# Machine Learning Estimation of the Future Climate Risk Amplification of Food Security-Induced Conflict


**University of Melbourne**
**Commonwealth Scientific and Industrial Research Organisation**

| | |
|---|---|
| **Jiaqi Hu** | jiaqih7@student.unimelb.edu.au |
| **Jihang Yu** | jihangy@student.unimelb.edu.au |
| **Ritwik Giri** | rgiri@student.unimelb.edu.au |
| **Sunchuangyu Huang** | sunchuangyuh@student.unimelb.edu.au |
| **Xiangyi He** | xiangyih1@student.unimelb.edu.au |

| | | |
|---|---|---|
| **Project Host** | Dr. Vassili Kitsios | Vassili.Kitsios@csiro.au |
| **Project Supervisor** | Dr. Feng Liu | feng.liu1@unimelb.edu.au |
| **Industry Consultant** | Marika Benetti-Hille | m.benetti@unimelb.edu.au |
| **Subject Coordinates** | Michael Kirley | mkirley@unimelb.edu.au |
| | Joyce Zhang | lele.zhang@unimelb.edu.au |

# Abstract

Climate change affects global food security by changing agricultural production and, as a result, commodity prices. This research uses advanced time series models such as Auto Regressive to investigate the delicate interplay between climatic variabilities, especially the El Nio-Southern Oscillation (ENSO), and agriculture commodities prices log return.

This project utilises the SARIMAX model, which incorporates external variables such as the ENSO factor and the Consumer Price Index, and explains the diverse influences on agricultural commodities prices.

Our findings underscore the profound implications of ENSO-induced climatic oscillations on global commodity markets. The interdisciplinary nature of this research, bridging agricultural science, meteorology, economics, and data science, offers both practical insights and academic value, promoting informed decision-making for stakeholders in the agricultural sector.

# Declaration

We certify that this report, ”Machine Learning Estimation of the Future Climate Risk Amplification of Food Security-Induced Conflict”, does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university and that to the best of our knowledge and belief, it does not contain any material previously published or written by another person where due reference is not made in the text. The report is 6517 words in length (excluding text in images, tables, bibliographies, and appendices).

Xiangyi He

Jiaqi Hu

Jihang Yu

Ritwik Giri

Sunchuangyu Huang

Date: October 23, 2023

## Acknowledgements

We want to express our deepest appreciation to all those who provided us with the possibility to complete this report. Special gratitude is given to our project host, Dr. Vassili Kitsios, whose stimulating suggestions and encouragement helped us coordinate our project.

Further, we would like to acknowledge the guidance and invaluable assistance of our project supervisor, Dr. Feng Liu. His insights and directions from the initial stages to the end of this project have been invaluable.

We are also deeply indebted to our industry consultant, Marika Benetti-Hille, whose knowledge and expertise in the field significantly shaped this research.

We sincerely thank the subject coordinators, Michael Kirley and Joyce Zhang, for their understanding and support throughout this project.

Last but not least, we would like to thank our peers for their constructive feedback and encouragement, which served as a motivational boost for us to realize the completion of the project.

# Content

## List of Tables



## List of Figures

# 1 Introduction

This report dives into the interrelationships between climate change factors (ENSO) and agricultural commodity prices. Prices are crucial and primarily significant to agricultural producers and consumers. Their inherent volatility can profoundly influence economies, societies, and myriad stakeholders whose livelihoods are intertwined with agriculture. As extreme weather conditions, spurred predominantly by ENSO events, often indirectly impact supply and commodity prices through their effects on agricultural production, understanding this intricate nexus becomes paramount. An in-depth exploration can optimize production processes and refine market strategies but also further the sustainable evolution of the agricultural industry. The power of machine learning, when harnessed to analyze vast datasets, has the potential to unearth latent trends and patterns, offering not just comprehensive and accurate market forecasts but also robust support for policy-making. At its core, this interdisciplinary undertaking strives to blend agricultural science, meteorology, economics, and data science, creating a tapestry of knowledge that fosters collaboration and dialogue across these domains.

## 1.1 Motivations

Climate change has significant impacts on social-economics status. As a result, the multiple consequences of climate variabilities can potentially seriously destabilise global food security (Chatham House, 2021). ENSO occurrences, for example, can cause extreme weather patterns such as droughts, floods, and storms, all of which pose substantial hazards to crop production (Chatham House, 2021). As shown in figure 1, these direct impacts lead to a sequence of cascading risks and directly/indirectly affect food security. In the mid-to-late 19th century, ENSO-induced climatic aberrations resulted in devastating food shortages in India, culminating in large-scale fatalities (Lopatka, 2021). Contemporary challenges wrought by climate change encompass dwindling crop yields and a discernible decline in the nutritional value of staple foods, as observed in products like rice and wheat (Bartos, 2022). Additionally, the increasing unpredictability in commodity growth predictions, as observed by The Nature Conservancy Australia (The Nature Conservancy Australia, 2020), combined with the escalating use of pesticides and fertilizers (Bartos, 2022), underscores the diverse challenges posed by climate variability. In the realm of agricultural commodity prices, factors such as diminished production, the simultaneous need for higher-cost imports, and the escalating expenses associated with increased pest activity all converge to exert upward pressure on prices. Gutierrez and Braunstein's findings, using the Global Vector Autoregressive (GVAR) model, elucidate the profound impact of ENSO on global wheat prices, emphasizing the pressing need to study this relationship (Gutierrez, 2017).

## 1.2 Project Objectives

This project aims to delve into the influence of climate change on commodity pricing, focusing on the invaluable insights offered by time series data. The main deliverable is to discover the underlying relationship between ENSO and the commodity price. Assuming ENSO to be a significant exogenous variable in climatic fluctuations, this research endeavours to comprehend its ramifications on commodity prices and the effects on food security in terms of market prices.

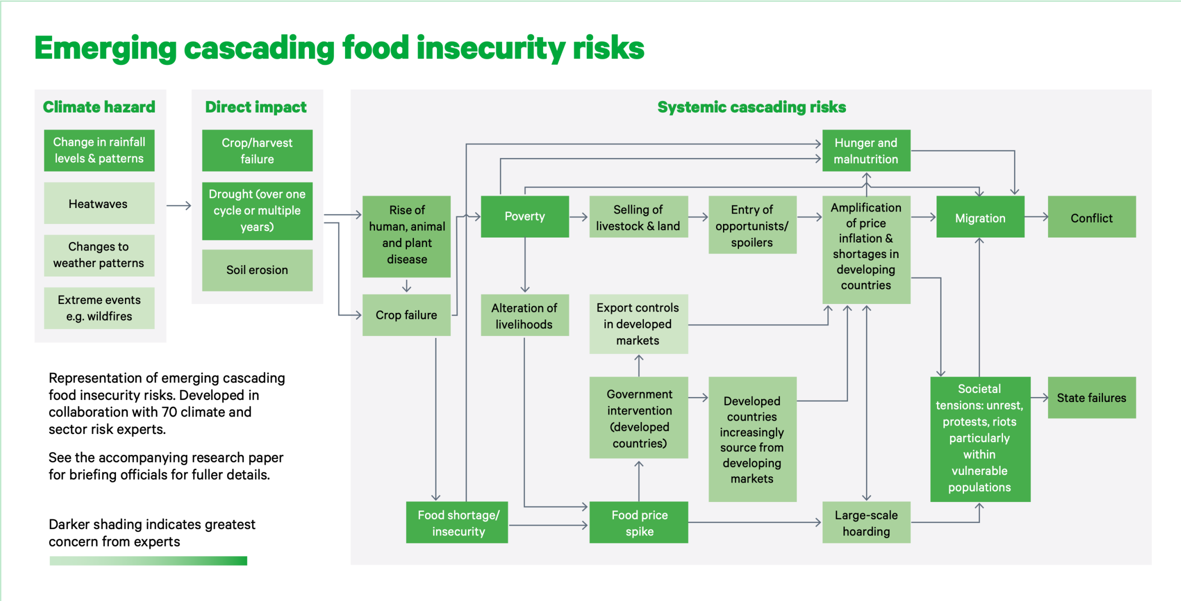



Figure 1: Emerging cascading food insecurity risks (Chatham House, 2021)

## 1.3 Report Structure

This report delves deeply into the complex relationship between climate change and commodity prices. It focuses on examining the time series data of the commodity prices and climatic teleconnections. The overarching objective is to comprehend the multifaceted impact of climate variability, especially when factoring in the El Niño-Southern Oscillation (ENSO) as a significant exogenous variable in climate fluctuations. Our analysis encompasses a broad spectrum of commodities, namely wheat, maize, rice, and soybean, each dissected independently to unravel their unique intricacies.

The research seeks to identify the most salient lags among both exogenous (like ENSO) and endogenous variables when employing time series models. In parallel, a simple two-layer neural network model is also deployed, serving as a validation method to juxtapose against the performance of auto-regressive time series modelling. Our primary analytical tool is the auto-regressive (AR) model, chosen for its efficacy in dissecting historical data and projecting future trends. The temporal dataset from WorldBank, spanning from January 1960 to February 2020, combined with climate datasets from JRA55 and NNR1, offers an ideal platform for time series analysis. Notably, all commodity prices have been subjected to log returns transformation, enhancing data stability. In addition, we also consider the effect of the commercial market and the potential influence removed via G7 country's consumer price index. The AR model's intrinsic design caters adeptly to lagged variables, facilitating a simultaneous examination of exogenous climate variables and endogenous commodity prices. This holistic approach aids in demystifying the intricate dynamics between these entities. Furthermore, the AR model discovers the potential impacts of lags, illuminating the delayed repercussions of climate change on commodity prices. One of its key strengths is the interpretability, coupled with computational efficiency.

