AMERICAN UNIVERSITY OF BEIRUT

Time Series Analysis and Forecasting of Food Prices in Lebanon

by

Wassim Radwan

Advisor(s)

Dr. Sirine Taleb

A Capstone Project
submitted in partial fulfillment of the requirements
for the degree of Master's in Business Analytics
to the Suliman S. Olayan School of Business
at the American University of Beirut

Beirut, Lebanon

August 2022

An Abstract of the Capstone Project of

Wassim Radwan for Master's in Business Analytics (MSBA)

Title: Time Series Analysis and Forecasting of Food Prices in Lebanon

The COVID-19 pandemic, financial crisis, port explosion, and the Russia-Ukraine war gave birth to a food crisis in Lebanon. Food prices have been skyrocketing during the last couple of years while more than half of the population has plunged into poverty leading to the destruction of the middle class in Lebanon. The capstone project is aimed at forecasting the price of three different food items which are sugar, rice, and meat. R programming language is used to build forecasting models. Benchmarking models such as the Average method, Naïve method, Seasonal Naïve method, and Random Walk Forecasts are implemented and their predictive performance is evaluated based on their RMSE value. ARIMA models with different parameters and with/without Box-Cox transformation are implemented as well as Simple Exponential Smoothing, Holt's Trend method (Damped & Undamped), Holt-Winters' Seasonal method (Additive & Multiplicative), and their predictive performance are evaluated based on their AICc value. The best model to predict the price of sugar for the upcoming year is the ARIMA (1,1,2) with transformation since it had the lowest AICc (-0.41241). The best model to predict the price of rice for the upcoming year is the ARIMA (0,1,1) with transformation since it had the lowest AICc (-129.2665). The best model to predict the price of meat for the upcoming year is the seasonal ARIMA (0,1,0)(0,0,1)[12] with transformation since it had the lowest AICc (-1.580995). These models could be used by UNESCWA and WFP to provide a better understanding of how the price of certain food items will change in the future. Future work could include experimenting with different forecasting models such as ARIMA with Fourier series.

Table of Contents

Introduction

In 2020, Lebanon was in the midst of a major financial crisis that saw the local currency weaken against the U.S. dollar by 90% as of today, suffering from the COVID-19 pandemic, and a fatal blast devastating the port of Beirut and the surrounding areas that resulted in 6000 people being injured and nearly 300,000 people losing their homes. The port comprised around 70% of Lebanon's trade shipments and served as a storage facility for grain. This has caused the decline in living conditions for millions of Lebanese citizens whereby approximately half the country's population is starving. The number of Syrian refugees has reached nearly 1.5 million which only exacerbated the food and economic crisis in Lebanon. 90% of Syrian refugees are classified as being extremely destitute to the point where the United Nations High Commissioner for Refugees (UNHCR) estimated that approximately two-thirds of Syrians seeking refuge in Lebanon are now skipping meals (Chehayeb, 2021). Food items are about ten times more expensive than they were back in 2019 whereby the price of 1 kilogram of rice was 1,100 LBP in 2019 but has drastically increased to a whopping 21,000 LBP in 2021 (Hussein, 2021). Lebanon's socioeconomic downturn is worsened by the fact that 80% of the country's wheat is imported from both Russia and Ukraine, two countries currently engaged in a deadly war that does not appear to be ending anytime soon, and millions of Lebanese people's lives are dependent on the silos filled with grain that are stuck in crucial Ukrainian ports such as those in Mykolaiv and Odesa (Melillo, 2022). To make matters worse, the Lebanese government's ability to import is increasingly being constrained because of the diminishing dollar reserves (Chehayeb, 2020). Also, many other countries import their wheat from Russia and Ukraine, rendering Lebanon unable to outbid more powerful competitors thus mitigating the needed bargaining power to obtain a sustainable supply. The country's policies, which could be described as economically neoliberal, caused trade and services to be promoted over domestic agricultural production, which led to a lacking food security infrastructure as well as elitist financial policies that neglected the poorer classes of society (Halabi, 2022). The main purpose of this project is to provide an accurate assessment of the progression of food prices in Lebanon over the last decade, as well as building diversly multiple prediction models to forecast what the price of certain food types will be next year based on their historical data performance. The food types to be analyzed are as follows:

- Sugar ($/Kg)
- Rice ($/Kg)
- Meat ($/200 g)

The prediction models and visualizations used are generated by using the R programming language, and they include Benchmarking models (Average, Naïve, Seasonal Naïve, Random Walk Forecast), and ARIMA (AutoRegressive Integrated Moving Average) models with and without Box-Cox transformation. Simple Exponential Smoothing, Holt's model (damped and non-damped), and Holt's-Winter model (Additive and Multiplicative). The benchmarking models will be evaluated based on their Root-Mean-Squared error (RMSE) performance while the rest of the models will be evaluated based on their Akaike's Information Criterion (corrected) performance (AICc). Other visualizations have been dashboarded using Tableau. Lastly, a user-friendly web application was made using the shiny package in R to be accessed by interested parties.

Background and Related Work

Forecasting food prices using prediction models has been performed many times. In 2018, a study was conducted by Ohyver and Pudjihastuti to develop ARIMA models to predict the price of medium-quality rice to forecast its fluctuation to aid the Indian government in controlling and stabilizing it and hence attain a decent development of domestic trade. Two ARIMA models were can be utilized to predict the price of medium-quality rice which are ARIMA (2,0,2) and ARIMA (1,1,2). The study concluded that the ARIMA (1,1,2) model was found to be better than the ARIMA (2,0,2) model since it had a lower RMSE, AIC, and MAPE (Mean Absolute Percentage Error), i.e., it had good accuracy (Ohyver & Pudjihastuti, 2018). Another study conducted in 2021 utilized different techniques to predict the sale prices of an Italian food wholesaler. The records in the wholesaler's weekly dataset were accumulated in a period between 2013 and 2021, and the study focused on three products that had the most frequency of records which were found to be Carnaroli rice 1 kg x 10, Gorgonzola cheese 1/8 of wheel 1.5 kg, and Cured aged ham 6.5 kg. Techniques such as ARIMA models, Prophet which is Facebook's adaptable forecasting tool, and deep learning models employing Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) were used to predict the price of products, and the results were compared between them. In general, ARIMA models provide a robust benchmark for econometric analyses. The study concluded that the predictive performance of LSTM neural networks and ARIMA models were analogous, whereas the combination of LSTMs and CNNs, which necessitates more time for tuning, yielded the best accuracy, while Prophet's accuracy was much lower compared to the rest of the techniques even though it was quicker to set up and tune and the data did not need to be pre-processed (Menculini et al., 2021). In 2022, a study was conducted to determine the elements that influence China's price fluctuations of vegetables in a horizontal dimension and vertical dimension during the COVID-19 pandemic, and then implement an ARIMA model and Naïve benchmarking model and evaluate their predictive performance. A vegetable price dataset comprising 51,567 records was used. The vegetables analyzed in the study were chosen as representatives of their respective categories and they are carrot (root vegetables), cabbage (leafy vegetables), and eggplant (solanaceous fruit vegetables). The study concluded that the ARIMA model outperformed the Naïve model since it had a lower RMSE, MAE (Mean Absolute Percentage Error), and MAPE (Mao et al., 2022). In 2017, a study was conducted to forecast the price of Indian agricultural food types, mainly Maize, Paddy, and Ragi. The time series data ranged from 2002 to 2016, and univariate ARIMA models were implemented and their predicted performance was evaluated. The study concluded that ARIMA (1,1,1), ARIMA (1,1,2), and ARIMA (1,2,1) were found to be the best models for paddy, ragi, and maize, respectively since their AIC and SBC (Shwartz Basic Criteria) values were the lowest (Gaddi, 2017).

Methodology

Approach

This section outlines the steps taken to render the analyses of the time series and modeling procedure as well.

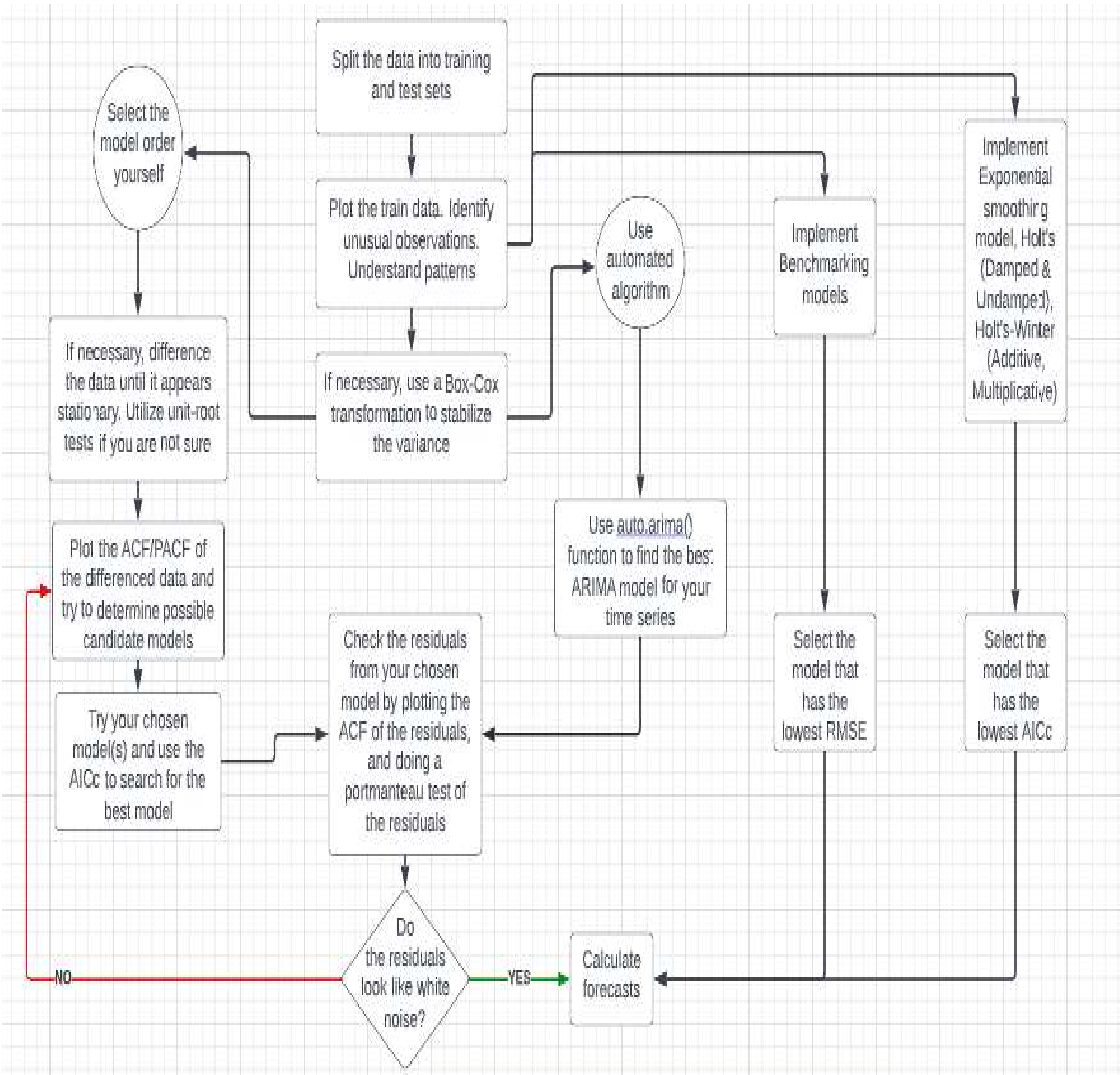


Figure 1- General Modeling Procedure

Tools and Software

This section discusses the tools and software used in the development of the Capstone project.
Rstudio: the main integrated development environment (IDE) for R, a programming language for statistical computing and graphics. The code was written using Rstudio.

Tableau: A visualization tool used to explore the characteristics of the dataset and render simple visualizations and dashboards for the report.

Shiny: an R package that enables building interactive web applications straight from R. The web application for the project was made using Shiny.

fpp2: A package in R that loads data required for the examples and exercises used in the book Forecasting: Principles and Practice (2nd edition) by Rob J Hyndman and George Athanasopoulos. It also loads several packages needed to do the analysis described in the book. It includes the following packages:

- **forecast**, for forecasting methods and some data sets.
- **ggplot2**, for data visualization.
- **fma**, for data taken from the book "Forecasting: methods and applications" by Makridakis, Wheelwright and Hyndman (1998).
- **expsmooth**, for data taken from the book "Forecasting with Exponential Smoothing" by Hyndman, Koehler, Ord and Snyder (2008).

Organization(s)

United Nations Economic and Social Commission for Western Asia (UNESCWA)

UNESCWA is one of the five regional commissions under the jurisdiction of the United Nations Economic and Social Council is comprised of a total of 20 member states, all from the regions of the Middle East and North Africa. The Commission's role is to promote the economic and social development of Western Asia through regional and subregional integration and cooperation.

World Food Programme (WFP)

Is the food-assistance branch of the United Nations and the largest organization in the world that focuses on food security and starvation. It was established in 1961, headquartered in Rome, Italy, and is present in 80 different countries across the globe. It offers other services such as development and technical assistance, managing supply chains and logistics, etc. Also, it is an executive member of the United Nations Sustainable Development Group whose goal is to satisfy the 17 Sustainable Development Goals (SDG), mainly SDG 2 for "zero hunger" by 2030.

Data

The dataset was kindly provided to me by Dr. Mohammad Zuheir Bakleh. It is a public dataset that is used by UNESCWA employees for research and analytical purposes. Since this project's main emphasis is univariate time series analysis and forecasting, then the dataset contains only a few features which are the following: date – code – usdprice. The variable code indicates the food type we are dealing with. The food types are sugar, rice, and meat. Since the goal is to analyze and predict the price of each food type, it was more fitting to subset the original datasets into three different datasets, i.e., one for each food type. Hence, the variable code was deemed irrelevant and was not included in the three new datasets. The period of data selected for each food type and its price ranged from the beginning of 2013 to the end of 2021. All datasets are monthly hence each dataset contained 108 records or rows.

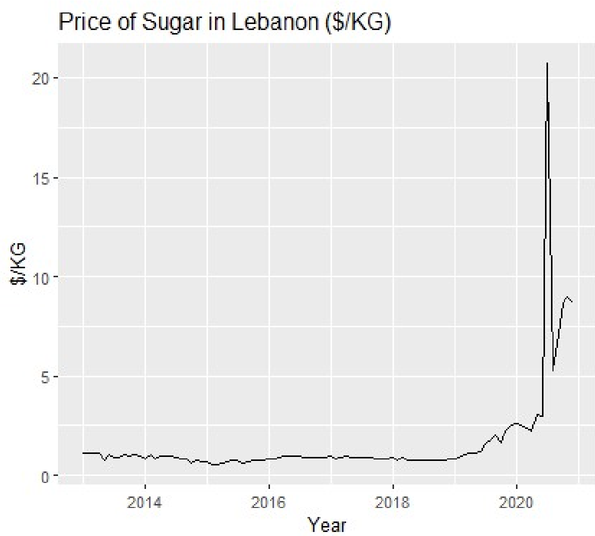
Summary Statistics

Food Type	Minimum	1st Quartile	Mean	Standard Deviation	3rd Quartile	Maximum
Sugar ($/kg)	0.5257	0.7998	1.8141	2.847872	1.1209	20.7221
Rice ($/kg)	1.051	1.187	2.099	2.324624	1.509	13.468
Meat ($/200 g)	0.6572	1.15	2.0227	1.955317	1.4965	9.9579

Table 1- Descriptive Statistics

Time-Series Analysis

Before plotting the time series, we split the data for each dataset into training and test splits whereby the training set period ranged from 2013 to the end of 2020, and the test split period ranged from 2021 to the end of 2021, i.e., each training set has 96 records and each test set has 12 records.



