with the statistics of log-return which shows that soybean has higher skewness and kurtosis indicating a higher change of extreme movements. Further, CCF and Granger-causality test shows that price volatility of soybean, which is a kharif crop, is significantly impacted by the meteorological variables at lag 3 and 4. This lag is consistent with the period between sowing and harvesting of soybean. We found that brinjal is not significantly impacted by the meteorological variables. The finding is also consistent as it is cultivated almost all year round and is not dependent on a particular season for cultivation. Lastly, SARIMAX and LSTM is applied to forecast the conditional volatility using the meteorological variables. The forecasting of price volatility of soybean is less accurate than brinjal. This may be because of the higher volatility in soybean during 2020.

Our findings could be beneficial in areas more vulnerable to climate change. It may also help policy-makers to make policies which are robust to external disruptions. Also, in future other variables in addition to meteorological variables may be incorporated in the model to understand the price dyanmices better and also improve the accuracy.

Acknowledgements

The authors thank Himani Gupta for downloading and preprocessing the data of meteorological variables.

References

1. Chakrabarti, A. S., Bakar, K. S. & Chakraborti, A. *Data science for complex systems* (Cambridge University Press, 2023).
2. Kwapień, J. & Drożdż, S. Physical approach to complex systems. *Phys. Reports* **515**, 115–226 (2012).
3. Chakraborti, A. *et al.* Emerging spectra characterization of catastrophic instabilities in complex systems. *New J. Phys.* **22**, 063043 (2020).
4. Chakraborti, A., Sharma, K., Pharasi, H. K. *et al.* Phase separation and scaling in correlation structures of financial markets. *J. Physics: Complex.* **2**, 015002 (2020).
5. Rai, A., Nath Sharma, B., Rabindrajit Luwang, S., Nurujjaman, M. & Majhi, S. Identifying extreme events in the stock market: A topological data analysis. *Chaos: An Interdiscip. J. Nonlinear Sci.* **34** (2024).
6. Rind, D. Complexity and climate. *Science* **284**, 105–107 (1999).
7. Kumar, R. & Gautam, H. R. Climate change and its impact on agricultural productivity in india. *J. Climatol. & Weather. Forecast.* **2**, 1–3 (2014).
8. Zhao, C. *et al.* Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci.* **114**, 9326–9331, DOI: 10.1073/pnas.1701762114 (2017).
9. Burney, J. & Ramanathan, V. Recent climate and air pollution impacts on indian agriculture. *Proc. Natl. Acad. Sci.* **111**, 16319–16324, DOI: 10.1073/pnas.1317275111 (2014).

10. Burney, J., McIntosh, C., Lopez-Videla, B., Samphantharak, K. & Maia, A. G. Empirical modeling of agricultural climate risk. *Proc. Natl. Acad. Sci.* **121**, e2215677121, DOI: 10.1073/pnas.2215677121 (2024).

11. Birthal, P., Khan, M. T., Negi, D. & Agarwal, S. Impact of climate change on yields of major food crops in india: Implications for food security. *Agric. Econ. Res. Rev.* **27**, 145, DOI: 10.5958/0974-0279.2014.00019.6 (2014).

12. Carleton, T. A. Crop-damaging temperatures increase suicide rates in india. *Proc. Natl. Acad. Sci.* **114**, 8746–8751, DOI: 10.1073/pnas.1701354114 (2017).

13. Isah, K. & Muse, B. Modelling the volatility inducement of climate change in food prices: The role of technological shocks. *Asian Econ. Lett.* **5**, DOI: 10.46557/001c.115719 (2024).

14. Palagi, E., Coronese, M., Lamperti, F. & Roventini, A. Climate change and the nonlinear impact of precipitation anomalies on income inequality. *Proc. Natl. Acad. Sci.* **119**, e2203595119, DOI: 10.1073/pnas.2203595119 (2022).

15. Zhao, G., Merder, J., Ballard, T. C. & Michalak, A. M. Warming may offset impact of precipitation changes on riverine nitrogen loading. *Proc. Natl. Acad. Sci.* **120**, e2220616120, DOI: 10.1073/pnas.2220616120 (2023).

16. Moore, F. C. & Lobell, D. B. The fingerprint of climate trends on european crop yields. *Proc. Natl. Acad. Sci.* **112**, 2670–2675, DOI: 10.1073/pnas.1409606112 (2015).

17. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to u.s. crop yields under climate change. *Proc. Natl. Acad. Sci.* **106**, 15594–15598, DOI: 10.1073/pnas.0906865106 (2009).

18. Tigchelaar, M., Battisti, D. S., Naylor, R. L. & Ray, D. K. Future warming increases probability of globally synchronized maize production shocks. *Proc. Natl. Acad. Sci.* **115**, 6644–6649, DOI: 10.1073/pnas.1718031115 (2018).

19. Paria, B., Mishra, P. & Behera, B. Climate change and transition in cropping patterns: District level evidence from west bengal, india. *Environ. Challenges* **7**, 100499, DOI: https://doi.org/10.1016/j.envc.2022.100499 (2022).

20. Mahawan, A., Jaiteang, S., Srijiranon, K. & Eiamkanitchat, N. Hybrid arimax and lstm model to predict rice export price in thailand. *2022 Int. Conf. on Cybern. Innov. (ICCI)* 1–6 (2022).

21. Prashnani, M., Dupare, B., Vadrevu, K. P. & Justice, C. Towards food security: Exploring the spatio-temporal dynamics of soybean in india. *PloS one* **19**, e0292005 (2024).

22. Agarwal, D. K., Billore, S., Sharma, A., Dupare, B. & Srivastava, S. Soybean: introduction, improvement, and utilization in india—problems and prospects. *Agric. Res.* **2**, 293–300 (2013).

23. Ghosh, S. K. Eggplant (solanum melongena l.) and climate resilient agricultural practices. In *in: Climate Change Dimensions and Mitigation Strategies for Agricultural Sustainability Vol 2* (New Delhi Publishers, 2022).

24. Rao, B. B. & Bhavani, B. Climate change–likely effects on the population dynamics of brinjal shoot and fruit borer (luecinodes orbonalis guen.). *Indian J. Dryland Agric. Res. Dev.* **25**, 58–62 (2010).

25. Agri commodities could offer high returns in 2014, says tushar rathod. https://www.businesstoday.in/magazine/cover-story/story/expert-outlook-agriculture-commodities-in-2014-tushar-rathod-135322-2014-01-10.

26. Sharma, K. *et al.* A complex network analysis of ethnic conflicts and human rights violations. *Sci. reports* **7**, 8283 (2017).