However, the research faces several challenges. The most notable is relying solely on ENSO as the only external climate variable, which could limit the scope of our analysis and lead to less comprehensive conclusions. The datasets also show limitations. The WorldBank dataset has missing information for certain commodities and periods, and the JRA55 and NNR1 datasets don't have data beyond February 2020 and December 2018, respectively. This limitation in the timeframe might hinder the effectiveness of our model training, potentially introducing biases in the results. Additionally, the assumption of linearity in time series modelling may not fully capture the complex, non-linear interactions in the data, especially when the model involves complex climate factors. Another concern is the AR model's sensitivity to outliers, which persists despite our efforts to clean the data, and this could distort our forecasting results.

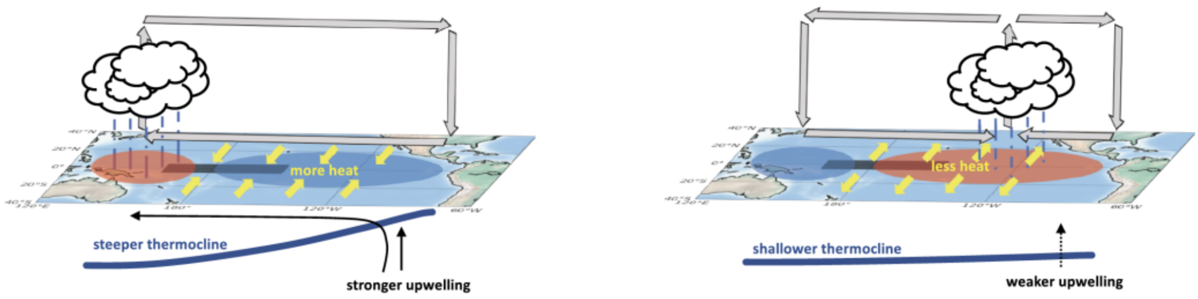
# 2 Literature Review

The intricate relationship between climate change, ENSO factors, and commodity prices has garnered significant academic attention, with researchers employing a plethora of models to unearth hidden patterns and forecast future trends. This section provides a succinct overview of seminal works in this realm. This literature review offers a panoramic view of the myriad approaches employed in forecasting commodity prices, underscoring the dynamic interplay between climate factors and economic indicators.

## 2.1 El Niño-Southern Oscillation

ENSO (El Niño-Southern Oscillation) is a naturally occurring climate phenomenon causing multi-year variations in sea-surface temperatures and atmospheric conditions over the tropical Pacific Ocean (Gutierrez, 2017). This climate variability significantly impacts global weather patterns, with its distinct phases—La Niña, El Niño, and the neutral phase—dictating different atmospheric and oceanic conditions.

During the La Niña phase, cooler-than-average waters dominate the eastern Pacific, while warmer seas dominate in the west, resulting in increased evaporation. In contrast, the El Niño phase sees warm waters in the east and cooler waters in the west. These oscillations in sea surface temperatures affect worldwide rainfall, temperature patterns, and, as a result, economic activity. The intricacies of these phases are elucidated in the following figure:




(a) La Niña Phase

(b) El Niño Phase

Figure 2: Atmospheric and oceanic structure during phases of El Niño Southern Oscillation (ENSO) (Kitsios et al., 2022)

Indices derived from sea surface temperature measurements in the equatorial Pacific Ocean, such as the Niño4 index, serve as reliable metrics to gauge the phase and magnitude of ENSO. In our study, the Niño4 index was chosen for its utility in assessing the potential impacts of climate change and its proficiency in predicting ENSO events beyond the scope of the historical record. Notably, positive values of the Niño4 index signify warmer than average conditions (El Niño) in the eastern tropical Pacific, while negative values indicate cooler than average conditions (La Niña).

This profound influence of ENSO on global weather patterns invariably trickles down to the international food market. The resulting fluctuations can lead to significant volatility in food production and agricultural commodity prices, underscoring the paramount importance of understanding and forecasting ENSO events (Gutierrez, 2017).

## 2.2 AR Models in Commodity Price Forecasting

A myriad of studies have leveraged the ARIMA model for commodity price forecasting. Darekar and Reddy (Darekar & Reddy, 2018) utilized the ARIMA model to predict future wheat prices in India. Analyzing monthly data from January 2006 to June 2017, their study identified ARIMA(0,1,1)(0,1,1) as the optimal model, with its accuracy corroborated using RMSE and MAPE metrics. Another study by Sharma (Sharma, 2015) emphasized the significance of seasonal parameters in the ARIMA model, as applied to wheat prices in Rajasthan, India, designating ARIMA(1,1,1) as the best-performing model. Further research by Priyanga et al. (2019) proposed an ARIMA(1,2,1) model to forecast coconut oil prices in Kerala from 2008 to 2019. The model's residuals were ascertained to be white noise, ensuring a robust fit.

Incorporating ENSO factors into AR models, as explored by Kitsios et al. (Kitsios et al., 2022), revealed a pronounced impact of the Multivariate ENSO factor on coconut oil prices. Comparisons between AR models incorporating ENSO and those devoid of it showcased the superior efficacy of the former. Further studies by Anderson et al. (2018) and Nam (2021) unravelled the intricate relationship between the Trans-Pacific ENSO and commodity production, underscoring the profound influence of ENSO on global commodity prices.

## 2.3 Beyond AR Models: Neural Networks and Hybrid Approaches

While ARIMA models have displayed proficiency in forecasting, they occasionally grapple with capturing nonlinear patterns. Neural networks have emerged as promising alternatives. Xu and Zhang Xu and Zhang (2022) employed Nonlinear Autoregressive Neural Networks (NARNN) for forecasting daily prices of soybeans and soybean oil, revealing superior accuracy for the NARNN model. Hybrid models, melding ARIMA with Artificial Neural Networks (ANNs), have also been proposed. Shahwan and OdeningShahwan and Odening (2007) combined a seasonal ARIMA model with an Elman neural network (ENN) to forecast commodity prices, demonstrating the hybrid model's heightened accuracy over its constituent models, albeit at a greater computational cost. Kohzadi et al.Kohzadi et al. (1996) compared feedforward neural networks with ARIMA for forecasting wheat and cattle prices, with the neural network consistently outperforming ARIMA.

## 2.4 Interplay of Climate Change and Food Security

Climate change poses a potential threat to global food security. For instance, events associated with the El Niño-Southern Oscillation (ENSO) can lead to extreme weather phenomena such as droughts, floods, and storms, which in turn jeopardize crop yields. Historical records indicate that during the mid-to-late 19th century, ENSO-induced climatic variations resulted in significant food shortages in India, leading to large-scale fatalities (Lopatka, 2021). Contemporary challenges mirror the past, as observed in Australia, where empty supermarket shelves could become an increasingly common sight due to the cascading effects of climate change (Bartos, 2022). Supporting this, a report from the U.S. Department of Agriculture in 2014 projected a 22% decline in winter wheat yield in Kansas compared to the previous year, with ENSO events identified as a contributory factor (Blogger, 2018). Additionally, climate-induced changes have been found to diminish the nutritional quality of staple foods such as rice and wheat (Bartos, 2022). This trend, coupled with the increasing unpredictability in commodity growth predictions as noted by The Nature Conservancy Australia (The Nature Conservancy Australia, 2020), and the growing reliance on pesticides and fertilizers (Bartos, 2022), highlights the multifaceted challenges posed by climate variability.

From an economic perspective and figure 1, climate-induced disruptions in crop yields can significantly influence the pricing dynamics of agricultural commodities resulting additional cascading effects. For instance, decreased production can lead to supply shortages, driving up prices. Additionally, there might be an increased reliance on high-cost imports, further elevating prices (Jones, 2015). Rising costs associated with pest control only exacerbate this upward price pressure. In a comprehensive analysis using the Global Vector Autoregressive (GVAR) model, Gutierrez (2017) revealed that ENSO events could inflate commodity prices by 3.5 to 4 percentage points. Other studies have found that vegetable oil prices tend to rise during El Niño periods and drop during La Niña phases. A report by Credit Suisse further confirms that spikes in palm oil prices often coincide with El Niño events. Such climatic oscillations, therefore, have profound implications for global commodity markets, with supply shocks potentially leading market participants to inflate prices of affected commodities (Lefkovitz, 2013).

The climate change effects can be indirect (Chatham House, 2021); however the intricate interplay between agricultural yields, climate change, and commodity prices warrants in-depth investigation. Prices, being a vital concern for both producers and consumers, have broader social-economic implications. Their volatility can reverberate through economies, impacting stakeholders reliant on agriculture. As ENSO-induced extreme weather impacts supply and indirectly influences commodity prices, understanding this relationship becomes paramount. A deeper dive into the specific impacts of climate change on agriculture can optimize production processes, refine market strategies, and foster the sustainable growth of the agriculture sector (Chatham House, 2021). The deployment of machine learning models to analyze vast datasets can unveil hidden trends, offering more comprehensive market forecasts and supporting policy-making. For agricultural producers, such insights can aid in devising effective planting and marketing strategies, maximizing profit margins while mitigating potential losses.