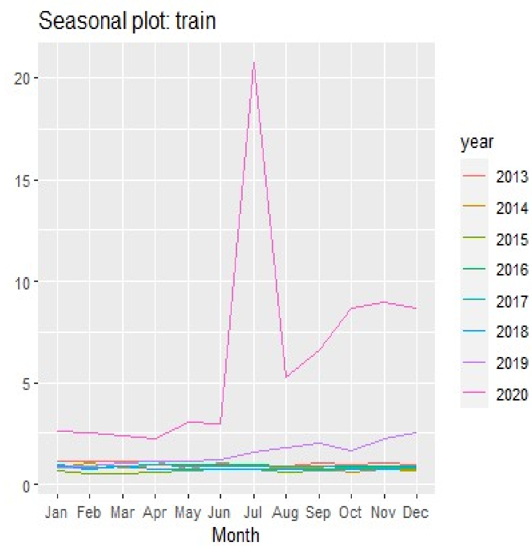
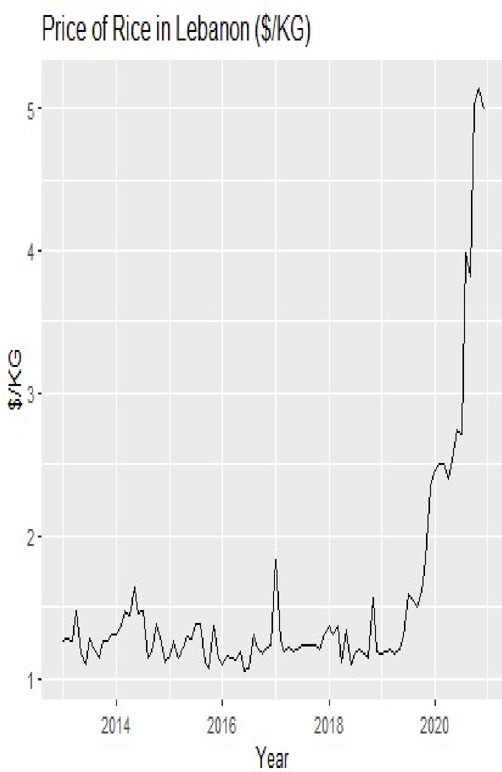


Figure 2- Time plot of the price of sugar in Lebanon

Figure 3- Seasonal plot of the price of sugar in Lebanon

From 2013 to 2019, the price has been relatively stable. However, we observe a rapidly increasing trend during 2020 followed by a sharp decrease shortly after, followed by a slight increase. Thus, a changing trend is observed. The data shows no signs of a significant seasonality



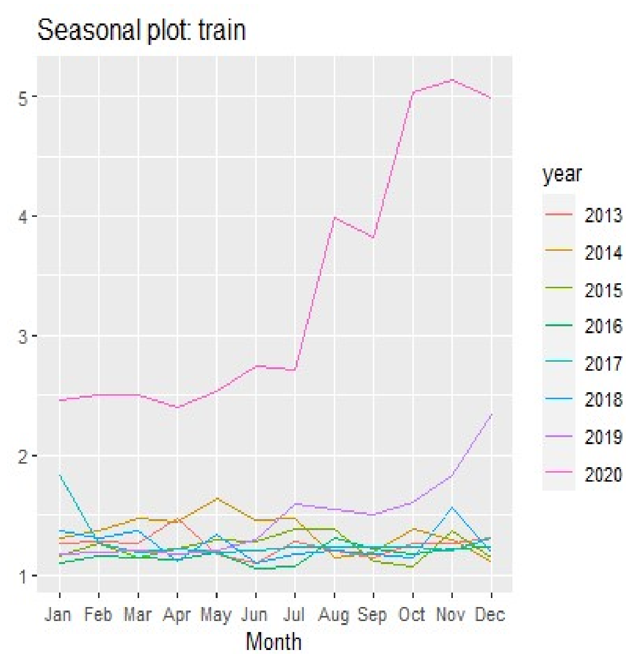


Figure 4- Time plot of the price of rice in Lebanon

Figure 5- Seasonal Plot of the price of rice in Lebanon

From 2013 to 2019, the price has been fluctuating. However, we observe a rapidly increasing trend during 2020 and onwards. The data shows no signs of significant seasonality.

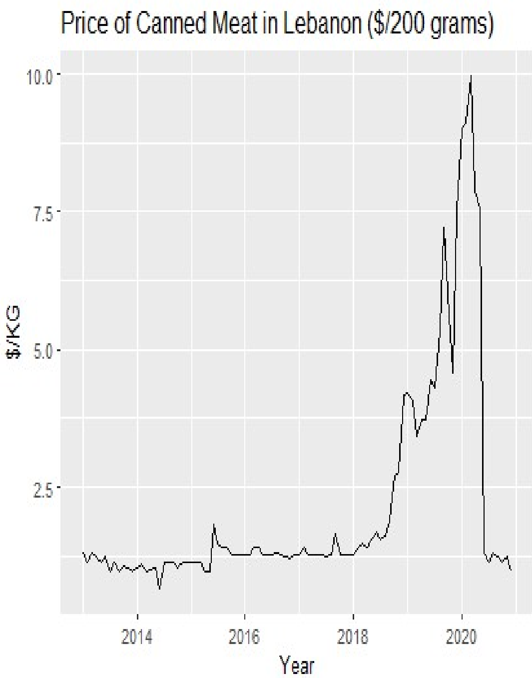


Figure 6- Time plot of the price of meat in Lebanon

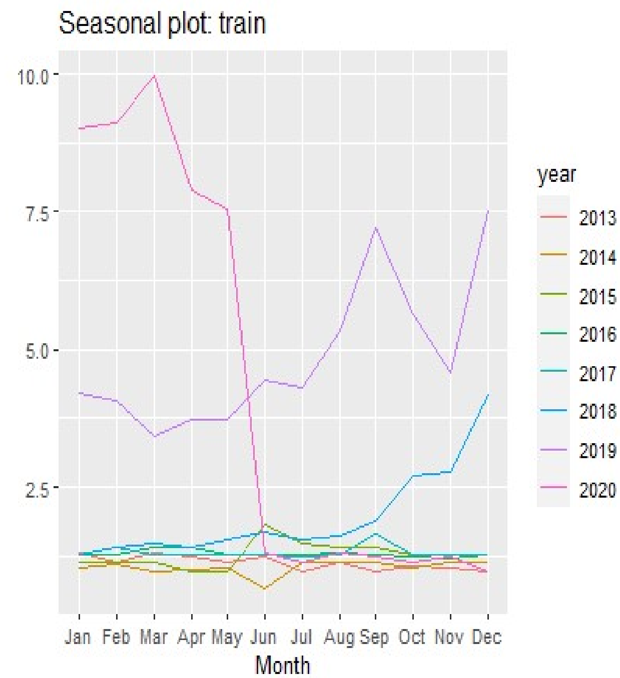


Figure 7- Seasonal plot of the price of meat in Lebanon

From 2013 to around 2018, the price has been relatively stable. However, we observe an increasing trend starting from 2019 followed by a decrease shortly after, followed by a slight increase, then finally a sharp decrease. Thus, a changing trend is observed. The data shows no signs of a significant seasonality.

Techniques and Approaches

After analyzing the time series and checking for trends and patterns for each dataset, several different types of models were built to predict what the price of each food type would be like next year. The prediction models to be used are the following:

1. **Benchmarking models**

- Average method: forecasts of all future values are equal to the average of the historical data.
- Naïve method: all forecasts are set to be the value of the last observation
- Seasonal Naïve method: each forecast is set to be equal to the last observed value from the same season of the year
- Random walk forecast: enables forecasts to increase or decrease over time, whereby the amount of change is called the drift and it is set to be the average change observed in the historical data.

2. ARIMA

A combination of differencing with autoregression and a moving average model. ARIMA stands for Autoregressive Integrated Moving Average. It has three parameters:

- p = order of the autorgressive part
- d = degree of first differencing involved
- q = order of the moving average part

We also implemented ARIMA modeling to the datasets after transforming using Box-Cox transformation, which is dependent on the lambda parameter. Transformation is done to stabilize the variance of the time series. We can transform the dataset automatically via R's BoxCox.lambda() function, but it was found to be ineffective since it only increased the variance of the time series rather than stabilizing it. Hence, a lambda value of 0 was set manually to transform the time series. The advantage of setting lambda to 0 (Natural logarithm or log transformation) is that it forces forecasts and prediction intervals to be positive.

3. Simple Exponential Smoothing

A simple method well-suited to forecast data that does not show signs of a clear trend or seasonal pattern. The only component included in the forecast equation is the level component.

4. Holt's linear trend method

Enables simple exponential smoothing to forecast data with a clear trend. The forecast equation includes a level component and a trend component. The method can also be **damped** by including a damping parameter that dampens the trend so that it approaches a constant some time in the future. For each food type in this project, the damping parameter was set to be 0.90.

5. Holt-Winters' seasonal method

An extension of Holt's method to capture the seasonality in the time series. The forecast equation includes three components which are level, trend, and seasonal components. This method has two variations which are the additive and multiplicative methods. The additive method is preferred if the seasonal variations in the time series are approximately constant. The multiplicative method is preferred if the seasonal variations are changing proportionally to the level of the series.

Accuracy Measures

The accuracy measures used to evaluate the predictive performance of the models are the AICc and RMSE. For benchmarking models, the RMSE will be used to evaluate their predictive performance, whereas the rest of the models' predictive performances will be evaluated based on the AICc. The reasoning for this is that benchmarking models are normally evaluated by their RMSE, but the other models (ARIMA, Holt's, etc.) are normally evaluated based on their AICc due to ease of interpretability and explainability which balances complexity and accuracy. Both are included in the project to allow the user to decide which model is the best based on his/her preferred accuracy measure.

Results

Sugar

Benchmarking Models

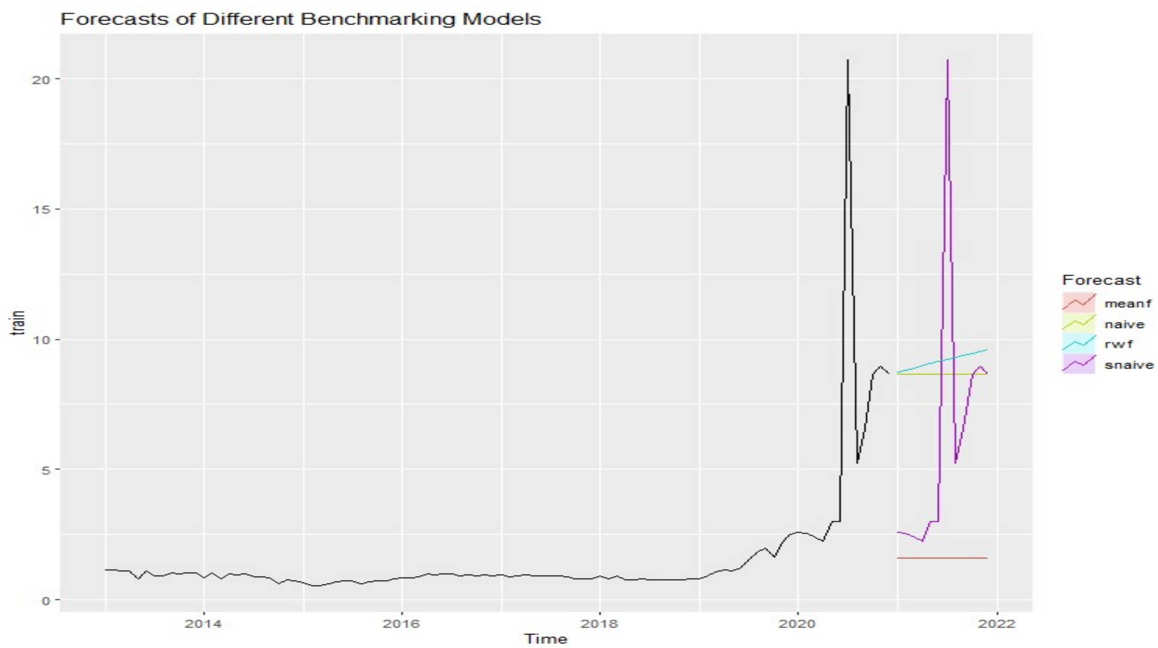


Figure 8- Forecasts of benchmarking models for the price of sugar

Benchmarking models	RMSE
Average method	4.684023
Naïve forecast	6.73861
Seasonal Naïve forecast	8.363276
Random walk forecast	7.266073

Table 2- Accuracy measures of benchmarking models for the price of sugar

For sugar, the best benchmarking model was the average method since it yielded the lowest RMSE value (4.684023)

ARIMA

To model the price of sugar data using ARIMA, we first differenced our data since the training data is not stationary because the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test yielded a test-statistic value of 0.8834 which is greater than the 1% critical value of 0.739. After differencing the data, it passed the KPSS test since the test statistic (0.128) is lower than all of the critical values indicating that the data is stationary and non-seasonal.

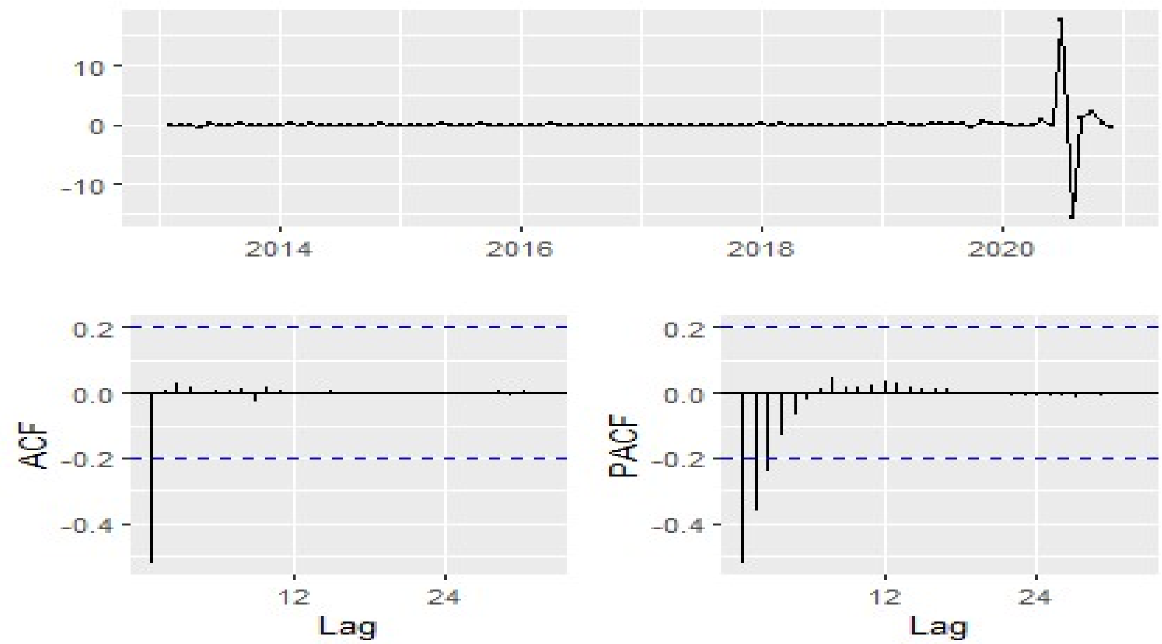


Figure 9- Time plot, ACF, and PACF of sugar data

Before considering which ARIMA model should be used, we first must check the ACF and PACF plots to figure out our starting point in the modeling procedure. After plotting the differenced training set (Figure 9), we can see PACF is sinusoidal and the ACF only has one significant spike. In our case, we have p=3 and d=1 as our initial model to start with. After starting with the initial model, we tested a total of 7 models, collected their AICc values, and tabulated them.

Models	**AICc**
ARIMA (1,1,0)	412.9359
ARIMA (0,1,1)	396.5468
ARIMA (1,1,1)	397.1608
ARIMA (2,1,1)	398.2497
ARIMA (0,1,2)	396.5687
ARIMA (2,1,0)	402.2507
AUTO.ARIMA (1,1,2) with drift	394.932

Table 3- Accuracy measures of sugar ARIMA models

As shown in Table 3, ARIMA (1,1,2) with drift is the best model since it had the lowest AICc (394.932).

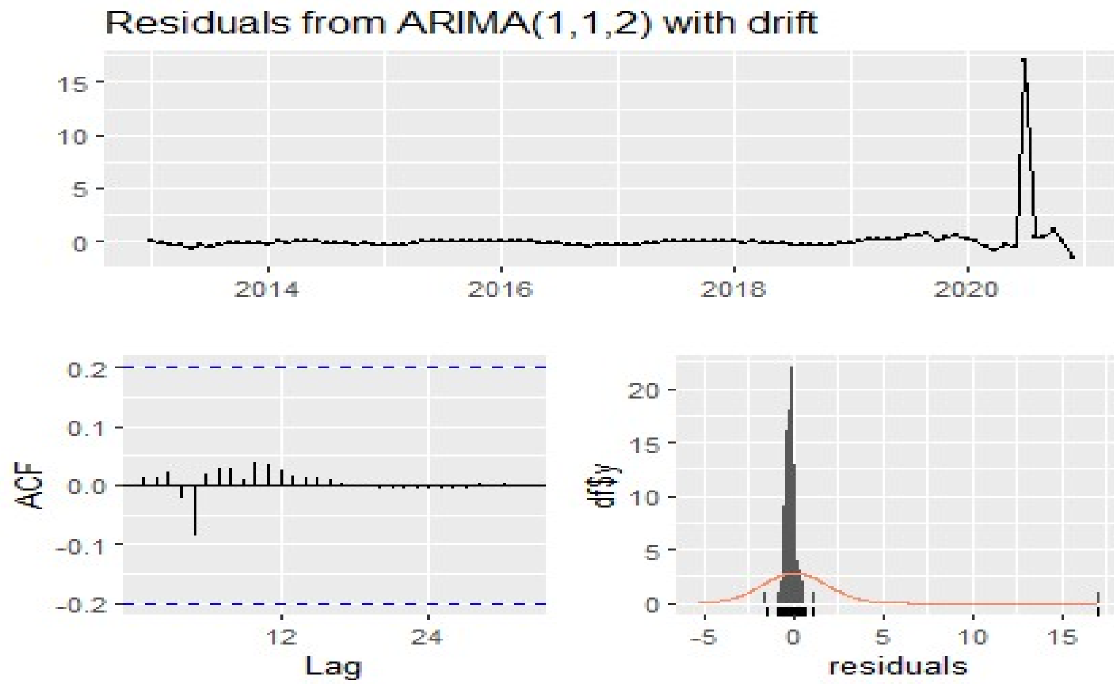
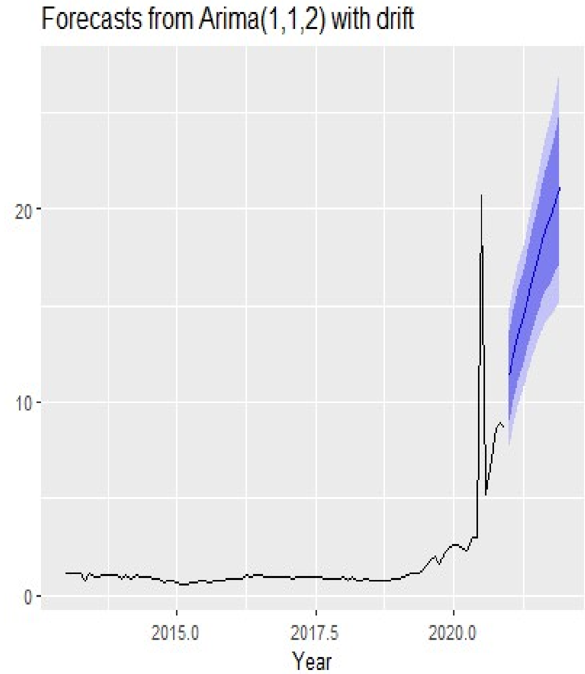


Figure 10- Time plot, ACF, and histogram of the best ARIMA model for the price of sugar

Variance is constant (Homoscedasticity) except for a few outliers past 2020. The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays no autocorrelations,i.e., all spikes are within the blue-dashed bounds. In addition, this model passed the Ljung- Box test with a p-value of 1 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.



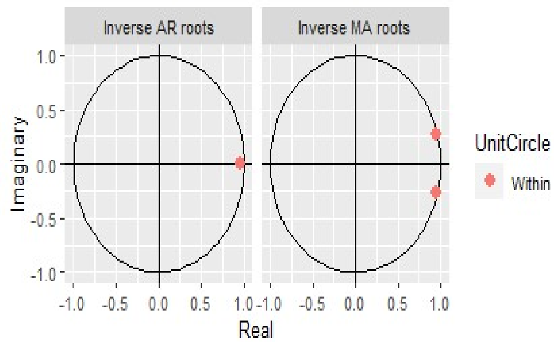


Figure 11-Forecasts of the best ARIMA model for the price of sugar/ Figure 12- Inverse characteristic roots for ARIMA(1,1,2) with drift fitted to the training set

All red dots lie inside the unit circles. Hence, the fitted model is both stationary and invertible.

ARIMA (with Box-Cox Transformation)

We first differenced our logarithmically transformed (i.e., lambda = 0) training data since it is not stationary because the KPSS test yielded a test-statistic value of 1.1483 which is greater then the 1% critical value of 0.739. After differencing the transformed data, it passed the KPSS test since the test statistic (0.5152) is lower than the 1% critical value of 0.739 indicating that the transformed data is stationary and non-seasonal.

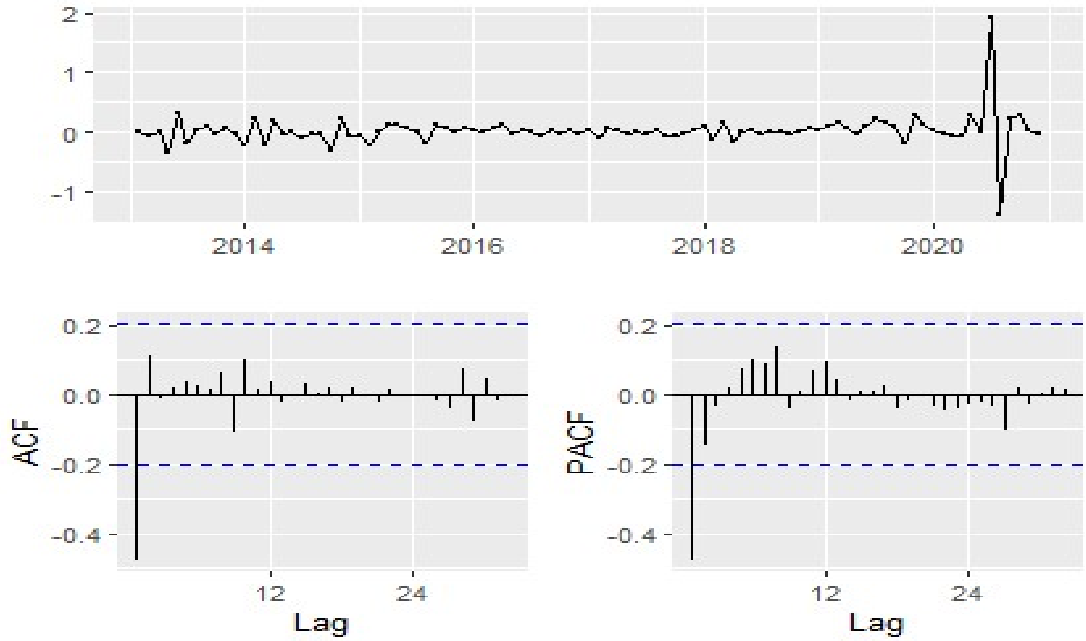


Figure 13- Time plot, ACF, and PACF of transformed sugar data

After plotting the transformed training dataset (Figure 13), we can see that the ACF and PACF are sinusoidal. In this case, p=1 and d=1 is the initial model to start with. After starting with the initial model, we tested a total of 8 models, collected their AICc values, and tabulated them.

Models	**AICc**
ARIMA (1,1,0)	3.476014
ARIMA (0,1,1)	4.352755
ARIMA (1,1,1)	4.209605
ARIMA (1,1,2)	-0.41241
ARIMA (2,1,1)	6.25246
ARIMA (0,1,2)	3.739238
ARIMA (2,1,0)	4.076606
AUTO.ARIMA (1,1,2)	-0.41241

Table 4- Accuracy measures of transformed sugar ARIMA models

As shown in Table 4, ARIMA (1,1,2) is the best model since it had the lowest AICc (-0.41241).

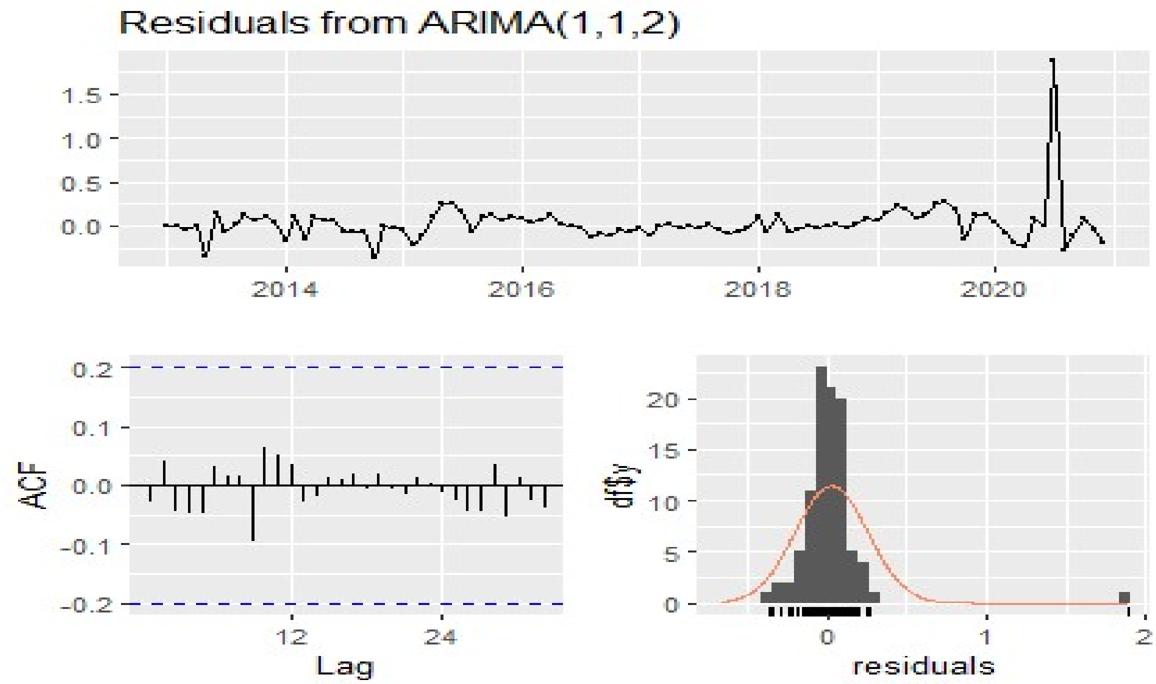
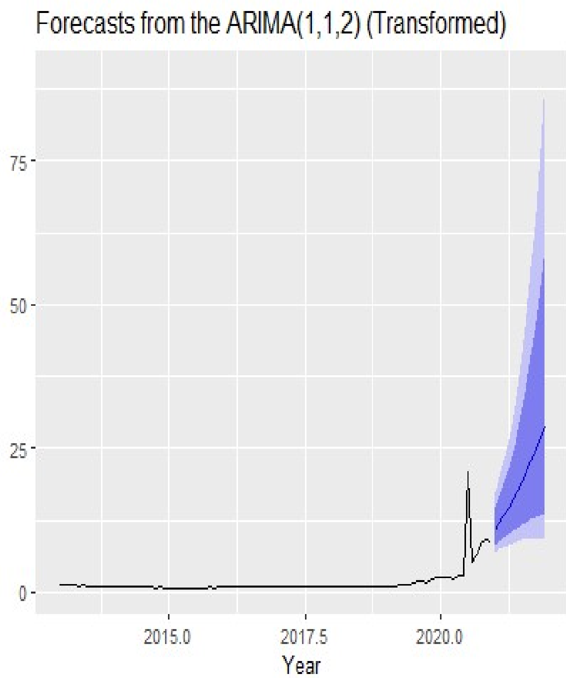


Figure 14 - Time plot, ACF, and histogram of the best ARIMA model for the transformed price of sugar

Variance is constant (Homoscedasticity) except for a few outliers past 2020. The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays no autocorrelations,i.e., all spikes are within the blue-dashed bounds. In addition, this model passed the Ljung- Box test with a p-value of 1 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.



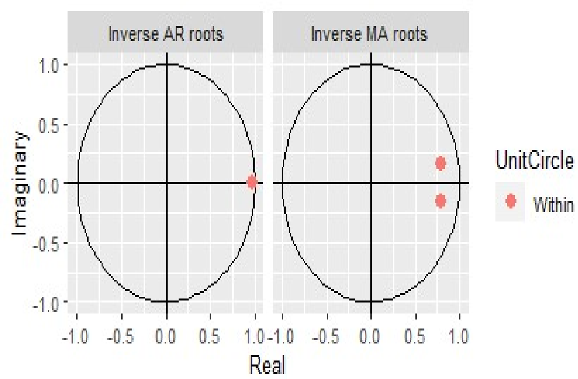


Figure 16- Inverse characteristic roots for ARIMA(1,1,2) fitted to the training set

Figure 15- Forecasts of the best ARIMA model with Transformation for the price of sugar

All red dots lie inside the unit circles. Hence, the fitted model is both stationary and invertible.

Simple Exponential Smoothing vs. Holt’s method vs. Holt-Winters’ method

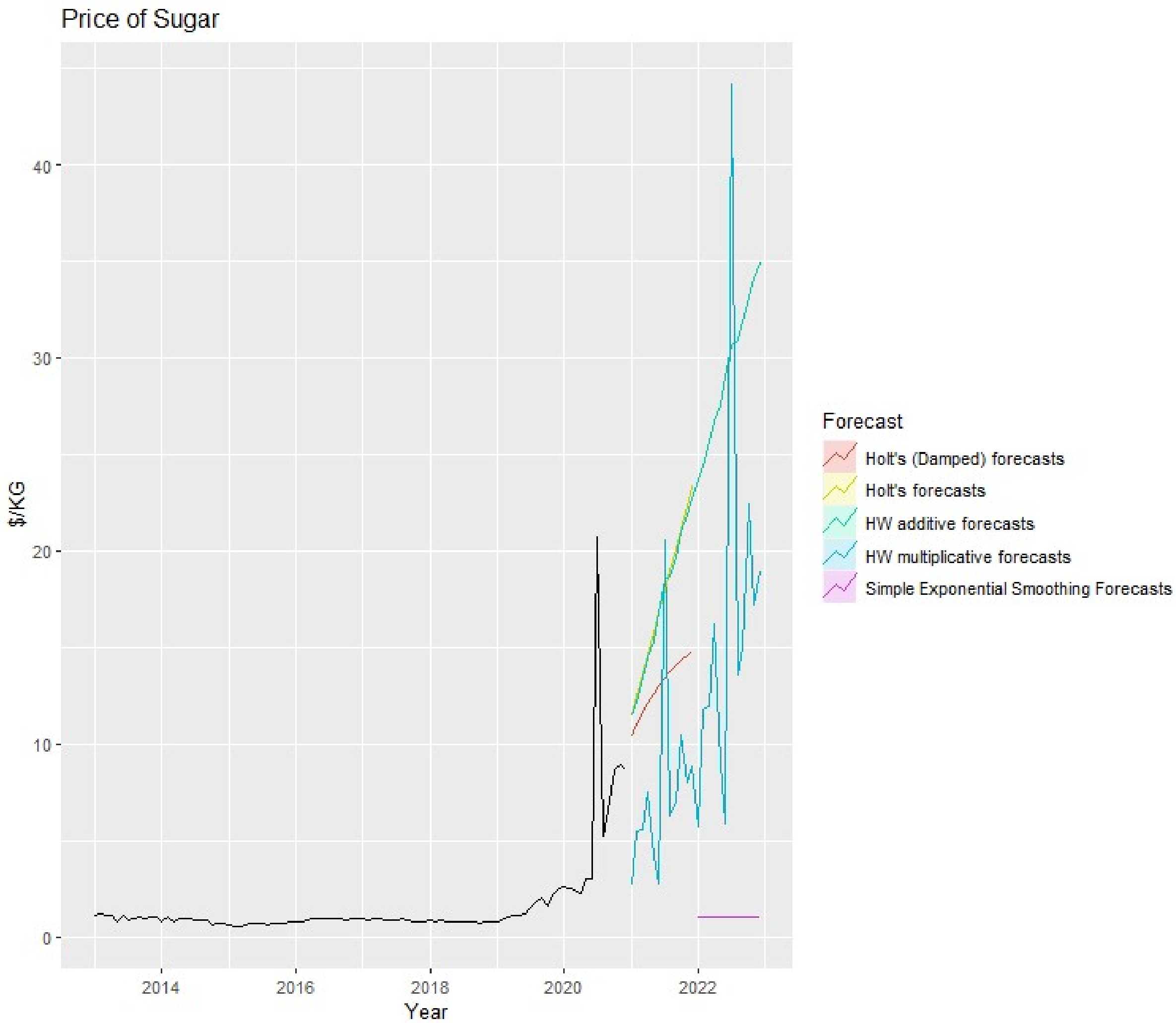


Figure 17- Forecasts of the sugar price from simple, trend, and seasonal models

Model	AICc
Simple exponential smoothing	679.8811
Holt’s trend	560.4550
Holt’s trend (Damped)	563.7557
Holt-Winter’s (Additive)	586.7083
Holt-Winter’s (Multiplicative)	1304.327

Table 5- Accuracy measures of simple, trend, and seasonal models of the price of sugar

Holt's trend model is the best since it had the lowest AICc (560.4550).

Thus, the best model across all models is the ARIMA (1,1,2) with transformation since it had the lowest AICc (-0.41241) across ALL models.