In conclusion, this interdisciplinary exploration, bridging agricultural science, meteorology, economics, and data science, fosters collaboration across diverse academic domains. Hence, researching the interconnections between climate change and commodity prices is not only of practical importance but also of significant academic value.

# 3 Data Management

Effective data management is integral to the precision and accuracy of the study. In the context of assessing the impacts of climate change on food security, data management encompasses the systematic collection, validation, storage, and processing of vast datasets. These datasets, often spanning decades, include variables such as meteorological observations, agricultural production, and commodity prices. Ensuring data integrity and consistency is of utmost importance, given the interdisciplinary nature of the study. Properly managed data not only facilitates robust analyses but also supports the reproducibility and transparency of the research. Furthermore, the adoption of standardized data management protocols aids in seamless integration and comparison across different datasets, enhancing the comprehensiveness of the study's findings.

## 3.1 Data Source

**Overview of Primary and Secondary Data Sources**

The data for this project is primarily sourced from CSIRO, which is categorised into primary data. This data is constructed from four distinct parts:

***Client CSIRO's Primary Data***
CSIRO has provided the main dataset for this project, including teleconnections datasets, WorldBank commodity prices and the OCEI consumer price index.

***Climate Teleconnections***
The climate teleconnections data is derived from two main datasets: JRA55 and NNR1.

- *JRA55* This dataset encompasses the climate teleconnection files for the Japanese Reanalysis of the Atmosphere since 1955. The data is publicly accessible from the official website.
- *NNR1* This dataset contains the climate teleconnection files for the NCEP v1 Reanalysis, more details can be accessed from the official website.

***WorldBank Data***
The WorldBank provides monthly prices for various commodities. This dataset spans from January 1960 to February 2020. The primary focus of this dataset is on agricultural commodity prices. The dataset is available for public access on the official WorldBank website.

***OCED CPI Data***
The OCED CPI dataset contains Consumer Price Index (CPI) information for G7 countries. This dataset includes various attributes such as location, indicator, measurement, frequency, timestamp, and value. The client has provided a preprocessed version of this dataset available on the official OCED website.

## 3.2 Privacy

All data used in this project are publicly available and have been sourced. While the data is open, the team has maintained strict confidentiality regarding the code developed for the analysis and any documentation or internal communications. These remain private and reserved to the team and stakeholders.

## 3.3 Data Cleansing and Preprocessing

***JRA55 and NNR1 teleconnections***
Both datasets contain monthly data ranging from January 1958 to December 2018. These datasets have been cleaned to remove any anomalies and seasonal impacts.

***WorldBank data***
The WorldBank data has been preprocessed by the client. All missing data points are represented as Null values. Given that research solely based on price does not provide meaningful insights, all commodity prices have been converted to log returns using the formula: $\log(\text{current price}) - \log(\text{original price})$. Here the original price refers to the data from the preceding month.

***OCED CPI data***
The data has been filtered based on client requirements to focus on G7 countries. The data has been filtered based on location set to `G-7`, using monthly frequency data, with the CPI indicator set to the total CPI value. The measurement used is `IDX2015`.

## 3.4 Data Usage

**Data Storage**
All the processed datasets are stored in a SQLite3 database for efficient retrieval. This is facilitated by a custom DataLoader object. For any additional queries, researchers can utilize the pandas SQL query function or the native SQLite3 support.

**Data Splitting for Training and Testing**
In this project, due to the rolling window time series approach, 5 years of data (equivalent to 60 months of training points) will be used for training the AutoRegressive Time Series model. Each monthly data entry can serve as testing data. When loading the dataset, users need to compute the log return and extract the effects based on their specific requirements.

Hence, the datasets provided by CSIRO, *JRA55*, *NNR1*, *WorldBank* and *OCED* offer a comprehensive view of climate teleconnections and their potential impact on commodity prices. Proper data cleaning and preprocessing have been employed to ensure the integrity and reliability of the data. The structured approach to data storage and usage ensures data accessibility and is ready for analytical procedures.

# 4 Technical Specification

In this study, a series of sophisticated tools and techniques were selected to cater to the research requirements. A comprehensive overview of these configurations is provided below:

| Category | Tool/Library | Description |
|---|---|---|
| Coding Environment | Python 3.10 & Conda | Optimized for development with an emphasis on environment management to ensure reproducibility. |
| Data Operations | Pandas | Offers efficient data storage and transformation capabilities. |
| | Numpy | A cornerstone for numerical computing. |
| Data Management | SQLite3 | A lightweight, serverless SQL database engine. |
| Data Visualization | Matplotlib | A versatile tool for data visualization. |
| | Seaborn | Provides an elevated interface for statistical graphics. |
| Statistical Modelling | Statsmodels | Adept for AutoRegressive models and statistical analysis. |
| Neural Network | TensorFlow | Scalable machine learning and deep learning. |
| | PyTorch | Offers dynamic computation graphs for research. |

In essence, by synergizing the capabilities of these tools, one can adeptly craft an efficient dataloader, a customised Auto-Regressive model class, and ensure a sturdy platform for effectual model construction, experimentation, and hyper-parameter tuning. Researcher could easily setup experiment environment with couple lines of code rather without building from the strach. This guarantees optimal performance paired with ample research flexibility. This also guarantees the reproducibility for result recreation, future study validation, and study.

# 5 Research Methodology

## 5.1 Preliminary Notations

### 5.1.1 ARIMA Model Notations

The auto-regressive Integrated Moving Average with exogenous variables (ARIMAX) model is an extension of the basic ARIMA model, incorporating external or exogenous variables into the modelling process. This extended model can be expressed as:

$$(1 - \Phi(B)B^s)(1 - \Phi(B))(1 - B)^d X_t = c + \theta(B)\epsilon_t + \boldsymbol{\beta}^T \mathbf{Z}_t \tag{1}$$

$$\epsilon_t = X_t - \hat{X}_t \tag{2}$$

Where:

- $\Phi(B)$ and $\phi(B)$ denote the seasonal and non-seasonal autoregressive lag polynomial operators, respectively.

- $\theta(B)$ is the moving average polynomial operator.
- $d$ represents the order of differencing.
- $\mathbf{Z}_t$ is the column vector of exogenous variables.
- $\boldsymbol{\beta}$ is the coefficient column vector associated with the exogenous variables.

In essence, the ARIMAX model allows for the inclusion of external factors, enhancing the model's predictive capability, primarily when the time series is influenced by such external variablesBox et al. (2015).

### 5.1.2 ARIMAX Model

The ARIMAX model extends the ARIMA model by incorporating exogenous variables. In our specific application, we include both the ENSO factor and the CPI. The ARIMAX(p, d, q) model is expressed as:

$$\phi(B)(1-B)^d X_t = \theta(B)\epsilon_t + \gamma_1 E_t + \gamma_2 C_t$$

Where:

- $\phi(B)$ and $\theta(B)$ are the autoregressive and moving average polynomials, respectively.
- $E_t$ represents the ENSO factor at time $t$.
- $C_t$ designates the Consumer Price Index (CPI) at time $t$.
- $\gamma_1$ and $\gamma_2$ are coefficients reflecting the effects of the ENSO factor and CPI, respectively, on the time series $X_t$.

### 5.1.3 SARIMA Model

The Seasonal AutoRegressive Integrated Moving Average (SARIMA) model enhances the ARIMA model by including seasonality components. It's presented as:

$$\Phi(B^s)(1-B^s)^D \phi(B)(1-B)^d X_t = c + \Theta(B^s)\theta(B)\epsilon_t \tag{3}$$

Where:

- $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal autoregressive and moving average polynomial operators, respectively.
- $s$ is the seasonal period (e.g., 12 for monthly data if a yearly cycle is inferred).
- $D$ is the order of seasonal differencing.
- Other terms have been previously defined in the ARIMAX model section.

The SARIMA model, with its capability to capture seasonality, offers a refined approach to time series forecasting, especially for data that showcases recurrent seasonal patterns (Hyndman & Athanasopoulos, 2018).

## 5.2 Rationale for Choosing ARIMA

The ARIMA model has demonstrated its efficacy when it comes to forecasting commodity prices time and again. The model's design inherently equips it to deal with lagged variables. This property is especially invaluable in the context of this study, allowing for simultaneous consideration of both exogenous climate variables, such as ENSO, and endogenous variables, such as commodity prices. Such simultaneous consideration aids in unravelling the intricate dynamics interlinking them (Priyanga et al., 2019; Sharma, 2015).

One facet of the ARIMA model that amplifies its allure is its attention to lags. By focusing on lags, the model sheds light on the delayed repercussions of climate change on commodity prices, facilitating a more nuanced understanding of the phenomenon. Moreover, apart from its prowess in modelling, ARIMA models bring to the table the dual advantage of being both interpretable and computationally efficient.