Rice
Benchmarking Models

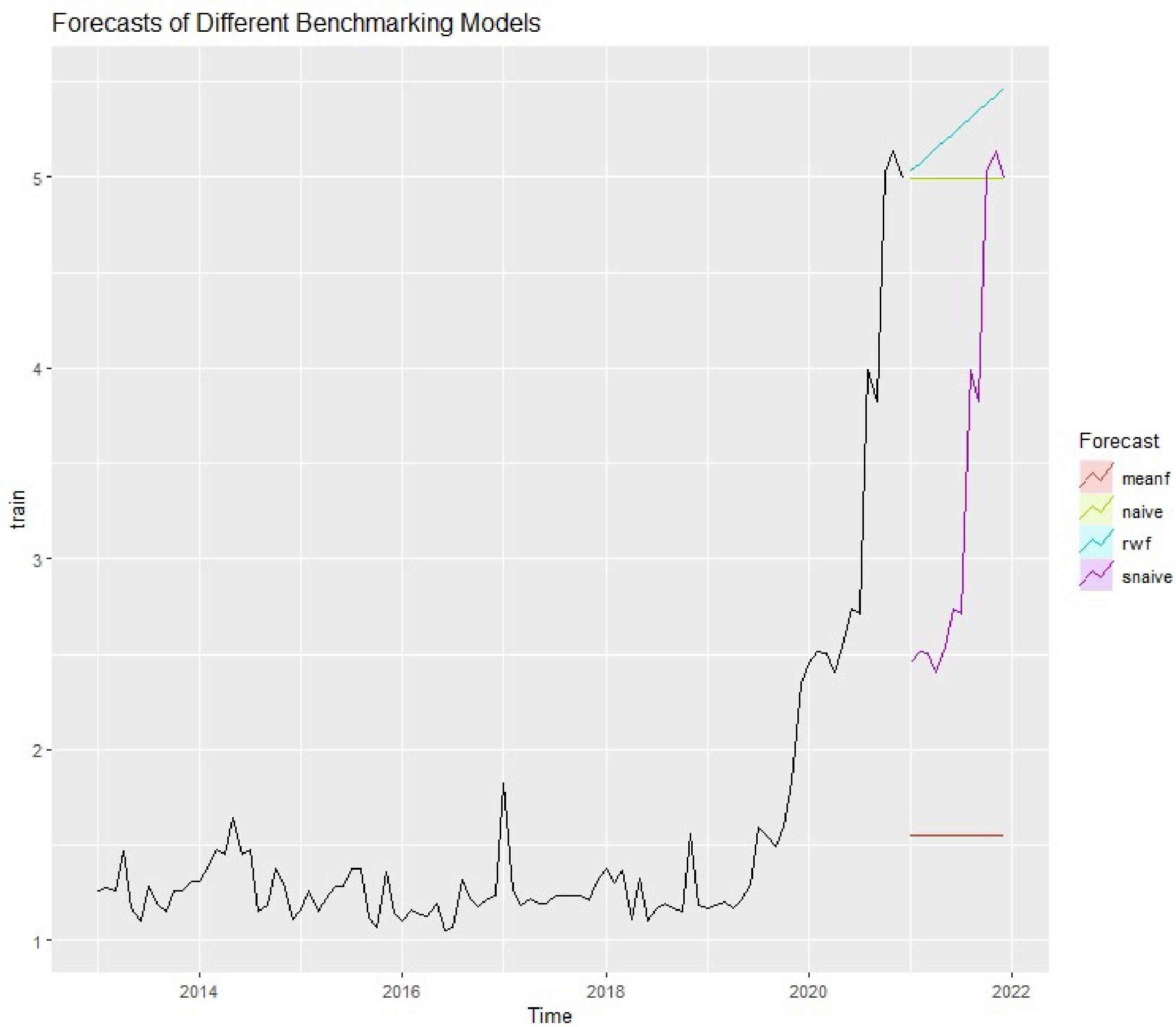


Figure 18- Forecasts of benchmarking models for the price of rice

Benchmarking models	RMSE
Average method	6.749286
Naïve forecast	4.844064
Seasonal Naïve forecast	6.364649
Random walk forecast	4.863799

Table 6- Accuracy measures of benchmarking models for the price of rice

For rice, the best benchmarking model was the Naive method since it yielded the lowest RMSE value (4.844064)

ARIMA

To model the price of rice data using ARIMA, we first differenced our data since the training data is not stationary because the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test yielded a test-statistic value of 0.9598 which is greater than the 1% critical value of 0.739. After performing second-order differencing, i.e., differencing the data for a second time, it passed the KPSS test since the test statistic (0.0261) is lower than all of the critical values indicating that the data is stationary and non-seasonal.

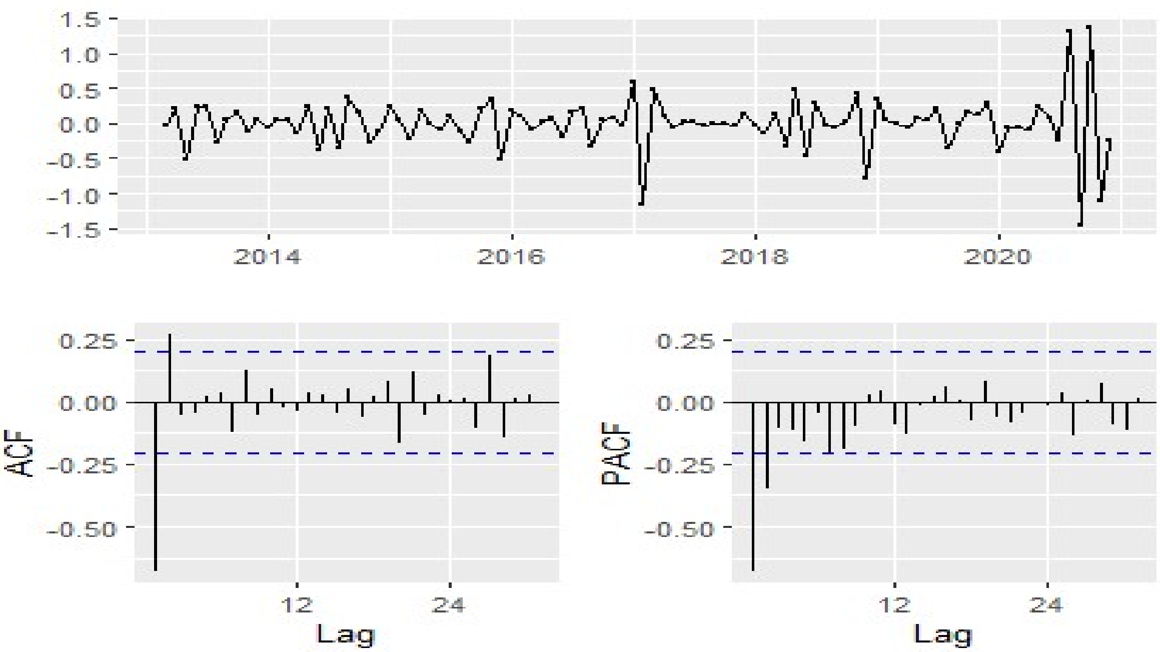


Figure 19- Time plot, ACF, and PACF of rice data

Before considering which ARIMA model should be used, we first must check the ACF and PACF plots to figure out our starting point in the modeling procedure. After plotting the differenced training set (Figure 19), we can see PACF is sinusoidal and the ACF only has one significant spike. In our case, we have p=2 and d=2 is the initial model to start with. After starting with the initial model, we tested a total of 7 models, collected their AICc values, and tabulated them.

Models	AICc
ARIMA (1,1,0)	3.11701048
ARIMA (0,1,1)	3.98229036
ARIMA (1,1,1)	3.54808746
ARIMA (1,1,2)	-0.5067714
ARIMA (0,1,2)	-1.9986324
ARIMA (2,1,0)	0.07403545
AUTO.ARIMA (1,2,1)	-3.3663024

As shown in Table 7, ARIMA (1,2,1) is the best model since it had the lowest AICc (-3.36630239).

Table 7- Accuracy measures of rice ARIMA models

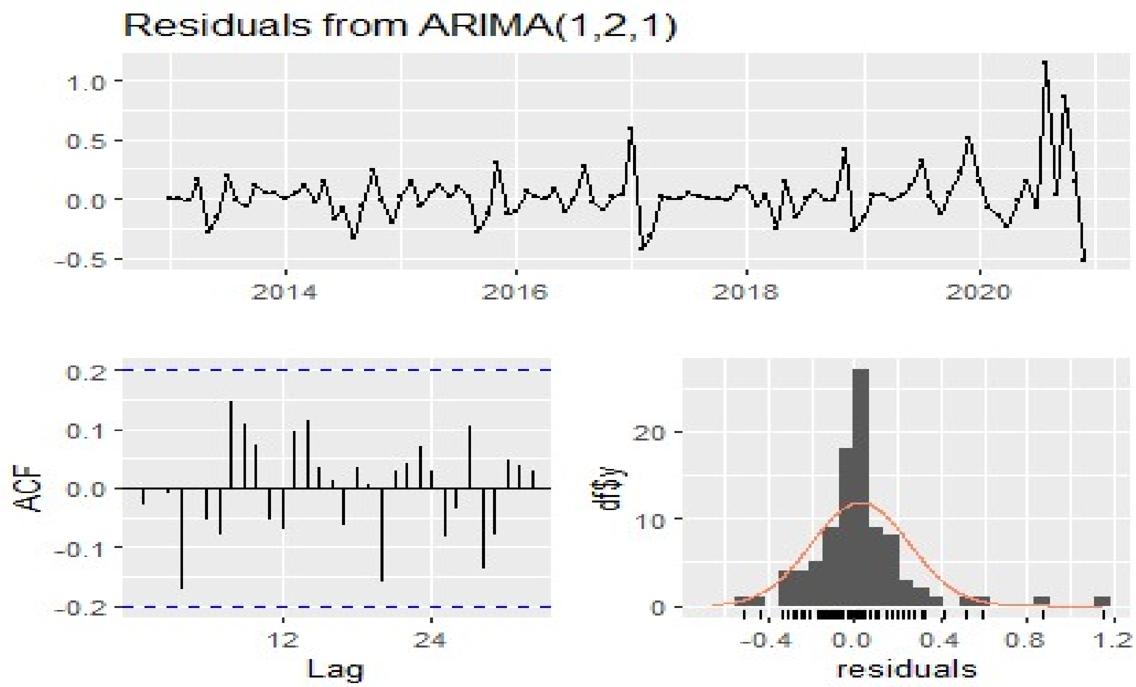


Figure 20- Time plot, ACF, and histogram of the best ARIMA model for the price of rice

Variance is constant (Homoscedasticity) except for a few outliers past 2020. The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays no autocorrelations,i.e., all spikes are within the blue-dashed bounds. In addition, this model passed the Ljung- Box test with a p-value of 0.7697 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.

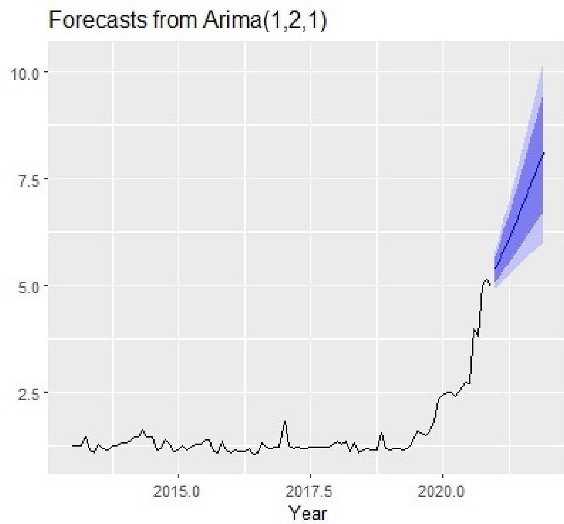


Figure 21 -Forecasts of the best ARIMA model for the price of rice

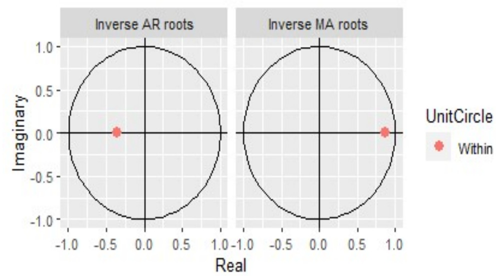


Figure 22- Inverse characteristic roots for ARIMA(1,2,1) fitted to the training set

All red dots lie inside the unit circles. Hence, the fitted model is both stationary and invertible.

ARIMA (with Box-Cox Transformation)

We first differenced our logarithmically transformed (i.e., lambda = 0) training data since it is not stationary because the KPSS test yielded a test-statistic value of 1.0398 which is greater than the 1% critical value of 0.739. After differencing the transformed data twice (second-order differencing), it passed the KPSS test since the test statistic (0.0229) is lower than the 1% critical value of 0.739 indicating that the transformed data is stationary and non-seasonal.

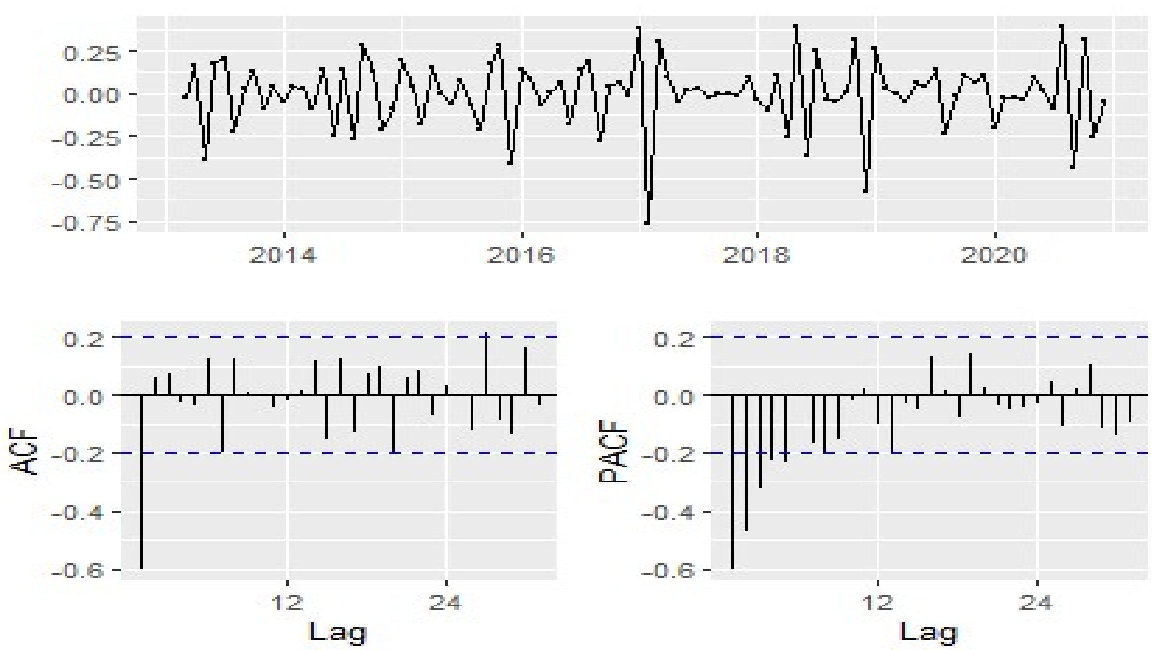


Figure 23- Time plot, ACF, and histogram of the best ARIMA model for the transformed price of rice

After plotting the transformed training dataset (Figure 23), we can see that the ACF and PACF are sinusoidal. In this case, p=3 and d=2 is the initial model to start with. After starting with the initial model, we tested a total of 8 models, collected their AICc values, and tabulated them.

Models	AICc
ARIMA (1,1,0)	-128.8224
ARIMA (0,1,1)	-129.2665
ARIMA (1,1,1)	-127.2422
ARIMA (2,1,1)	-125.6305
ARIMA (0,1,2)	-127.3392
ARIMA (2,1,0)	-127.6203
AUTO.ARIMA (0,2,2)	-128.4593

Table 8- Accuracy measures of transformed rice ARIMA models

As shown in Table 8, ARIMA (0,1,1) is the best model since it had the lowest AICc (AICc = -129.2665).

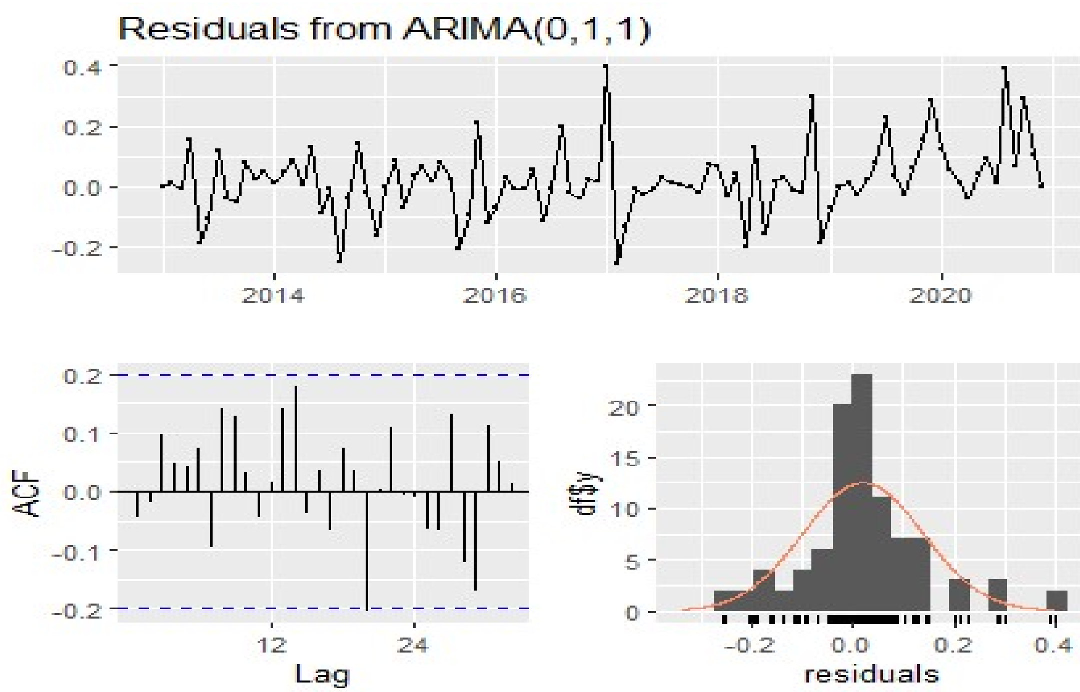
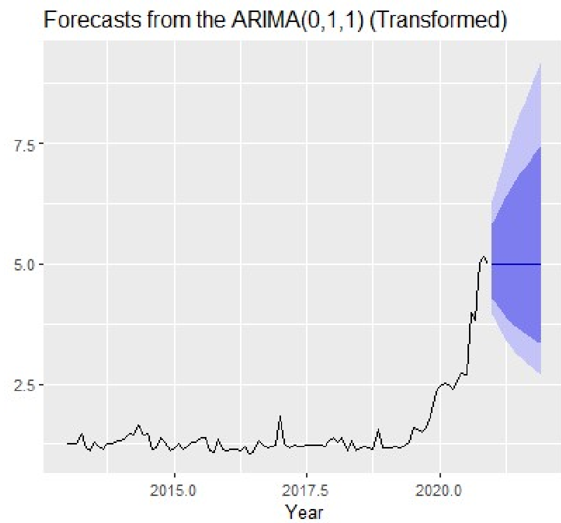


Figure 24- Time plot, ACF, and histogram of the best ARIMA model for the transformed price of rice

Variance is constant (Homoscedasticity). The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays one autocorrelation, but it is not significant. In addition, this model passed the Ljung- Box test with a p-value of 0. 0.6554 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.



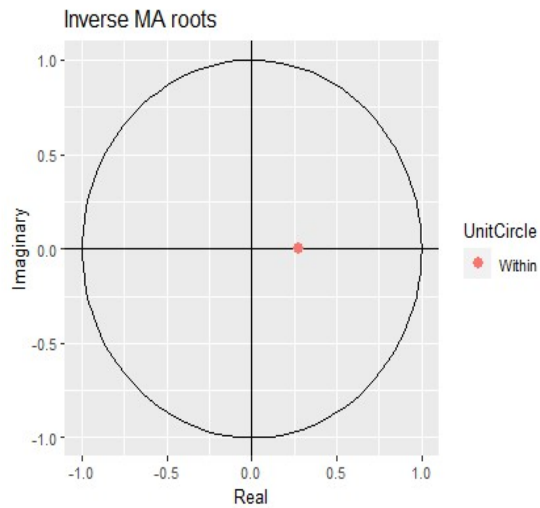


Figure 25- Forecasts of the best ARIMA model for the transformed price of rice / Figure 26- Inverse characteristic roots for ARIMA(0,1,1) fitted to the training set

All red dots lie inside the unit circle. Hence, the fitted model is both stationary and invertible.

Simple Exponential Smoothing vs. Holt's method vs. Holt-Winters' method

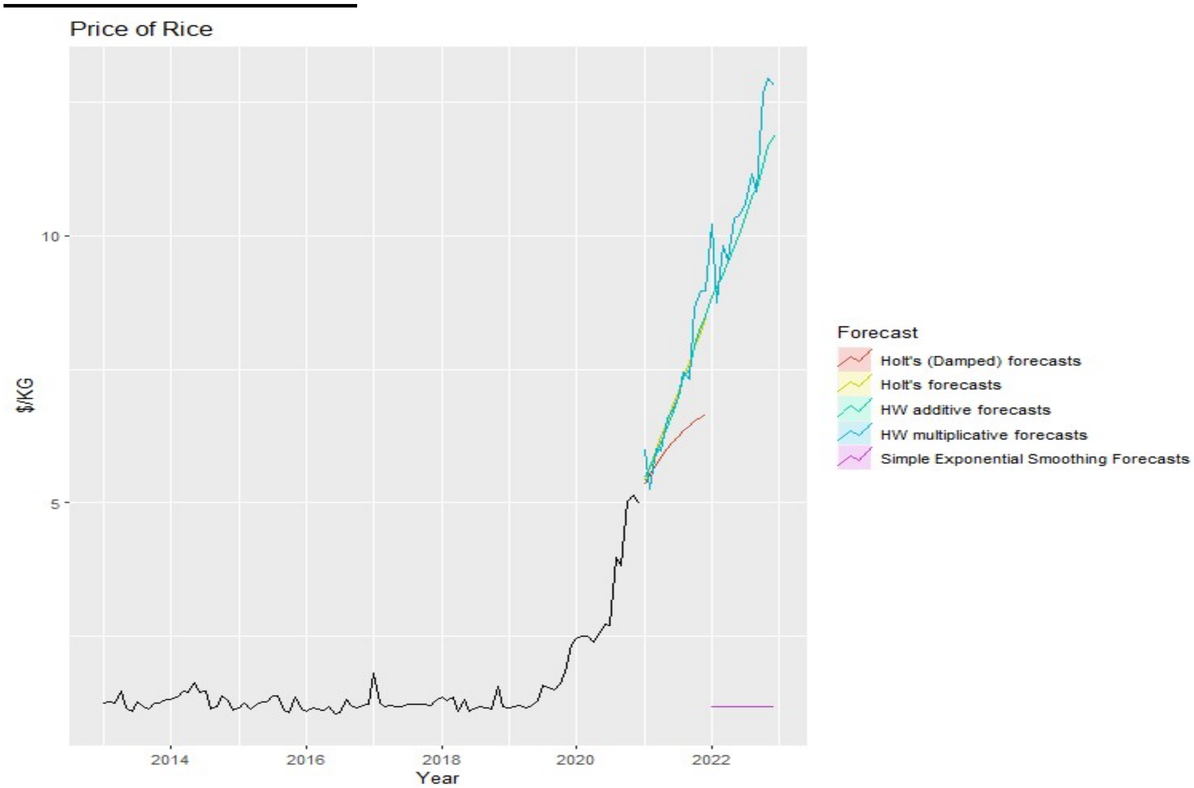


Figure 27- Forecasts of the rice price from simple, trend, and seasonal models

Model	AICc
Simple exponential smoothing	561.3947
Holt's trend	163.6313
Holt's trend(Damped)	166.6481
Holt-Winter's (Additive)	186.3091
Holt-Winter's (Multiplicative)	160.2228

Table 9- Accuracy measures of simple, trend, and seasonal models of the price of rice

Holt's-Winter (Multiplicative) is the best model since it had the lowest AICc (160.2228).

Thus, the best model across all models is the ARIMA (0,1,1) with transformation since it had the lowest AICc (-129.2665) across ALL models.

Meat

Benchmarking Models

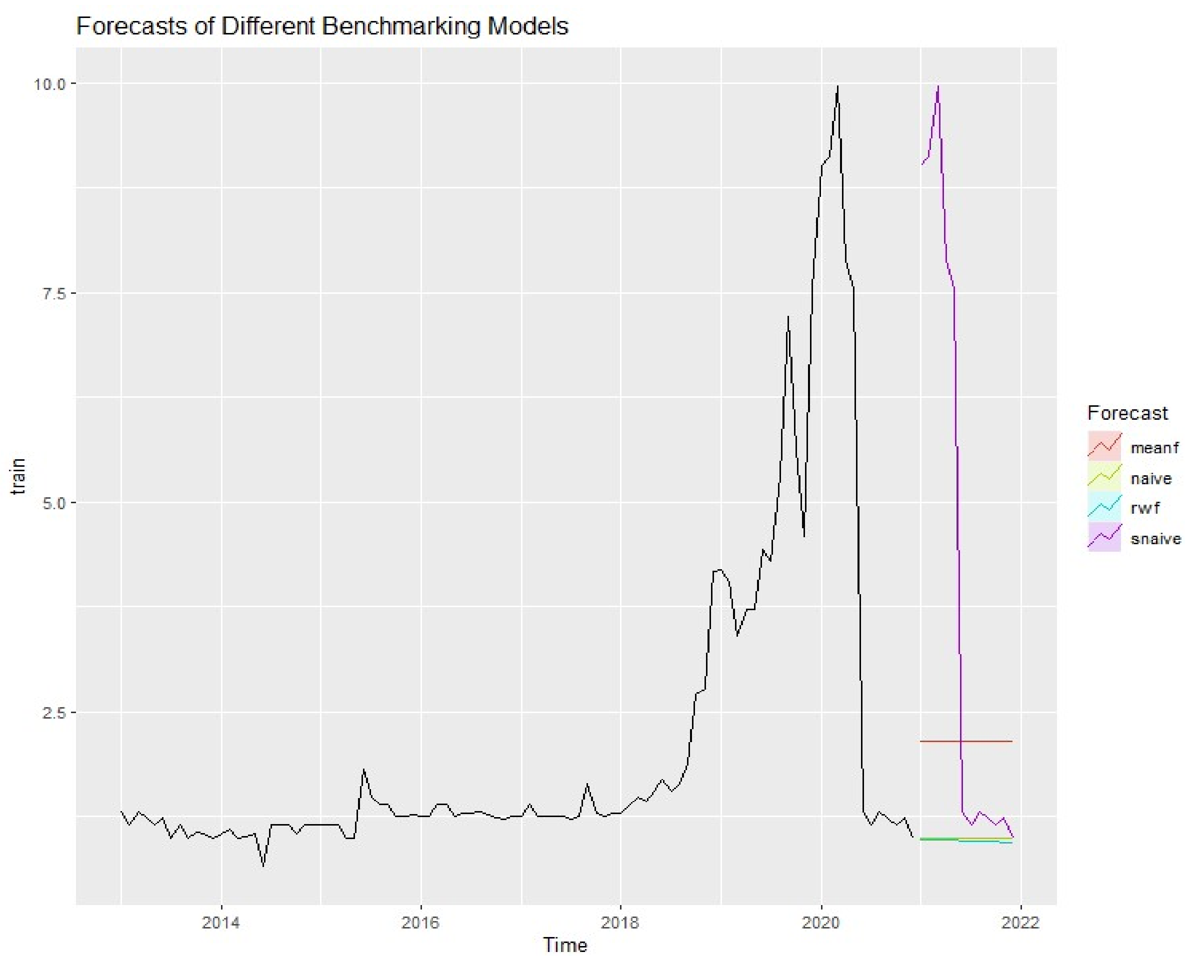


Figure 28- Forecasts of benchmarking models for the price of meat

Benchmarking models	RMSE
Average method	1.136631
Naïve forecast	0.126083
Seasonal Naïve forecast	4.980811
Random walk forecast	0.130691

Table 10- Accuracy measures of benchmarking models for the price of meat

For meat, the best benchmarking model was the Naive method since it yielded the lowest RMSE value (0.126083).

ARIMA

To model the price of meat data using ARIMA, we first differenced our data since the training data is not stationary because the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test yielded a test-statistic value of 1.0312 which is greater than the 1% critical value of 0.739. After differencing the data, it passed the KPSS test since the test statistic (0.0696) is lower than all of the critical values indicating that the data is stationary and non-seasonal.

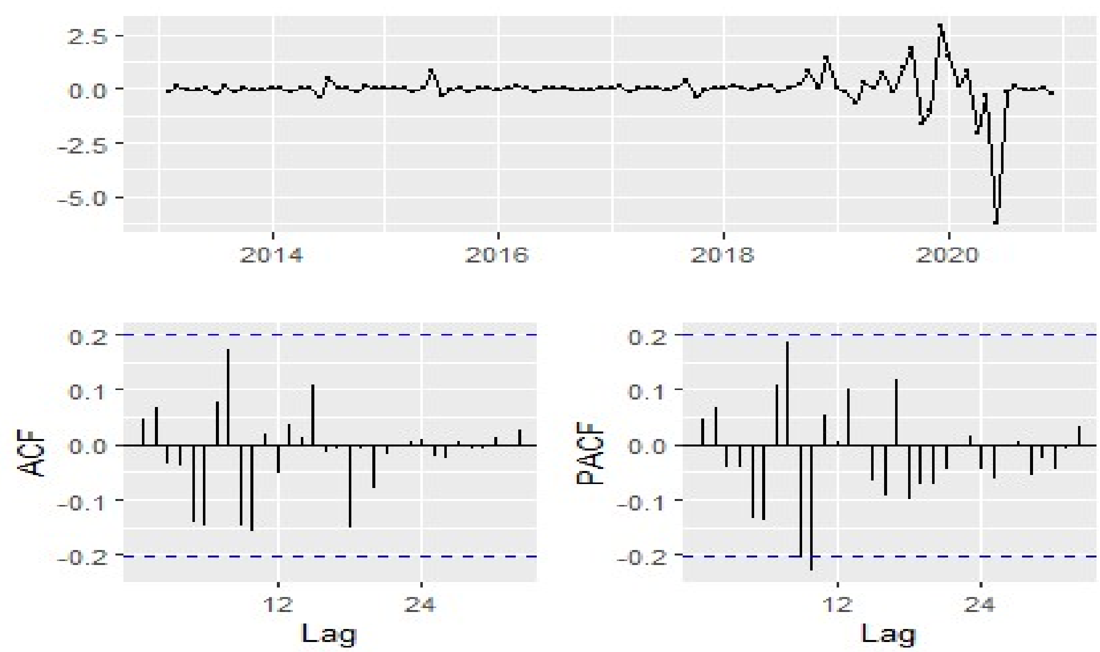


Figure 29- Time plot, ACF, and PACF of meat data

Before considering which ARIMA model should be used, we first must check the ACF and PACF plots to figure out our starting point in the modeling procedure. After plotting the differenced training set (Figure 29), we can see PACF and ACF are sinusoidal. In this case, d=1 is the initial model to start with. After starting with the initial model, we tested a total of 9 models, collected their AICc values, and tabulated them.

Models	AICc
ARIMA (1,1,0)	243.2096
ARIMA (0,1,1)	243.2355
ARIMA (1,1,1)	245.2317
ARIMA (1,1,2)	247.0013
ARIMA (2,1,1)	247.007
ARIMA (0,1,2)	244.8787
ARIMA (2,1,0)	244.931
ARIMA (1,0,1)	246.2441
AUTO.ARIMA (0,1,0)	241.3379

As shown in Table 10, AUTO.ARIMA (0,1,0) is the best model since it had the lowest AICc (241.3379).

Table 11- Accuracy measures of meat ARIMA models

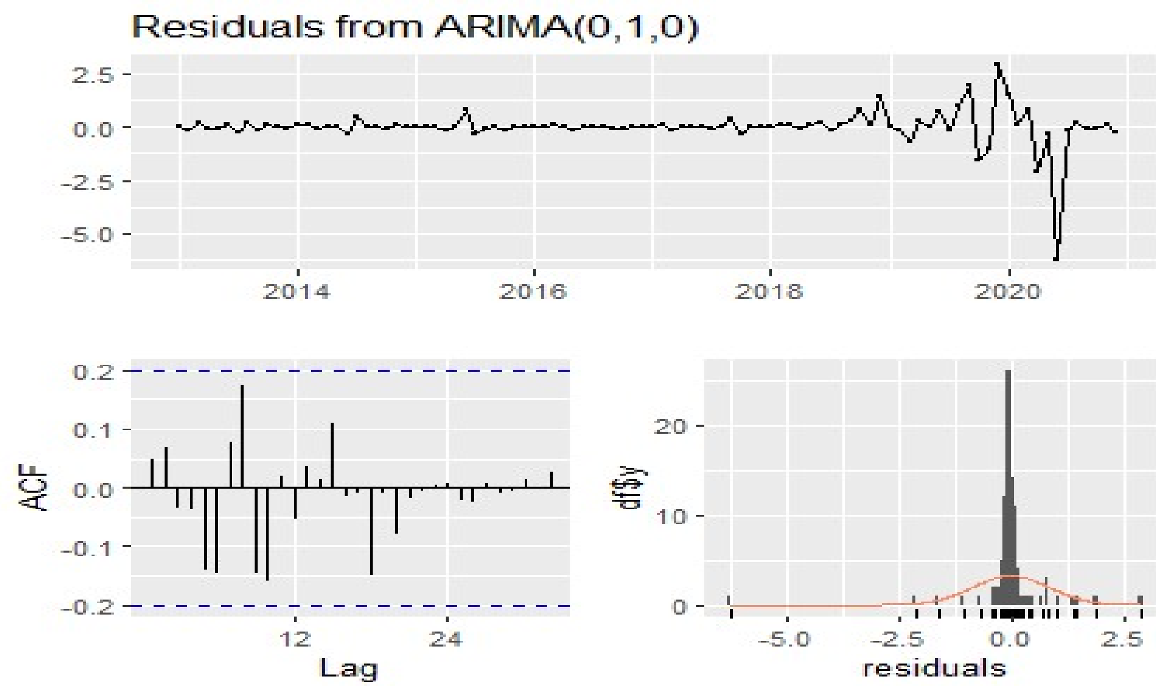


Figure 30- Time plot, ACF, and histogram of the best ARIMA model for the price of meat

Variance is constant (Homoscedasticity) except for a few outliers between 2019 and 2020. The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays no autocorrelations,i.e., all spikes are within the blue-dashed bounds. In addition, this model passed the Ljung- Box test with a p-value of 0.4597 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.