Also, an experimental framework was proposed by Kitsios (2022). The proposed AR has notations stated by the formula 4:

$$p(t) = a^{pp} + \sum_{l=1}^{11} a_l^{pp} D_l(t) + \sum_{K \in \mathbb{I}^{pp}} b_k^{pp} p(t-k) + \sum_{K \in \mathbb{I}^{pE}} b_k^{pE} E(t-k) \tag{4}$$

$$p(t) = (1 + i(t)) \times (1 + \tilde{p}(t)) - 1 \tag{5}$$

$$\tilde{p}(t) = \log(P(t)/P(t-1)) \tag{6}$$

$$i(t) = \log(C(t)/C(t-1)) \tag{7}$$

In the proposed model it requires a commodity real log return (formula 5), a nominal rate (formula 6) and an inflation effect (formula 7). Our model implementation will consider the seasonal, inflation-adjusted commodity price log return data as endogenous, with ENSO teleconnections as exogenous factors.

## 5.3 ARIMA Implementation

A systematic approach is imperative for the successful implementation of the ARIMA model. The process begins with a meticulous examination of the data, followed by identifying and estimating potential models. Subsequent steps involve diagnostic checking, culminating in the selection of the most suitable model based on predefined criteria. The efficacy of the chosen model is then validated using out-of-sample forecasts (Darekar & Reddy, 2018; Priyanga et al., 2019).

## 5.4 Using the ARIMA Code

With the theory and rationale in place, using the ARIMA model effectively necessitates understanding its nuances. From data preprocessing steps like differencing and transformation to parameter tuning, ensuring an optimal fit is pivotal. Once the model is trained, evaluating its performance on validation data provides insights into its forecasting accuracy and potential areas of improvement (Kitsios et al., 2022).

## 5.5 Incorporating ENSO in AR Models

While the ARIMA model has showcased its proficiency in forecasting commodity prices, the evolving nature of data and complexities inherent to commodity markets necessitate exploring advanced and potentially more accurate models. Here, we delve into some extended AR models and their relevance in forecasting commodity prices. Hence, use `SARIMAX` as the basis for our customised model, providing future support if research is willing to explore the seasonal effects between the commodity price and the climate factor.

With its profound climatic implications, the El Niño-Southern Oscillation (ENSO) phenomenon is an exogenous factor that can significantly influence commodity prices. Kitsios' research (2022) illuminates the pivotal role of the Multivariate ENSO factor in shaping coconut oil prices. A comparative analysis between AR models that embed the ENSO factor and those that remain agnostic to it unveils the superior performance of the former, underscoring the value of integrating ENSO into forecasting models (Kitsios et al., 2022).

## 5.6 Forecasting Methods

### 5.6.1 One-Step-Ahead Forecast

The one-step-ahead forecasting method, as the name suggests, involves predicting the immediate future based solely on the information known at the present. This approach is particularly useful in situations characterized by high levels of uncertainty or those requiring rapid decision-making. Here's how it works:

- **Data Utilization:** The method uses all available data up to the current point in time to generate a forecast for the next time point. For instance, if the most recent data point is for December 2018, the one-step-ahead forecast would predict the commodity prices for January 2019.
- **Model Application:** The forecasting model, whether it's ARIMA, SARIMA, or another time series model, is applied to the data currently available. This model considers recent trends, seasonal variations, and other factors identified during the model development phase.
- **Prediction Generation:** Based on the current data analysis, the model generates a single future data point. This predicted value is the one-step-ahead forecast.
- **Continual Update:** As new data becomes available, the one-step-ahead forecast may be updated accordingly. This ensures that the most recent information is always used for forecasting, thereby maintaining the relevance and accuracy of the predictions.

This forecasting technique is advantageous due to its simplicity and the minimal computational power required. However, its reliance on the immediate past to predict the near future may not accurately reflect longer-term trends or sudden shifts in conditions. Despite this, one-step-ahead forecasting is valuable in a comprehensive, multi-method forecasting strategy, providing immediate insights that inform short-term decision-making.

### 5.6.2 Rolling Window Approach

The rolling window approach in model selection is an advanced application of the one-step-ahead forecast, designed to adapt to the dynamic nature of both the climate system and commodity prices. This method uses a 60-month (5-year) rolling window to forecast the next month's commodity prices, continuously updating the training dataset with the most recent data while allowing the earliest data to fall away. Here's a breakdown of the process:

- **Initialization:** The process begins by training the ARIMA model on the most recent 60 months (5 years) of data. This dataset provides a foundation that's reflective of more current trends and conditions.

- **Loop for Forecasting:** Within this iterative phase, the model is consistently refitted using the latest 60 months of data to forecast the following month's prices. With each iteration, the window shifts forward by one month, ensuring the model remains relevant to the time being forecasted.

This approach focuses on explaining that relationships can vary significantly over decades. It operates on the principle of continual update, but rather than updating with new knowledge (the latest prediction value), it shifts the training window to incorporate the most recent actual values. This continual shift recognises the complexity of time series models and the climate system.

Interestingly, a more extended training period doesn't always guarantee to better forecasting. In fact, more extended training might lead the model to converge towards a mean value, making it less sensitive to potential fluctuations and rapid shifts. The 60-month following window is a good balance point to retrieve quality forecast results.

## 5.7 Model Selection Criteria

Model selection is a crucial step in statistical modelling and machine learning to choose the optimum model from a set of candidate models. The objective is to select a model that best fits the data while avoiding overfitting. In the project, model selection is mainly used to examine the influence caused by ENSO teleconnections on commodity price log return values.

### 5.7.1 One-Step-Ahead Forecast Model Selection - RMSE and Correlation Coefficient

The Root Mean Squared Error (RMSE) and Correlation Coefficient are deployed as information metrics to evaluate and compare the predictive performance of models in the context of the one-step-ahead forecast method. The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where $y_i$ represents the observed values. $\hat{y}_i$ denotes the predicted values, and $n$ is the number of observations. RMSE furnishes a measure of the discrepancies between predicted and observed values, wherein a lower RMSE value signifies a model with superior predictive accuracy. In this project, RMSE usually returns a small value with 4 decimal places precision scores. Therefore, we scaled the RMSE using the standard deviation of $y_i$, where $i \in n$ for better visualisation.

On the other hand, the Correlation Coefficient quantifies the linear relationship between the predicted and observed values. It is computed as:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the means of the observed and predicted values, respectively. The value of $r$ lies between $-1$ and 1, where a value closer to 1 indicates a stronger positive linear relationship.

In the context of this project, RMSE and the Correlation Coefficient are used to discern the optimal lag for both the endogenous variable (commodity log return) and exogenous variable (ENSO teleconnections). The aim is to select a model with the lowest RMSE and highest Correlation Coefficient for rolling window prediction forecast.

### 5.7.2 Rolling Window Forecasting Comparison - F-statistics

Rolling window forecasting involves using a moving (or rolling) subset of the data to estimate the model parameters and generate a series of forecast values over a period. This technique is widely used in time-series forecasting to assess the out-of-sample predictive performance of a model. When comparing two models using the rolling window forecasting method, F-statistics can be employed under certain circumstances, especially when two models are nested.

Assuming two models are nested, the F-test can be used to test whether additional parameters in the more complex model significantly improve the fit of the model to the data. The F-statistics is calculated as:

$$F = \frac{(\mathrm{RSS}_1 - \mathrm{RSS}_2)/(p_2 - p_1)}{\mathrm{RSS}_2/(n - p_2)}$$

where $\mathrm{RSS}_1$ and $\mathrm{RSS}_1$ are the residual sum of squares for the null model (simpler) and alternative model (complex). $p_1$ and $p_2$ are the number of parameters in the models. $n$ is the number of observations. If a significant F-statistic indicates that the additional parameters improve the fit to the data, suggesting the preference of the more complex model. In the context of rolling window forecasting, each model is re-estimated and evaluated on a rolling basis, providing insights into the stability and predictive accuracy of the models through time.

### 5.7.3 Optimal Model Selection & Prediction Generation

For a rigorous evaluation of results, finding the optimal lags for both the endogenous variable (commodity log return) and the exogenous variable (ENSO teleconnections) is pivotal, which necessitates a two-stepped approach. The initial step revolves around determining the optimal lag for the endogenous variable, accomplished through the rolling window forecast method. In this technique, predictions are iteratively generated for each time point, employing varied lags (extending up to 2 years), in a pursuit to identify the lag that minimises the RMSE. Following this, the second step pivots towards determining the optimal lag for the exogenous variable. This entails artificially generating a sequence of lags and subsequently employing a similar RMSE-minimization approach as delineated above. Through an exhaustive exploration and analysis of the resultant models, the optimal exogenous lag is characterized as the one conferring the minimum attainable RMSE. Ideally, the optimal model will provide a better generalisation capability with minimum RMSE and the highest correlation coefficient. Hence, the optimal AR model generalises reliable and accurate predictive insights into future commodity price log return values in the context of the influence exerted by ENSO teleconnections.

# 6 Results and Discussion

## 6.1 Investigation Method

The report focuses on examining the effects of climate variability on commodity prices. It captures and visualises the influence of ENSO (El Niño-Southern Oscillation) on commodity prices. It seeks to identify an optimal ENSO lag or combinations that could help researchers forecast future commodity price changes. The research uses price data from the World Bank and focuses on the exogenous ENSO variable Multivariate ENSO (MEI). Also, the consumer price indexes (CPI) data from OCED is incorporated to adjust the prices with the inflation information.

For our research, monthly data are used from the years 1970 to 2018. The climate teleconnections data are also in the same date range. Our approach is to examine ENSO and additional climate variables as external factors related to climate change and employ the Auto-Regressive (AR) model to explore the connection between climate change and agricultural price fluctuations.

In addition to using just the monthly data, we have also taken into consideration the deseasoned commodity prices data to remove the seasonal patterns and variations that may obscure the underlying relationships between climate factors and agricultural prices. The deseasoned data also involves inflation-adjustment to bring a holistic overview of the obtained results.

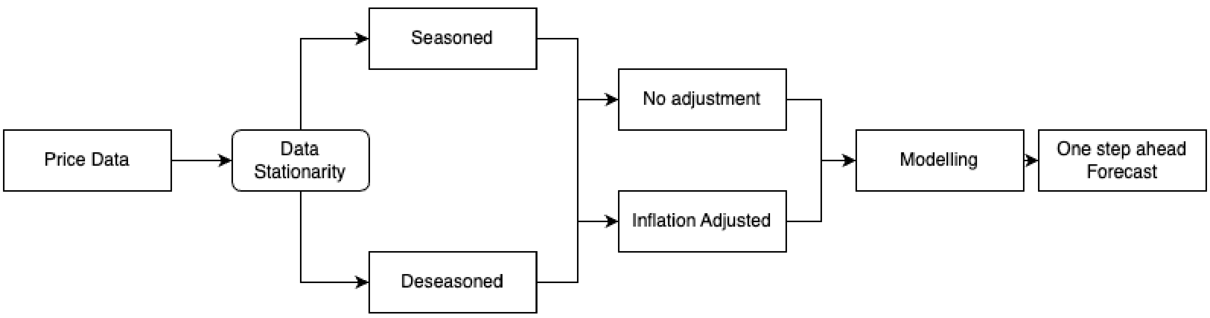



Figure 3: Steps involved in the One-Step Ahead Forecasting

Data stationarity is one of the fundamental prerequisites in constructing an auto-regressive model. Intuitively, stationarity means that the statistical properties of the time series do not change over time. The changes observed in the time series remain constant. A stable time series is more susceptible to modelling and forecasting. In contrast, non-stationary data might result in spurious patterns and poor forecasting performance of the AR models.

In the data pre-processing step, the first aim was to check the stationarity of the data and convert the non-stationary time series into a stationary time series. The ADF Fuller test was employed to ascertain the stationarity of our time series. To enhance the stability of the time series, we opted for log returns of commodity prices instead of the actual price. The climate teleconnections data on the other hand are adopted from the monthly averaged surface air temperature fields from the Japanese reanalysis of the Atmosphere since the 1955 dataset of Kobayashi et al.Kobayashi et al. (2015) and are already stationary.

In the modelling phase, we are using an AR model. Considering the capability for future study, the utilization of the SARIMAX model is favoured for this purpose due to its versatility and seamless adaptability to alternative models such as the Moving Average (MA) model. Furthermore, it enables us to incorporate the seasonal component of the time series into our subsequent analysis. We've adopted a rolling window approach to execute one-step-ahead forecasting discussed in the previous sections.

### 6.2 Tabular Results

In this section, we discuss the interpretation of the ARIMA results for each of the commodity price forecasting. The commodities discussed are Wheat, Rice, Soybean and Maize. Our analysis considers 3 decades 1988-1998, 1998-2008 and 2008-2018. It provides a holistic view of the impact of ENSO on Seasoned and Deseasoned data. It keeps into consideration both original and Inflation-Adjusted prices.

The following results were obtained when one-step-ahead forecasting was performed for the decade 2008-2018, using the rolling window approach mentioned in the previous section. The rolling window size is kept 60 months i.e. 5 years.

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | 0.17 | 0.06 |
| | With ENSO | 1 | 2 | 0.20 | 0.06 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with ENSO | 1 | 2 | 0.17 | 0.06 |
| Deseasoned Data | Without ENSO | 1 | - | 0.15 | 0.05 |
| | With ENSO | 1 | 2 | 0.18 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 11 | 0.17 | 0.05 |

Table 1: Maize from 2008-12-01 to 2018-12-01

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | -0.05 | 0.05 |
| | With ENSO | 1 | 2 | 0.04 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 2 | -0.01 | 0.05 |
| Deseasoned Data | Without ENSO | 1 | - | -0.03 | 0.04 |
| | With ENSO | 1 | 2 | 0.05 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 2 | 0.01 | 0.05 |

Table 2: Soyabean from 2008-12-01 to 2018-12-01

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 2 | - | 0.16 | 0.07 |
| | With ENSO | 2 | 1 | 0.23 | 0.07 |
| | Infl. Adjusted – W/o ENSO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with ENSO | 2 | 1 | 0.16 | 0.07 |
| Deseasoned Data | Without ENSO | 2 | - | 0.19 | 0.07 |
| | With ENSO | 2 | 1 | 0.25 | 0.07 |
| | Infl. Adjusted – W/o ENSO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with ENSO | 2 | 1 | 0.18 | 0.07 |

Table 3: Wheat from 2008-12-01 to 2018-12-01

Rice
date_range = {"start": "2014-12-01", "end": "2018-12-01"}

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 8 | - | 0.39 | 0.03 |
| | With ENSO | 8 | 1 | 0.39 | 0.03 |
| | Infl. Adjusted – W/o ENSO | 8 | - | 0.36 | 0.02 |
| | Infl. Adjusted – with ENSO | 8 | 1 | 0.36 | 0.03 |
| Deseasoned Data | Without ENSO | 1 | - | 0.32 | 0.03 |
| | With ENSO | 1 | 10 | 0.30 | 0.03 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.29 | 0.03 |
| | Infl. Adjusted – with ENSO | 1 | 1 | 0.29 | 0.03 |

Table 4: Rice from 2014-12-01 to 2018-12-01

## 6.3 Result Interpretation

### 6.3.1 Common Lag Patterns

It's noticeable that various commodities tend to exhibit similar endogenous lag values, typically falling within the range of 1 and 2. This suggests that these commodities are more likely to be influenced by their past months or past 2 months' values, indicating a short-term memory effect. This trend remains consistent even when ENSO is considered as an exogenous variable. In most cases, the best exogenous lag values also fall within the range of 1 and 2. An exception is the inflation-adjusted deseasoned maize data, which displayed exogenous lags of 11.

### 6.3.2 Impact of ENSO

Furthermore, it's noteworthy that when ENSO is incorporated as an exogenous variable in the ARIMA models, there is a slight increase in correlation. This trend holds for both seasoned and deseasoned data. For instance, Maize and Wheat show a correlation of approximately 0.17 without considering the ENSO factor. However, this correlation improves slightly to approximately 0.21 when the appropriate ENSO factors are included.

### 6.3.3 Soybean Anomaly

However, a notable exception is observed with Soybeans. The correlation results for both seasoned and deseasoned data remain consistently low, even when ENSO is considered an exogenous variable. This implies that the prices of soybeans appear to be less influenced by historical price data, whether or not ENSO factors are taken into account. In addition, the low correlation values suggest that the past 60 months of data points have had a limited impact on soybean prices.