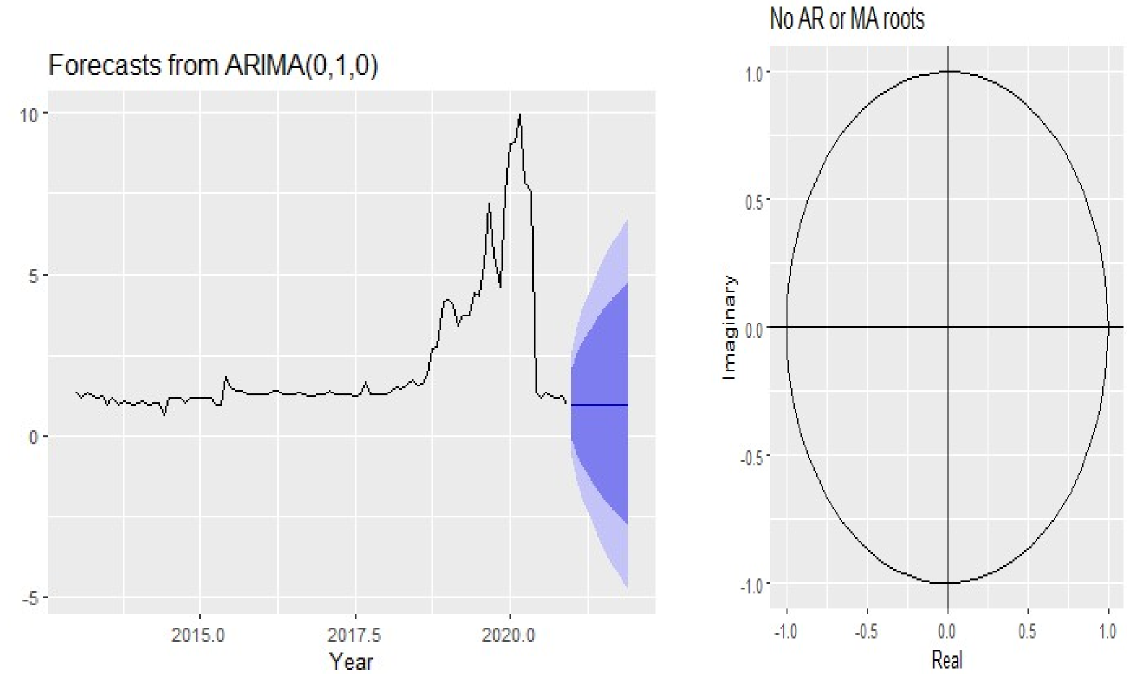


Figure 31- Forecasts of the best ARIMA model for the price of meat

ARIMA (with Box-Cox Transformation)

We first differenced our logarithmically transformed (i.e., lambda = 0) training data since it is not stationary because the KPSS test yielded a test-statistic value of 1.2523 which is greater than the 1% critical value of 0.739. After differencing the transformed data twice (second-order differencing), it passed the KPSS test since the test statistic (0.1165) is lower than the 1% critical value of 0.739 indicating that the transformed data is stationary and non-seasonal.

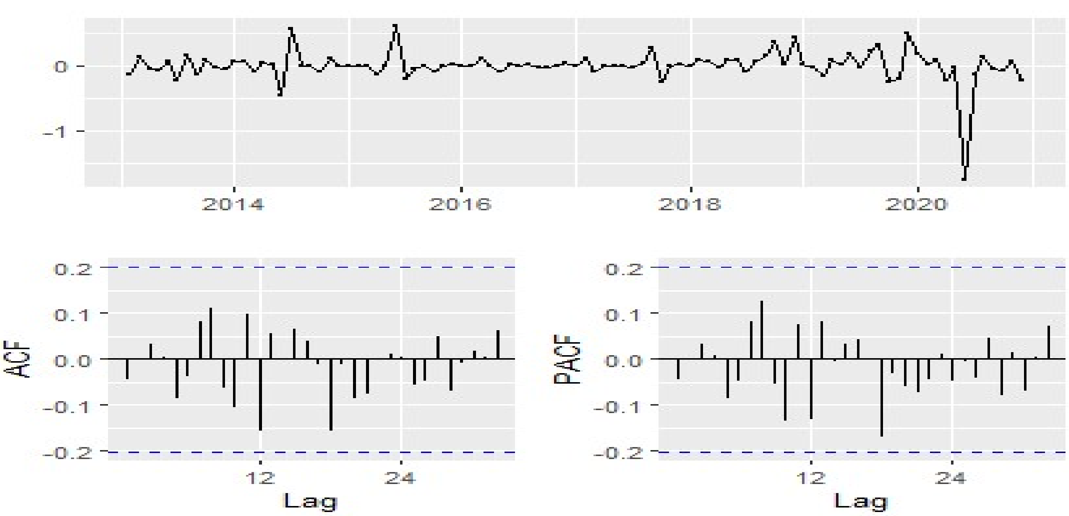


Figure 32- Time plot, ACF, and PACF of transformed meat data

After plotting the transformed training dataset (Figure 32), we can see that the ACF and PACF are sinusoidal. In this case, d=1 is the initial model to start with. After starting with the initial model, we tested a total of 9 models, collected their AICc values, and tabulated them.

Models	AICc
ARIMA (1,1,0)	3.170762
ARIMA (0,1,1)	3.171697
ARIMA (1,1,1)	5.30426
ARIMA (1,1,2)	7.484482
ARIMA (2,1,1)	7.482589
ARIMA (0,1,2)	5.302656
ARIMA (2,1,0)	5.304055
ARIMA (1,0,1)	5.287113
AUTO.ARIMA (0,1,0)(0,0,1)[12]	-1.581

As shown in Table 12, AUTO.ARIMA (0,1,0)(0,0,1)[12] is the best model since it had the lowest AICc (-1.580995).

Table 12- Accuracy measures of transformed meat ARIMA models

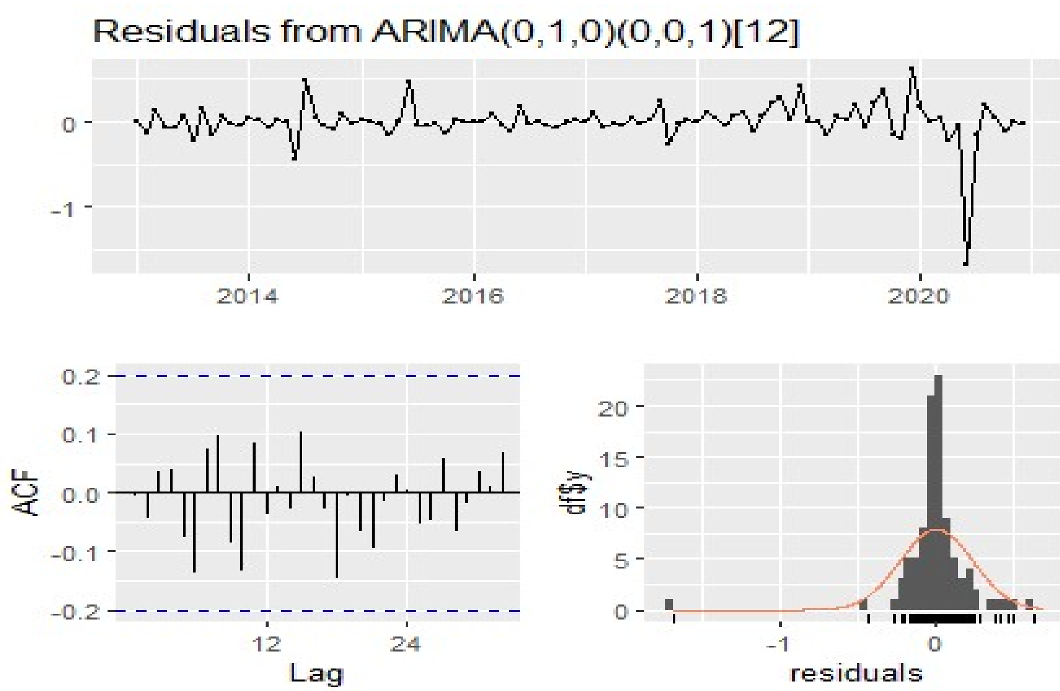


Figure 33- Time plot, ACF, and histogram of the best ARIMA model for the transformed price of meat

Variance is constant (Homoscedasticity) except for a few outliers past 2020. The histogram shows that residuals are normally distributed with a mean = 0, and the ACF plot displays no autocorrelations,i.e., all spikes are within the blue-dashed bounds. In addition, this model passed the Ljung- Box test with a p-value of 0.8189 (> 0.05), which insinuates that the autocorrelation between variables is statistically equal to zero. Thus, the residuals are white noise.

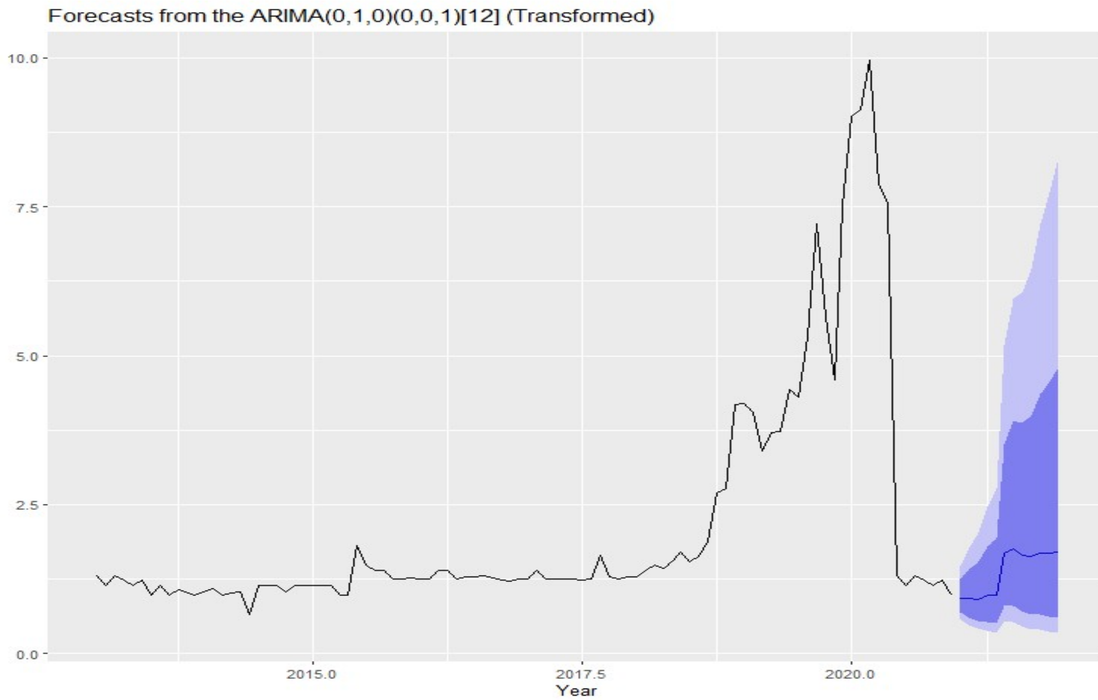


Figure 34- Forecasts of the best ARIMA model for the price of meat

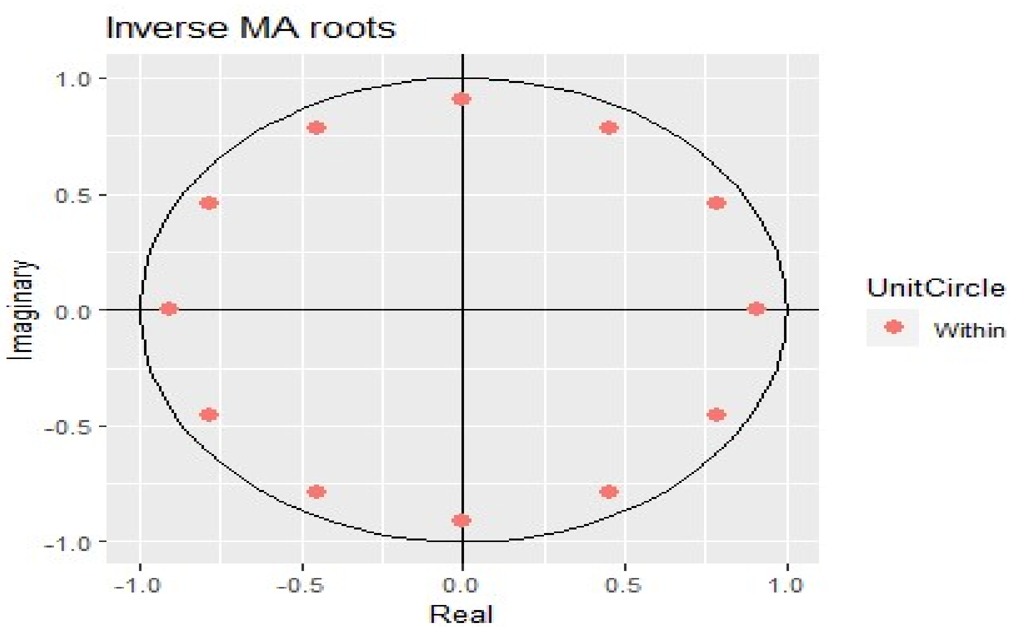


Figure 35- Inverse characteristic roots for ARIMA (0,1,0)(0,0,1)[12] fitted to the training set

All red dots lie inside the unit circle. Hence, the fitted model is both stationary and invertible.

Simple Exponential Smoothing vs. Holt's method vs. Holt-Winters' method

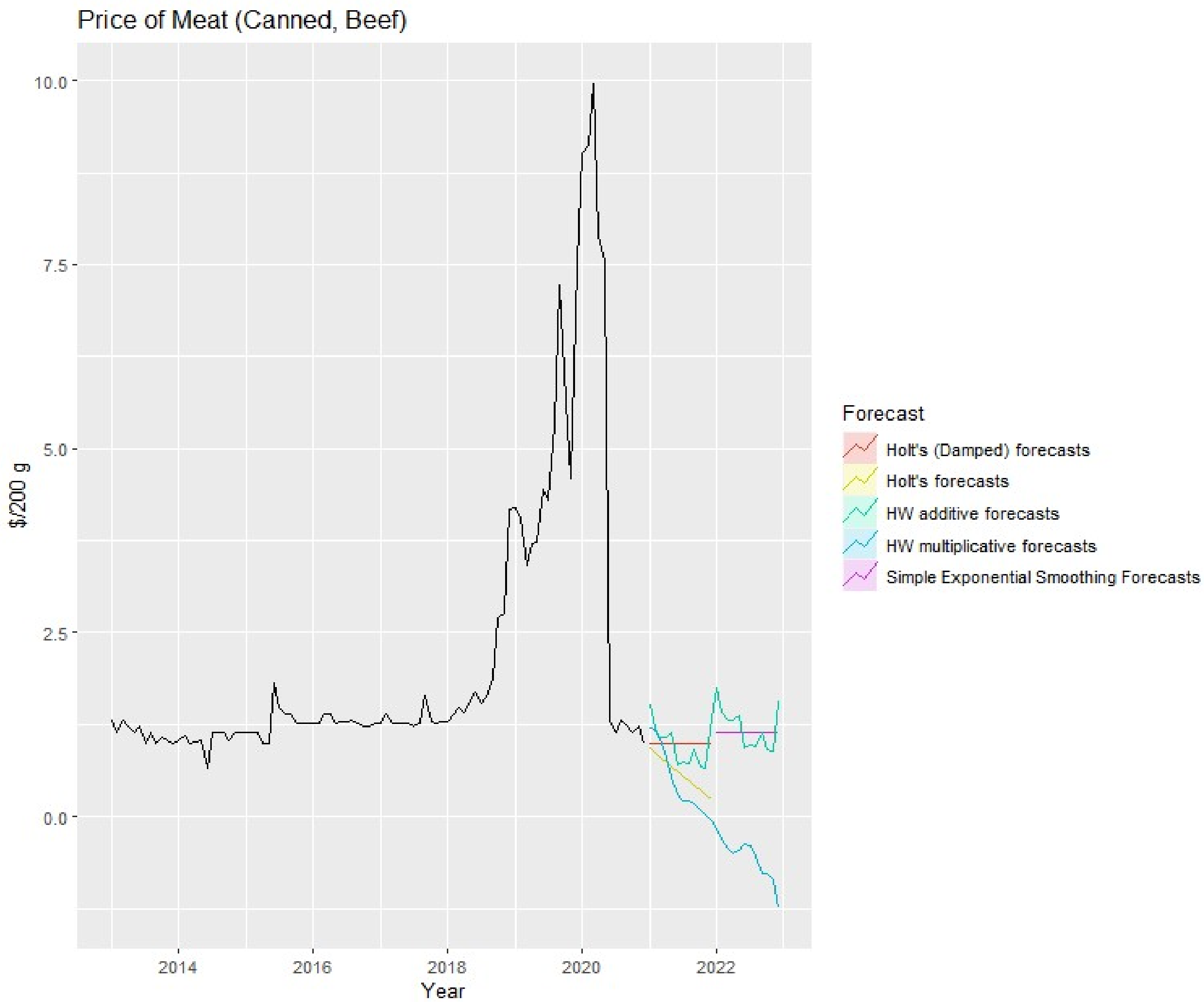


Figure 36- Forecasts of the meat price from simple, trend, and seasonal models

Model	AICc
Simple exponential smoothing	464.3388
Holt's trend	419.3414
Holt's trend (Damped)	417.6462
Holt-Winters' (Additive)	463.9886
Holt-Winters' (Multiplicative)	435.4940

Holt's trend model (Damped) is the best since it had the lowest AICc (417.6462).