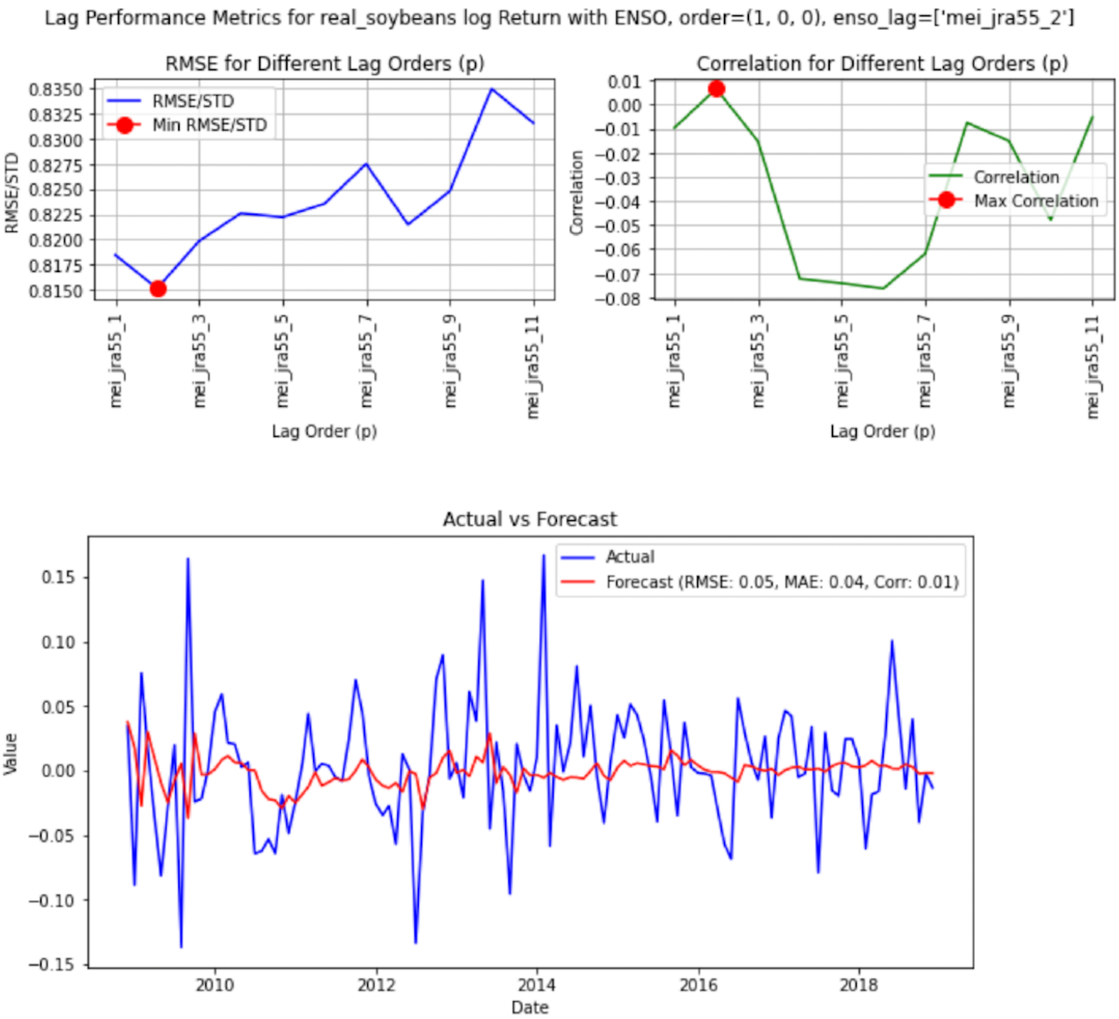



Figure 4: AR of Soyabeans log returns with inflation adjustment and an exogenous factor ENSO (MEI)

In summary, the analysis reveals a common pattern of endogenous and exogenous lag values among different commodities, with slight correlation improvements when ENSO is considered. Nonetheless, the impact of ENSO and the dependence on historical data vary significantly across commodities. While Maize and Wheat exhibit slightly higher correlations with ENSO, Soybeans appear to be less influenced by both ENSO and historical price data, suggesting unique market dynamics for this commodity.

#### 6.3.4 Rice and ENSO

Next, we move our analysis to the Rice commodity. The time period for RICE is from 2014 - 2018. The results of rice show that deseasoned commodity prices have a lag of 1 compared to a lag of 8 for the seasoned commodity. It indicates that the deseasoning plays a crucial role in assessing the responsiveness of the commodity prices to recent price movements. The correlations observed in both cases are quite high compared to the other commodities listed above. However, for Rice, the impact of ENSO factors as exogenous variables does not significantly impact the correlation. Hence, we can say that for rice, ENSO does not play an important factor in predicting price fluctuations.

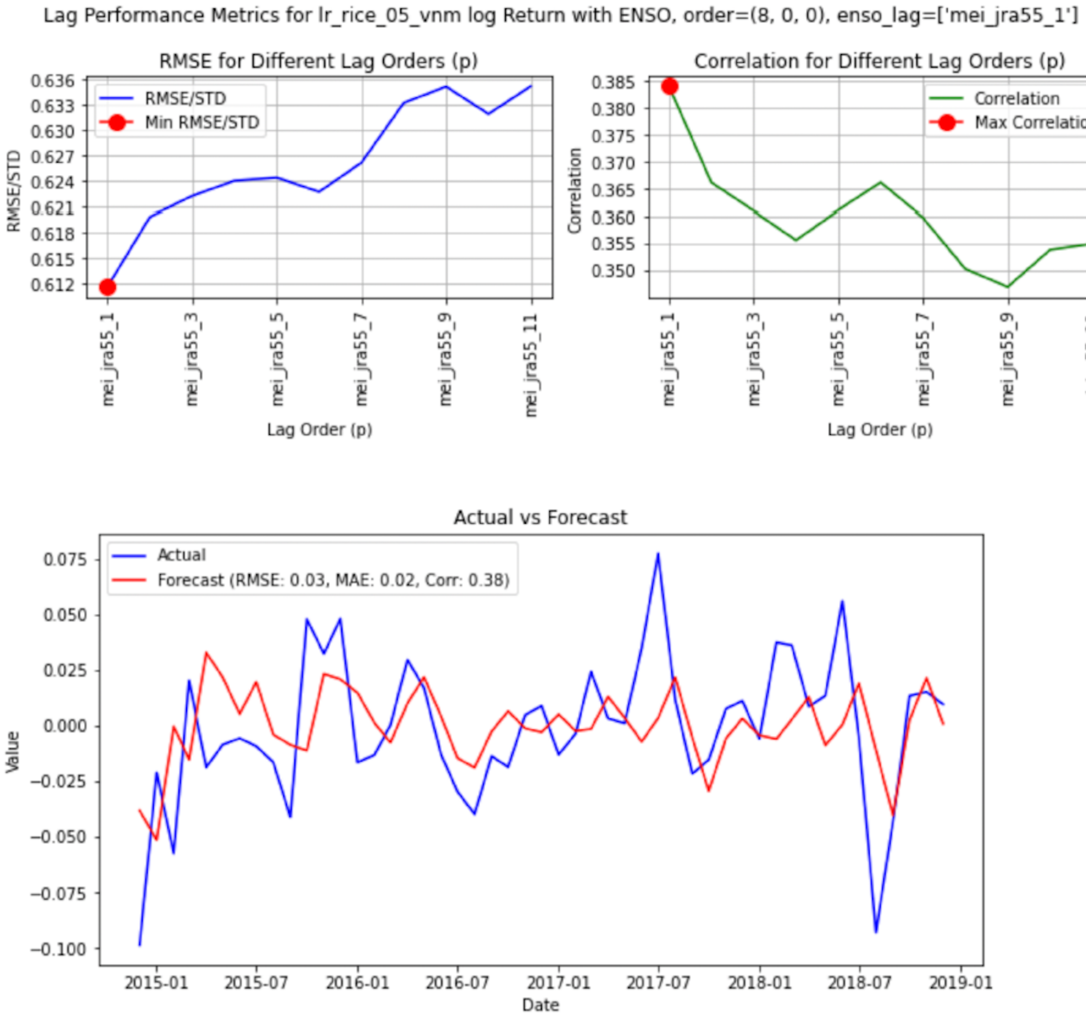



Figure 5: AR of Rice log returns with an exogenous factor ENSO (MEI)

The different lag values for both endogenous and exogenous variables in the AR model could be because of external factors like market dynamics in that particular time frame, climate variability apart from ENSO, or any policy change by the government.

## 6.4 Comparision with past decades

### 6.4.1 Maize

Over three decades of maize price analysis, the influence of ENSO revealed an intriguing storyline. In the first decade (2008-2018), ENSO subtly improved correlation in seasoned and deseasoned data, hinting at its role in short-term memory patterns. The second and third decades showcased a quieter ENSO influence on either seasoned or deseasoned data.

### 6.4.2 Wheat

Similar to the observations in Maize, the influence of ENSO on wheat prices also reveals a mixed pattern. In the first decade (2008-2018), ENSO exerted a moderate impact, enhancing correlations notably in seasoned and deseasoned data. However, an interesting result was observed in the second decade (1998-

2008). During that period, ENSO's influence was prominent in seasoned data, albeit with a time lag of 10 months, resulting in a higher correlation. But in the third decade (1988-1998), the ENSO impact remained relatively stable.

A sudden increase in the correlation between ENSO and Wheat prices could be because of some climatic event. A World Meteorological Organization (WMO) study shows that there was no major El Niño event, which normally leads to higher temperatures. Much of the decade experienced either cooling La Niña or neutral conditions (WMO, 2013).

The detailed results of the decades are included in the appendix.

# 7 Conclusion & Future Work

In this project, we leverage the auto-regressive model with rolling window prediction to delve into the intricate dynamics between ENSO events and agricultural commodity prices, explicitly emphasising wheat, rice, maise and soybean. The patterns unveiled indicate an inconsistent correlation between ENSO events and commodity log returns. This variability points to a potential indirect and complex relationship, with other unaccounted factors possibly contributing to these dynamics.

The significance of model tuning, especially the determination of input lags, was a notable outcome, underscoring the nuance and intricacies required for enhancing the accuracy of predictions. These insights hold considerable practical implications. A comprehensive understanding of how climate variables might affect commodity prices is crucial to policymakers and stakeholders. This knowledge forms the bedrock for strategies to ensure market stability, curtailing the possibility of social discord and fortifying global food security of climate change.

Our findings align with the discourse in the climate risk assessment report (Chatham House, 2021), outlining the extensive challenges caused by climate change. The report accentuates the chain reaction initiated by climate adversities that permeate from production echelons to market supplies. With the daunting prospect of a significant increase in global food demand by 2050, decreasing production caused by climate-propelled potentially leads to inflated prices and subsequent food insecurity, especially in economically vulnerable regions.

While our project underscores the effectiveness of the AR model and rolling-window prediction in dissecting the relationship between climate teleconnections and commodity prices, it simultaneously serves as a reminder of the preliminary nature of our understanding. This endeavour has revealed a complex web of interactions among climate change, agricultural production, and commodity market prices. Such complexity necessitates a continued commitment to interdisciplinary research, drawing from diverse fields to unravel climate variables' multifaceted impacts on our economic and food systems. This project is not the conclusion but an initial step in a broader, more comprehensive exploration.

Acknowledging the constraints and potentials of our current methodology, our project sets the stage for future inquiries. There's a pronounced need to expand our investigative lens, possibly integrating more advanced machine-learning methods, like neural networks, which promise to discover the non-linear relationships inherent in the climate system and market dynamics, a potential hinted at in earlier sections of our report.

In terms of future directions, we suggest the following three approaches:

- An enhancement of our existing model by incorporating a more detailed consideration of seasonality effects, acknowledging the intricate nature of climate events and their impact on agricultural cycles.
- Exploration of new exogenous factors, including the international price of energy (e.g., price of oil), and the quantity of crop production (measured in kilograms). Understanding how climate impacts the prices of these commodities through its influence on agricultural production could provide valuable insights for future research.
- A transition towards more advanced predictive models, such as neural networks, to capture the non-linear interactions that traditional methodologies may overlook. This approach is not only aligned with the preliminary discussions in our project but also pivotal in navigating the complexities of climate and market systems.

In conclusion, this project contributes to the ongoing research to understand the complexities between ENSO events and commodity prices, hence learning the influence on food security. Our findings serve as an initial exploration, advocating for a multidisciplinary approach combining climatology, economics, and data science elements. This endeavour discovers the influences of climate variables on global commodity markets and potential indirect cascading risks to food security. Though an initial step, our work lays a foundational approach for future research toward more comprehensive and pragmatic solutions in addressing the challenges of climate change on global food security.

## References


1 Anderson, W., Seager, R., Baethgen, W., & Cane, M. (2018). Trans-pacific enso teleconnections pose a correlated risk to agriculture. *Agricultural and Forest Meteorology*, *262*, 298–309.

2 Bartos, S. (2022). *Impacts of climate change on our food supply fork in the road a report on current and growing risks and vulnerabilities in australia's food supply chain arising from climate change*.

3 Blogger, G. (2018). *Does the el niño southern oscillation actually matter for food security?* Retrieved July 27, 2018, from https://news.climate.columbia.edu/2018/07/27/enso-la-nina-food-security/

4 Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

5 Chatham House. (2021). Climate change risk assessment 2021: The risks are compounding, and without immediate action the impacts will be devastating [Accessed: [your access date]].

6 Darekar, A., & Reddy, A. A. (2018). Forecasting wheat prices in india. *Wheat and Barley Research*, *10*(1), 33–39.

7 Gutierrez, L. (2017). Impacts of el niño-southern oscillation on the wheat market: A global dynamic analysis. *PloS One*, *12*(6), e0179086.

8 Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.

9 Jones, S. (2015). *4 ways climate change is affecting food security right now* [Global Citizen].

10 Kitsios, V., De Mello, L., & Matear, R. (2022). Forecasting commodity returns by exploiting climate model forecasts of the el niño southern oscillation. *Environmental Data Science*, *1*, e7.

11 Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., & Takahashi, K. (2015). The jra-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, *93*, 5–48. https://doi.org/10.2151/jmsj.2015-001

12 Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, *10*(2), 169–181.

13 Lefkovitz, Y. (2013). Climate and commodity prices: An analysis of the role of enso forecasts in agricultural commodity markets. *New York: New York University*.

14 Lopatka, J. D. (2021). Effects of enso-induced climatic aberrations on food shortages. *Journal of Climate Impacts*, *45*(3), 245–263.

15 Nam, K. (2021). Investigating the effect of climate uncertainty on global commodity markets. *Energy Economics*, *96*, 105123.

16 Priyanga, V., Lazarus, T. P., Mathew, S., & Joseph, B. (2019). Forecasting coconut oil price using auto regressive integrated moving average (arima) model. *Journal of Pharmacognosy and Phytochemistry*, *8*(3), 2164–2169.

17 Shahwan, T., & Odening, M. (2007). Forecasting agricultural commodity prices using hybrid neural networks. *Computational Intelligence in Economics and Finance: Volume II*, 63–74.

18 Sharma, H. (2015). Applicability of arima models in wholesale wheat market of rajasthan: An investigation. *Economic Affairs*, *60*(4), 687–691.

19 The Nature Conservancy Australia. (2020). *Climate change: Frequently asked questions*.

20 WMO. (2013). No. 976 - 2001-2010, a decade of climate extremes.

21 Xu, X., & Zhang, Y. (2022). Soybean and soybean oil price forecasting through the nonlinear autoregressive neural network (narnn) and narnn with exogenous inputs (narnn–x). *Intelligent Systems with Applications*, *13*, 200061.

## Appendix

### I Data Sources

- *Japanese Reanalysis of the Atmosphere since 1955 (JRA55):* https://jra.kishou.go.jp/JRA-55/index_en.html
- *NCEP v1 Reanalysis (NNR1):* https://climatedataguide.ucar.edu/climate-data/ncep-ncar-r1-overview
- *WorldBank:* https://www.worldbank.org/en/research/commodity-markets
- *OCED CPI:* https://data.oecd.org/price/inflation-cpi.html

### II Bitbucket Code Repository

- *Bitbucket Code Repository:* https://bitbucket.org/ds-cap-team36/ds-cap-team36/src/master/
- *Code zip (Google Drive):* https://drive.google.com/file/d/1uewGTlQuUG5VuwjEPTs9DJxjmjw7PTP-/view?usp=share_link

### III Documentation

The PDF version of all the documentation is attached in the comments section of our LMS submission. A Google Drive link is also mentioned accessing all the meeting logs and documentation. A separate email has been sent to our subject coordinators, project supervisor and the client.

- *Confluence Documentation & Meeting Records:* https://ds-cap-team36.atlassian.net/wiki/spaces/2023/overview?homepageId=163911
- *Confluence Documentation & Meeting Records:(Google Drive):* https://drive.google.com/file/d/1WzK1K9C5D1rJl0ReKlg_BDSeTChIDLWz/view?usp=share_link

#### III.i Overleaf Capstone Reports

- *Capstone Part 1 Report:* https://www.overleaf.com/read/srpnhffghbhh
- *Capstone Part 2 Report:* https://www.overleaf.com/read/pqbdgrynncny

#### III.ii External Links

- *Discord*: https://discord.gg/KPZMG4ytdm
- *Google Drive*: https://drive.google.com/drive/folders/1jOl69VxlQHf852pdPQAE0giSAc0qm9T8?usp=sharing

## IV Decade-Wise Results

### IV.i 1988-1998

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | 0.33 | 0.06 |
| | With ENSO | 1 | 1 | 0.31 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.34 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 1 | 0.32 | 0.05 |
| Deseasoned Data | Without ENSO | 1 | - | 0.30 | 0.05 |
| | With ENSO | 1 | 1 | 0.27 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.31 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 1 | 0.29 | 0.05 |

Table 5: Maize from 1988-12-01 to 1998-12-01

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | -0.03 | 0.04 |
| | With ENSO | 1 | 2 | 0.08 | 0.04 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.00 | 0.04 |
| | Infl. Adjusted – with ENSO | 1 | 2 | 0.10 | 0.04 |
| Deseasoned Data | Without ENSO | 1 | - | -0.04 | 0.04 |
| | With ENSO | 1 | 2 | 0.09 | 0.04 |
| | Infl. Adjusted – W/o ENSO | 1 | - | -0.03 | 0.04 |
| | Infl. Adjusted – with ENSO | 1 | 2 | 0.10 | 0.04 |

Table 6: Soyabean from 1988-12-01 to 1998-12-01

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | 0.29 | 0.04 |
| | With ENSO | 1 | 3 | 0.29 | 0.04 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.31 | 0.04 |
| | Infl. Adjusted – with ENSO | 1 | 3 | 0.31 | 0.04 |
| Deseasoned Data | Without ENSO | 1 | - | 0.27 | 0.04 |
| | With ENSO | 1 | 3 | 0.26 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.28 | 0.04 |
| | Infl. Adjusted – with ENSO | 1 | 3 | 0.27 | 0.04 |

Table 7: Wheat from 1988-12-01 to 1998-12-01

### IV.ii 1998-2008

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 8 | - | 0.32 | 0.06 |
| | With ENSO | 8 | 11 | 0.30 | 0.06 |
| | Infl. Adjusted – W/o ENSO | 7 | - | 0.31 | 0.06 |
| | Infl. Adjusted – with ENSO | 7 | 11 | 0.30 | 0.06 |
| Deseasoned Data | Without ENSO | 7 | - | 0.28 | 0.06 |
| | With ENSO | 7 | 4 | 0.29 | 0.06 |
| | Infl. Adjusted – W/o ENSO | 7 | - | 0.28 | 0.06 |
| | Infl. Adjusted – with ENSO | 7 | 4 | 0.28 | 0.06 |

Table 8: Maize from 1998-12-01 to 2008-12-01

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | 0.34 | 0.05 |
| | With ENSO | 1 | 10 | 0.34 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.30 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 11 | 0.32 | 0.06 |
| Deseasoned Data | Without ENSO | 1 | - | 0.36 | 0.05 |
| | With ENSO | 1 | 10 | 0.36 | 0.05 |
| | Infl. Adjusted – W/o ENSO | 1 | - | 0.33 | 0.05 |
| | Infl. Adjusted – with ENSO | 1 | 11 | 0.35 | 0.05 |