Thus, the best model across all models is the seasonal ARIMA (0,1,0)(0,0,1)[12] with transformation since it had the lowest AICc (-1.580995) across ALL models.

Shiny Web Application

To enhance the project even further, a user-friendly web application was generated using the shiny package. The web application's purpose is to simply organize the main findings of the project and enable the user to interact with different tables and graphs. Snapshots of the web application can be found in Appendix A of the report.

Discussion

Based on the obtained results, it was determined that to forecast the price of sugar for the upcoming year, the organization should implement an ARIMA (1,1,2) model with transformation since it performed the best in terms of AICc. ARIMA (0,1,1) with transformation was found to be the best model to forecast the price of rice for the upcoming year. Seasonal ARIMA (0,1,0)(0,0,1)[12] was found to be the best model to forecast the price of meat for the upcoming year. It would seem that transforming the datasets before model implementation yielded more accurate results. UNESCWA and WFP can use this gained insight to render more well-informed decisions when another food crisis hits Lebanon in the future. This would allow the Lebanese government to prepare in advance when another food scarcity tragedy is beset on the economically destitute civilians living in Lebanon, both citizens, and refugees. The main purpose that was expressed in the proposal was to forecast the food prices using historical data about their performance. I suspected that an ARIMA model would end up being the most suitable model to forecast the price of each food type since ARIMA models typically outperform other models if the problem is of the univariate econometric or financial type.

Conclusion

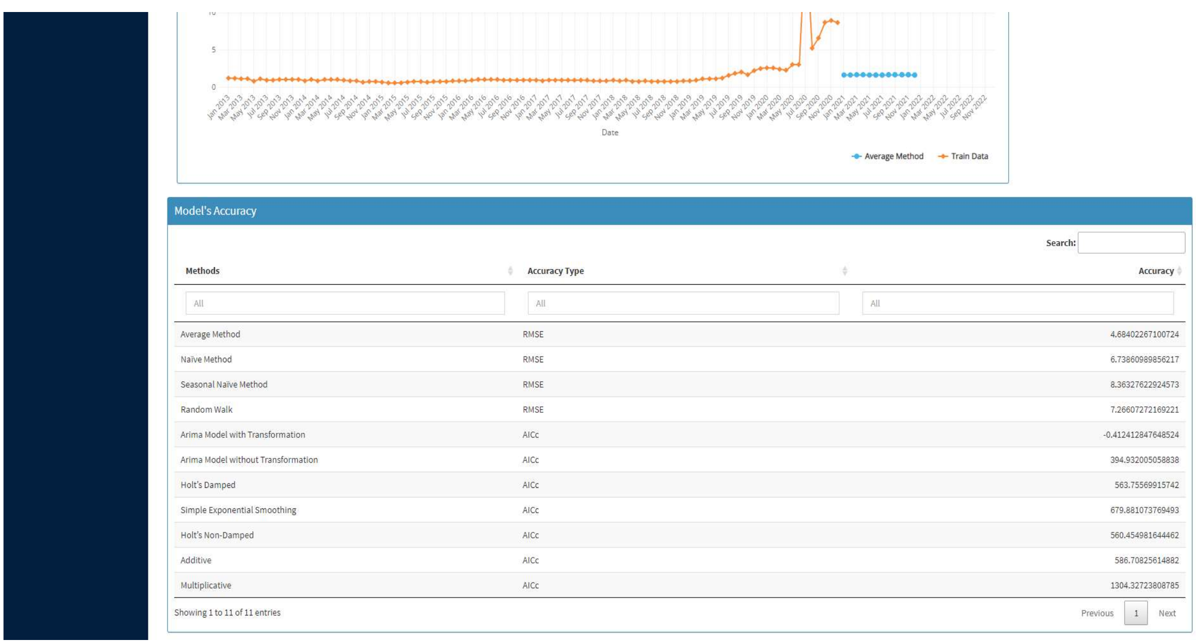
The COVID-19 pandemic, financial crisis, port explosion, and the Russia-Ukraine war gave birth to a food crisis in Lebanon. Food prices have been skyrocketing during the last couple of years while more than half of the population has plunged into poverty leading to the destruction of the middle class in Lebanon. The objectives of the project have been met remarkably well. However, when building and implementing prediction models, there is always room for improvement. Originally, I had hoped to conduct a multiple linear regression analysis to complement the univariate analysis as well as conduct a geospatial analysis of the data, but alas, the dataset was not designed to be worked on by programming languages. Also, other types of models may be implemented to predict the price of certain commodities such as an ARIMA with Fourier terms. We could also take different splits of the data to analyze the true effect of the COVID-19 pandemic on the price of food items in Lebanon. Hopefully, the models built for this project may prove useful in predicting what the food crisis will be like in the coming future since the Lebanese people have endured great pain and loss to the point of despair, and anything that can help them prepare in advance for the next food crisis would be advantageous to have. It is recommended that Lebanon use simple measures such as implementing a strategic grain silo policy that could have enabled the country to build and purchase stocks at low prices to serve as a bulwark against the food crisis. A policy such as this would have enabled the country to capitalize on the supply capacity of the Beirut port's national grain silos.

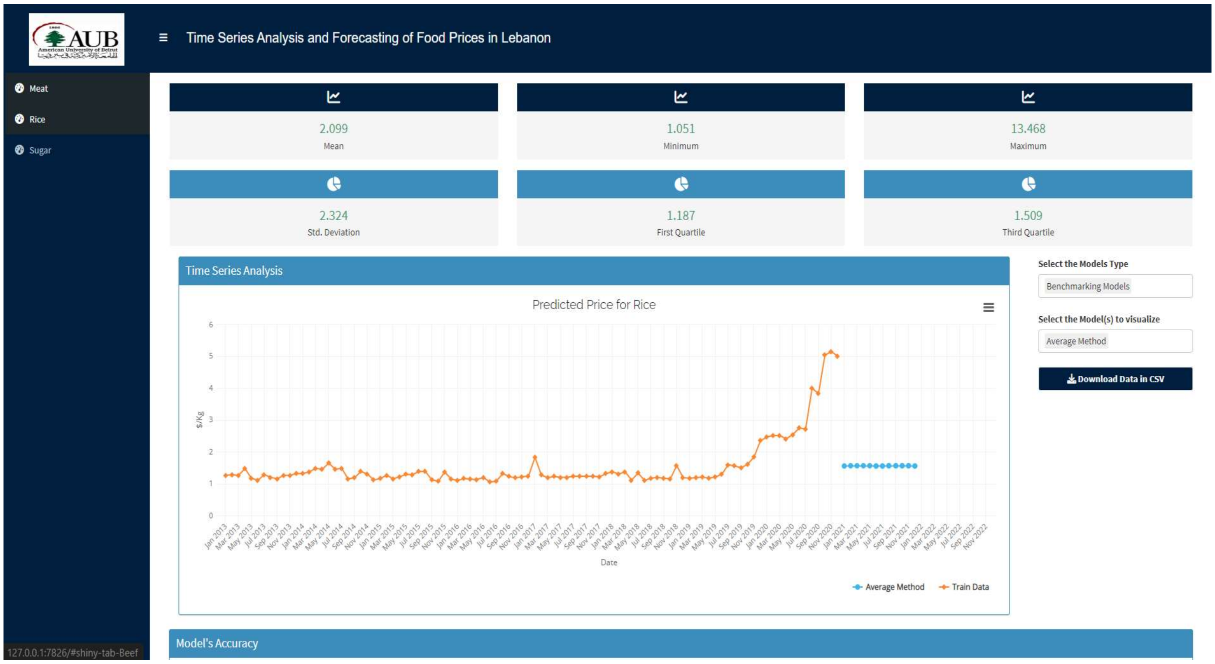
References

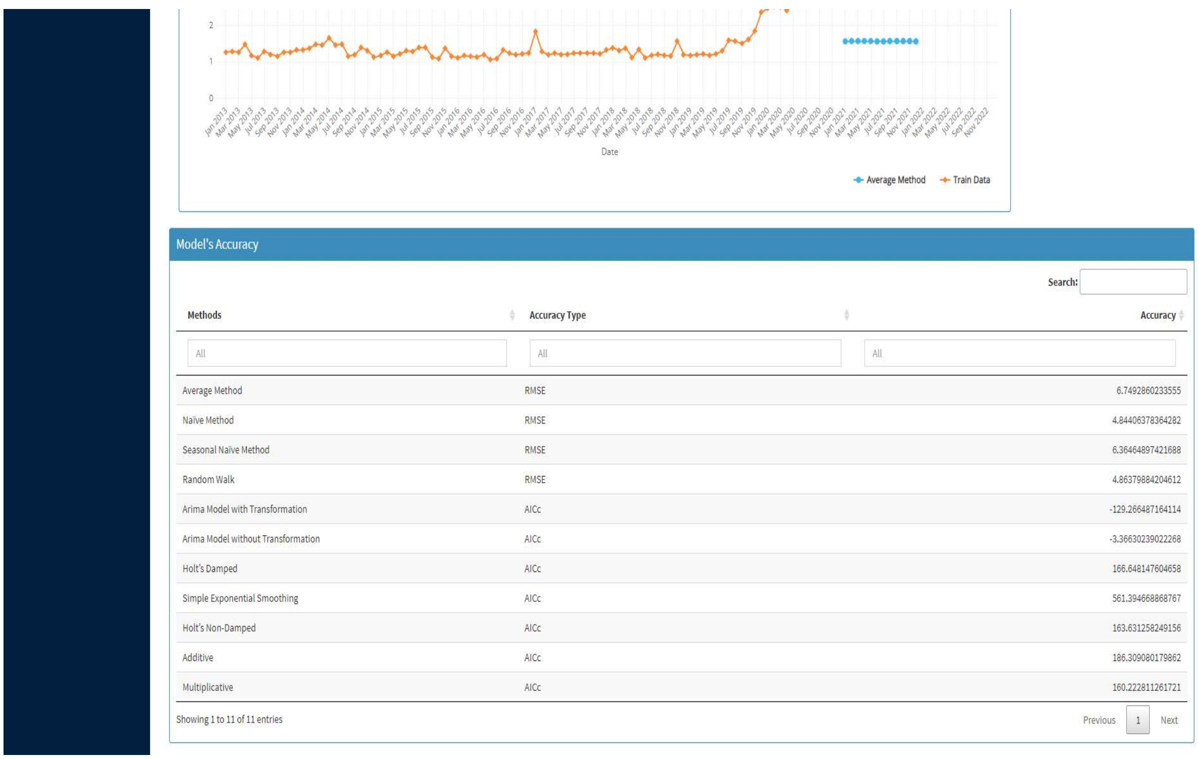
- *Food crisis in Lebanon: Country profile and overview of relief efforts*. World Food Program USA. (2022, August 25). Retrieved August 26, 2022, from https://www.wfpusa.org/countries/lebanon/
- Chehayeb, K. (2021, October 12). *'unprecedented' hunger in Lebanon as fuel crisis hikes food costs*. Hunger News | Al Jazeera. Retrieved August 26, 2022, from https://www.aljazeera.com/news/2021/10/11/lebanon-fuel-crisis-hunger-food-prices
- Hussein, M. (2021, August 4). *Infographic: How much do basic necessities cost in Lebanon?* Infographic News | Al Jazeera. Retrieved August 26, 2022, from https://www.aljazeera.com/news/2021/4/12/infographic-how-much-do-everyday-items-costs-in-lebanon
- Melillo, F. (2022, June 10). *Global Food Crisis, war in Ukraine sinks Lebanon*. INTERSOS. Retrieved August 27, 2022, from https://www.intersos.org/en/global-food-crisis-war-in-ukraine-sinks-lebanon/
- Chehayeb, K. (2020, April 11). Coronavirus: Wheat shortages and rising food prices spell disaster for cash-strapped Lebanon. Middle East Eye. Retrieved August 27, 2022, from https://www.middleeasteye.net/news/coronavirus-lebanon-shortages-rising-food-prices-economic-crisis
- Halabi, S. (2022, April 26). Lebanon's food crisis will get worse. ISPI. Retrieved August 27, 2022, from https://www.ispionline.it/en/pubblicazione/lebanons-food-crisis-will-get-worse-34732
- Ohyver, M., & Pudjihastuti, H. (2018, January). *Arima model for forecasting the price of medium quality rice to ...* Retrieved August 28, 2022, from https://www.researchgate.net/publication/327295318_Arima_Model_for_Forecasting_the_Price_of_Medium_Quality_Rice_to_Anticipate_Price_Fluctuations
- Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021, September 15). Comparing prophet and deep learning to Arima in forecasting wholesale food prices. MDPI. Retrieved August 28, 2022, from https://www.mdpi.com/2571-9394/3/3/40/htm
- Mao, L., Huang, Y., Zhang, X., Li, S., & Huang, X. (2022, July 28). Arima model forecasting analysis of the prices of multiple vegetables under the impact of the COVID-19. PLOS ONE. Retrieved August 29, 2022, from https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0271594
- Gaddi, G. (2017, January 1). *Application of Arima model for forecasting agricultural prices*. Journal of Agricultural Science and Technology. Retrieved August 29, 2022, from https://www.academia.edu/68697295/Application_of_ARIMA_Model_for_Forecasting_Agricultural_Prices?show_translation=true