Table 9: Soyabean from 1998-12-01 to 2008-12-01

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without ENSO | 1 | - | 0.18 | 0.04 |
| | With ENSO | 1 | 10 | 0.29 | 0.06 |
| | Infl. Adjusted – W/o ENSO | 7 | - | 0.22 | 0.05 |
| | Infl. Adjusted – with ENSO | 7 | 11 | 0.40 | 0.06 |
| Deseasoned Data | Without ENSO | 1 | - | 0.15 | 0.05 |
| | With ENSO | 1 | 10 | 0.28 | 0.04 |
| | Infl. Adjusted – W/o ENSO | 7 | - | 0.20 | 0.05 |
| | Infl. Adjusted – with ENSO | 7 | 10 | 0.34 | 0.06 |

Table 10: Wheat from 1998-12-01 to 2008-12-01

# V Observations and Results with other Teleconnections

# Observations

| Climate Teleconnection | Decade | Observations |
|---|---|---|
| AO Index | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| SAM Index | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO not significant **Soy:** EXO not significant |
| IOD Index | 2008-2018 | **Maize:** EXO not significant **Wheat:** EXO not significant **Soy:** EXO not significant |
| nhtele_index_1 | 2008-2018 | **Maize:** EXO not significant **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| nhtele_index_2 | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| nhtele_index_3 | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| nhtele_index_4 | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| pna_index | 2008-2018 | **Maize:** EXO Increases the correlation **Wheat:** EXO Increases the correlation **Soy:** EXO not significant |
| psa_index_1 | 2008-2018 | **Maize:** EXO not significant **Wheat:** EXO not significant **Soy:** EXO not significant |
| psa_index_2 | 2008-2018 | **Maize:** EXO not significant **Wheat:** EXO not significant **Soy:** EXO not significant |

1. **AO Index** - The Arctic Oscillation Index measures the atmospheric pressure patterns in the Arctic region, influencing weather and climate in mid-latitudes.

2. **SAM Index** - The Southern Annular Mode Index characterizes the north-south movement of westerly winds in the Southern Hemisphere, affecting weather and climate patterns.

3. **IOD Index** - The Indian Ocean Dipole Index gauges sea surface temperature differences in the Indian Ocean, impacting monsoons and climate in surrounding regions.

4. **nhtele_index_1** - Northern Hemisphere Teleconnection Index 1 represents a specific teleconnection pattern in the Northern Hemisphere affecting weather.

5. **nhtele_index_2** - Northern Hemisphere Teleconnection Index 2 denotes another teleconnection pattern in the Northern Hemisphere influencing climate variability.

6. **nhtele_index_3** - Northern Hemisphere Teleconnection Index 3 signifies an additional teleconnection pattern in the Northern Hemisphere linked to atmospheric circulation.

7. **nhtele_index_4** - Northern Hemisphere Teleconnection Index 4 corresponds to yet another teleconnection pattern in the Northern Hemisphere impacting weather and climate.

8. **PNA Index** - The Pacific/North American (PNA) Pattern Index describes the atmospheric circulation pattern over the Pacific and North America, influencing weather in those regions.

9. **psa_index_1** - Pacific South America Teleconnection Index 1 relates to a teleconnection pattern affecting weather and climate in the Pacific and South America.

10. **psa_index_2** - Pacific South America Teleconnection Index 2 represents an additional teleconnection pattern in the Pacific and South America region, influencing climate variability.

**Different indexes for the same teleconnection is possibly because of the different quantification methods used.**

**Artic Oscillations**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 6 | 0.19 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.20 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 7 | 0.24 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 7 | 0.23 | 0.05 |

**Artic Oscillations**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.04 |
| | With EXO | 2 | 7 | 0.25 | 0.04 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.04 |
| | Infl. Adjusted – with EXO | 2 | 7 | 0.23 | 0.04 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.04 |
| | With EXO | 2 | 7 | 0.25 | 0.04 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.04 |
| | Infl. Adjusted – with EXO | 2 | 7 | 0.23 | 0.04 |

**Artic Oscillations**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 2 | -0.04 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 2 | -0.04 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 8 | 0 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 8 | 0.01 | 0.05 |

**SAM Index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 3 | 0.21 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 3 | 0.21 | 0.04 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 3 | 0.22 | 0.04 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 3 | 0.20 | 0.05 |

**SAM Index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.04 |
| | With EXO | 2 | 6 | 0.17 | 0.04 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.04 |
| | Infl. Adjusted – with EXO | 2 | 8 | 0.15 | 0.04 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.04 |
| | With EXO | 2 | 6 | 0.21 | 0.04 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.04 |
| | Infl. Adjusted – with EXO | 2 | 6 | 0.18 | 0.04 |

**SAM Index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 8 | 0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 8 | 0.02 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 8 | 0.02 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 8 | 0.02 | 0.05 |

**nhtele_index_4**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 11 | 0.20 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 11 | 0.21 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 4 | 0.19 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 11 | 0.16 | 0.05 |

**nhtele_index_4**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 5 | 0.25 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 5 | 0.23 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 5 | 0.27 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 5 | 0.24 | 0.07 |

**nhtele_index_4**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 4 | 0.02 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | -0.06 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 9 | -0.03 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 9 | -0.03 | 0.05 |

**pna_index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 7 | 0.24 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 7 | 0.24 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 7 | 0.23 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 7 | 0.22 | 0.05 |

**pna_index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 7 | 0.29 | 0.06 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 7 | 0.27 | 0.06 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 7 | 0.3 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 7 | 0.28 | 0.06 |

**pna_index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 10 | 0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 9 | 0.06 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 10 | 0.02 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 10 | 0.02 | 0.05 |

**psa_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 5 | 0.17 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.19 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 5 | 0.16 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.17 | 0.05 |

**psa_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 11 | 0.20 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 11 | 0.16 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 11 | 0.21 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 11 | 0.17 | 0.07 |

**psa_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 3 | -0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 3 | -0.01 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 3 | -0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 3 | -0.01 | 0.05 |

**psa_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 3 | 0.22 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 10 | 0.16 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 3 | 0.20 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 10 | 0.12 | 0.05 |

**psa_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 4 | 0.20 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 4 | 0.18 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 4 | 0.20 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 8 | 0.17 | 0.07 |

**psa_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 10 | -0.03 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 10 | -0.02 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 10 | -0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 10 | -0.01 | 0.05 |

**nhtele_3 index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 1 | 0.23 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 1 | 0.24 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 3 | 0.20 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 3 | 0.18 | 0.05 |

**nhtele_3 index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 5 | 0.21 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 5 | 0.19 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 5 | 0.24 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 5 | 0.22 | 0.07 |

**nhtele_3 index**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 6 | 0.04 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.06 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 6 | 0.05 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.07 | 0.05 |

**IOD_INDEX**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 4 | 0.15 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 4 | 0.15 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 11 | 0.16 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 4 | 0.12 | 0.05 |

**IOD_INDEX**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 8 | 0.18 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 8 | 0.15 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 8 | 0.20 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 1 | 0.13 | 0.07 |

**IOD_INDEX**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 9 | 0.02 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 1 | -0.04 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 9 | 0.05 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 1 | -0.02 | 0.05 |

**nhtele_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 6 | 0.18 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.19 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 6 | 0.17 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.17 | 0.05 |

**nhtele_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 4 | 0.20 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 4 | 0.18 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 4 | 0.22 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 4 | 0.20 | 0.07 |

**nhtele_index_1**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 8 | -0.00 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 2 | 0.09 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 8 | 0.01 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 8 | 0.01 | 0.05 |

**nhtele_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Maize

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | 0.17 | 0.06 |
| | With EXO | 1 | 4 | 0.21 | 0.06 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.17 | 0.06 |
| | Infl. Adjusted – with EXO | 1 | 4 | 0.22 | 0.06 |
| Deseasoned Data | Without EXO | 1 | - | 0.15 | 0.05 |
| | With EXO | 1 | 4 | 0.19 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | 0.14 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 4 | 0.19 | 0.05 |

**nhtele_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Wheat

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 2 | - | 0.16 | 0.07 |
| | With EXO | 2 | 4 | 0.22 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.14 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 4 | 0.21 | 0.07 |
| Deseasoned Data | Without EXO | 2 | - | 0.19 | 0.07 |
| | With EXO | 2 | 4 | 0.24 | 0.07 |
| | Infl. Adjusted – W/o EXO | 2 | - | 0.16 | 0.07 |
| | Infl. Adjusted – with EXO | 2 | 4 | 0.22 | 0.07 |

**nhtele_index_2**

date_range = {"start": "2008-12-01", "end": "2018-12-01"}

Soyabean

| Type of Data | | Best ENDO lag | Best EXO Lag | Correlation | RMSE |
|---|---|---|---|---|---|
| Seasoned Data | Without EXO | 1 | - | -0.05 | 0.05 |
| | With EXO | 1 | 6 | -0.03 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.07 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | -0.03 | 0.05 |
| Deseasoned Data | Without EXO | 1 | - | -0.03 | 0.05 |
| | With EXO | 1 | 6 | -0.00 | 0.05 |
| | Infl. Adjusted – W/o EXO | 1 | - | -0.04 | 0.05 |
| | Infl. Adjusted – with EXO | 1 | 6 | 0.00 | 0.05 |