Appendices

Appendix A: Shiny Web Application

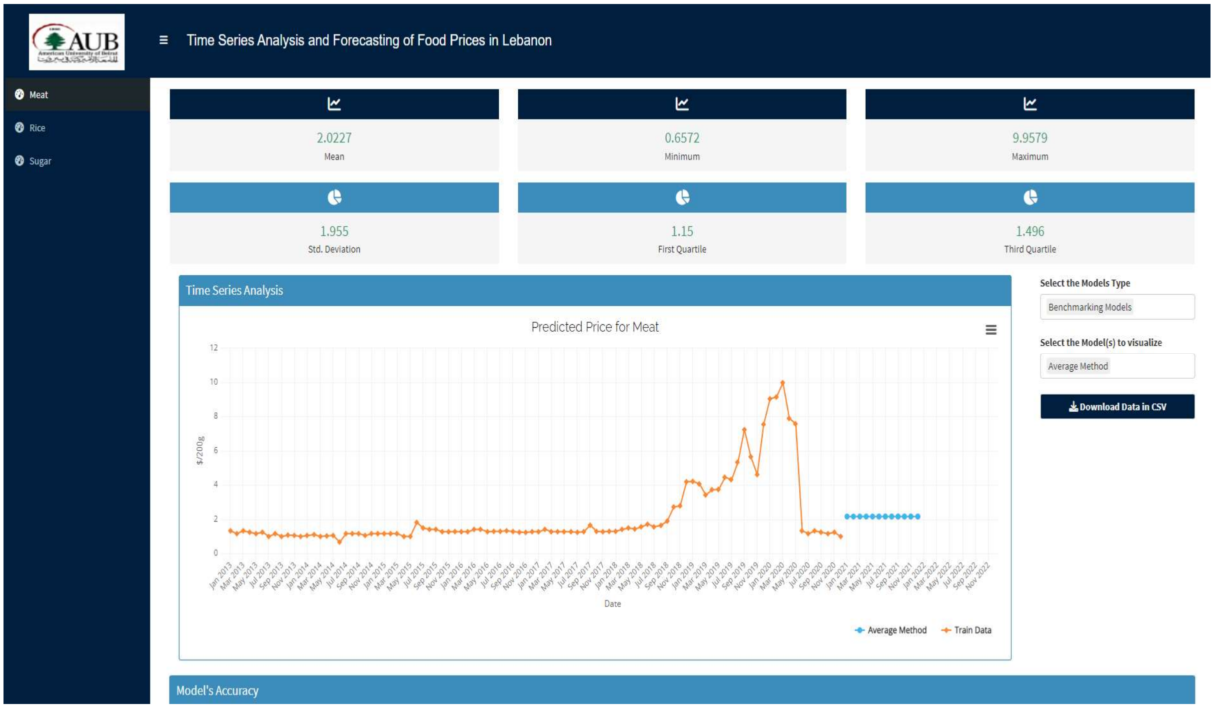


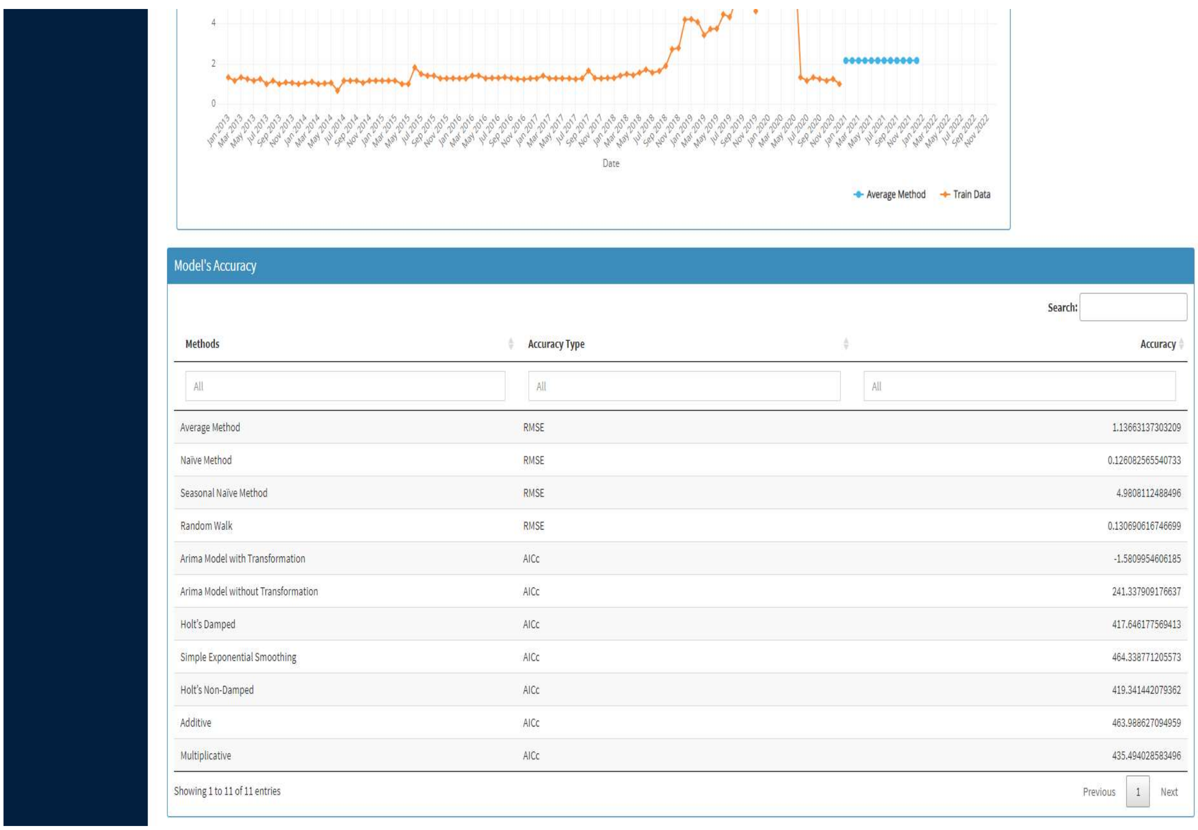




Methods	Accuracy Type	Accuracy
Average Method	RMSE	6.7492860233555
Naïve Method	RMSE	4.84406378364282
Seasonal Naïve Method	RMSE	6.36464897421688
Random Walk	RMSE	4.86379884204612
Arima Model with Transformation	AICc	-129.266487164114
Arima Model without Transformation	AICc	-3.36630239022268
Holt's Damped	AICc	166.648147604658
Simple Exponential Smoothing	AICc	561.394668868767
Holt's Non-Damped	AICc	163.631258249156
Additive	AICc	186.309080179862
Multiplicative	AICc	160.222811261721

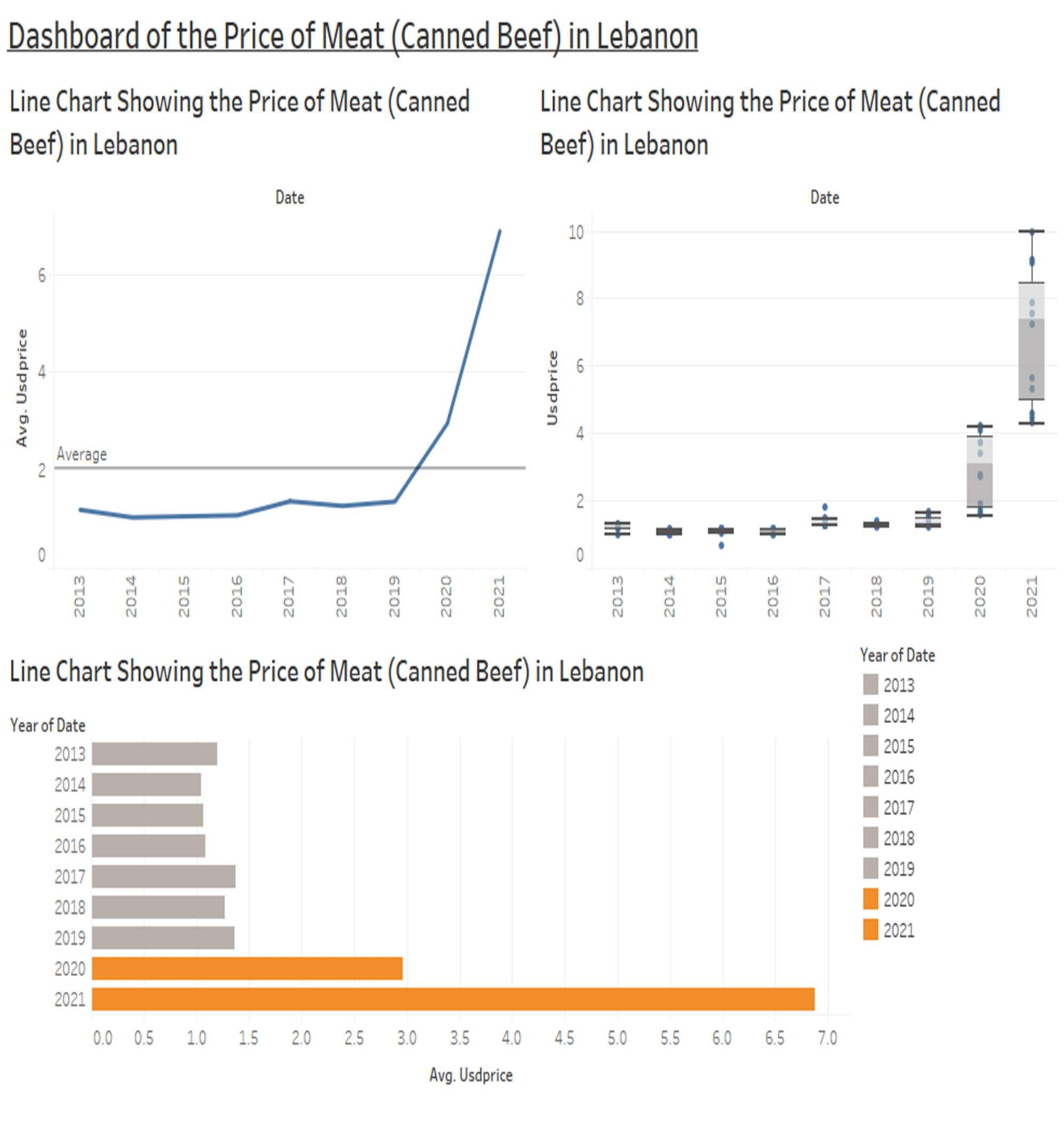
Showing 1 to 11 of 11 entries





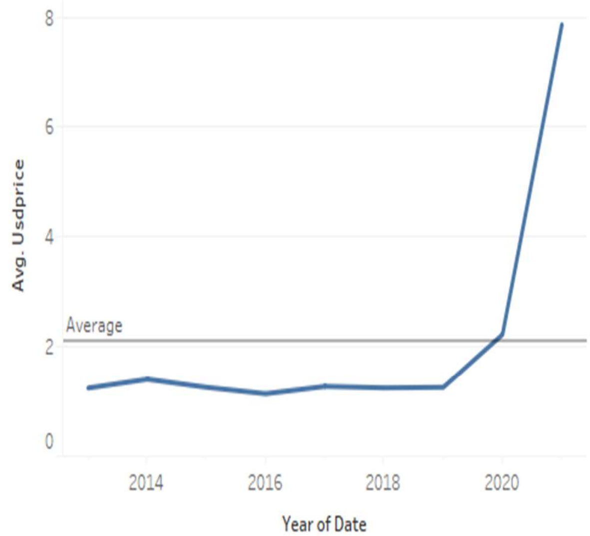
Methods	Accuracy Type	Accuracy
Average Method	RMSE	1.13663137303209
Naive Method	RMSE	0.126082565540733
Seasonal Naive Method	RMSE	4.9808112488496
Random Walk	RMSE	0.130690616746699
Arima Model with Transformation	AICc	-1.5809954606185
Arima Model without Transformation	AICc	241.337909176637
Holt's Damped	AICc	417.646177569413
Simple Exponential Smoothing	AICc	464.338771205573
Holt's Non-Damped	AICc	419.341442079362
Additive	AICc	463.988627094959
Multiplicative	AICc	435.494028583496

Appendix B: Tableau Dashboards

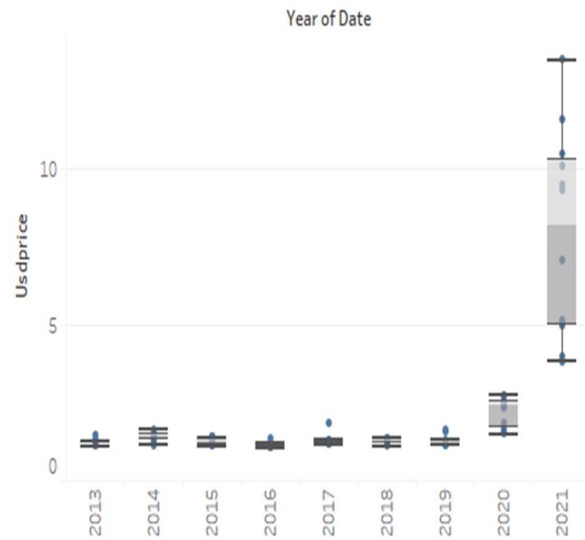


Dashboard of the Price of Rice in Lebanon

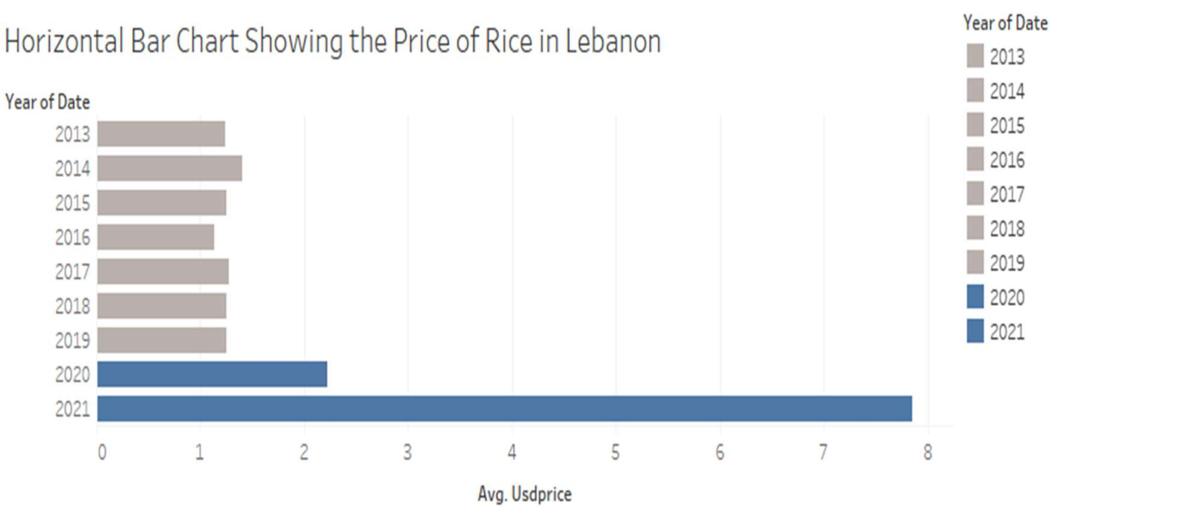
Line Chart Showing the Price of Rice in Lebanon

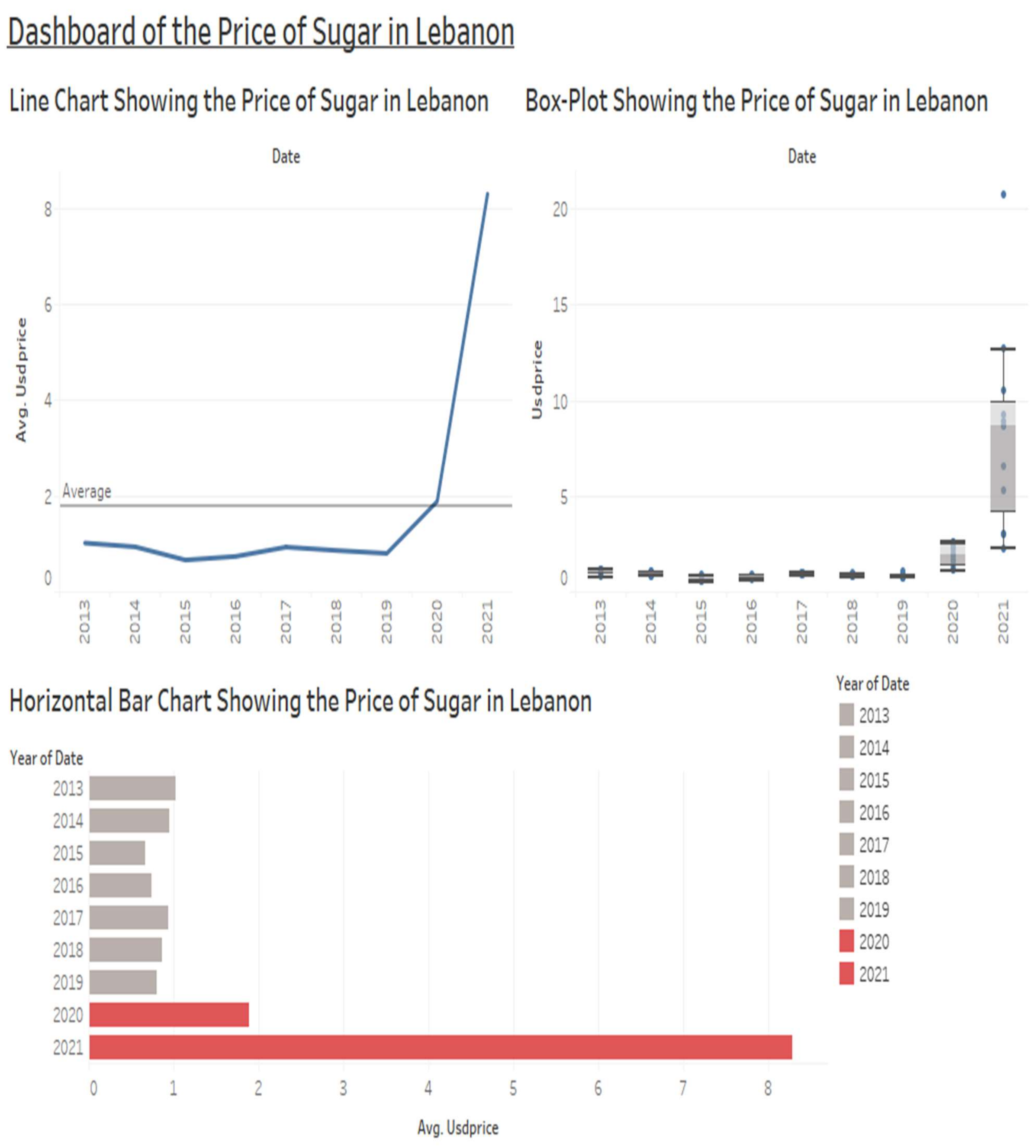


Box-Plot Showing the Price of Rice in Lebanon



Horizontal Bar Chart Showing the Price of Rice in Lebanon




Dashboard of the Price of Sugar in Lebanon
Line Chart Showing the Price of Sugar in Lebanon
Date
Avg. Usdprice
Average
0
2
4
6
8
2013
2014
2015
2016
2017
2018
2019
2020
2021
Box-Plot Showing the Price of Sugar in Lebanon
Date
Usdprice
0
5
10
15
20
2013
2014
2015
2016
2017
2018
2019
2020
2021
Horizontal Bar Chart Showing the Price of Sugar in Lebanon
Year of Date
2013
2014
2015
2016
2017
2018
2019
2020
2021
0
1
2
3
4
5
6
7
8
Avg. Usdprice
Year of Date
2013
2014
2015
2016
2017
2018
2019
2020
2021