2023

Application of Hybrid seasonal ARIMA-GARCH Model in modelling and forecasting fertilizer prices in Kenya.

Okello, Elizabeth Akinyi
Strathmore Institute of Mathematical Sciences
Strathmore University

Application of Hybrid Seasonal ARIMA-GARCH Model in Modelling and Forecasting Fertilizer Prices in Kenya

Okello, Elizabeth Akinyi

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Statistical Science of Strathmore University

Strathmore Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

July 2023

Declaration

I declare that this work has not been previously submitted and approved for award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Name: .... **Okello Elizabeth Akinyi** ....

Signature: ....

Date: .... June 18, 2023 ....

Approval

The thesis of Okello Elizabeth Akinyi was reviewed and approved by the following:

Dr. Evans Otieno Omondi
Supervisor,
Institute of Mathematical Sciences, Strathmore University.

Dr. Titus Okello Orwa
Supervisor,
Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu
Dean,
Institute of Mathematical Sciences, Strathmore University.

Dr. Bernard Shibwabo
Director,
Office of Graduate Studies, Strathmore University.

Abstract

Volatility in fertilizer prices pose a huge risk to both farmers and suppliers. To manage fertilizer price volatility, a more efficient price risk management model is necessary. Stand alone models have been criticized for failing to capture the true market conditions by capturing only the unilateral information. Better outcomes have been credited to combined models, such time series models. Existing models have factored in variables such as natural gas, transport, volumes traded, crude oil prices, corn prices, ethanol, market concentration and regions. In this study, the port through which fertilizer is imported is taken into account while creating a Hybrid SARIMA-GARCH model, which is then used to anticipate pricing. Using RMSE, MAE, and MASE, the model's predictive abilities were assessed. The findings of this study suggest that the best model for the port of Gulf is SARIMA models (1, 1, 0) (2, 1, 0)12, with an AIC = 997.53, and RMSE = 5.6015, and can efficiently capture the pricing behaviour in this port. In Yuzhny, Hybrid SARIMA (2, 1, 0) (2, 1, 0)12–GARCH (1, 1) turned out to be the best fit with AIC = 7.4389, RMSE = 7.5802, MAE=5.4797 and MASE=0.6885. The study concludes that the port through which fertilizer is imported has an effect on the price placed as each of the ports under study yielded a unique model.

KEY WORDS: Nonlinear time series, Heteroscedasticity, SARIMA model, GARCH model, Hybrid SARIMA GARCH model, Ljung–Box test, Augmented Dickey Fuller test.

Table of contents

List of figures

List of tables

List of abbreviations

AME	Absolute Maximum Error	ARX	Auto regressive with exogenous variables
ACF	Autocorrelation Function	ARCH	Auto Regressive Conditional Heteroskedasticity
IID	Independent and identically distributed	LB	Ljung Box
MAE	Mean Absolute Error	MAPE	Mean absolute percentage error
MLL test	McLeod–Li Test	OLS	Ordinary Least Squares
PACF	Partial Autocorrelation Coefficient	RMSE	Root Mean Square Error
RAE	Relative Absolute Error	MRE	Mean Relative Error
SARIMA	Seasonal auto regressive integrated moving average	VAR	Vector Auto Regressive
SARIMAX	Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors	ANN	Artificial Neural Networks
Bi-LSTM	Bi-directional long short term memory	GARCH	Generalized Auto Regressive Conditional Heteroskedasticity

Acknowledgement

First and above all, I am grateful to the Almighty God for His provision, grace, and for granting me good health throughout the study period.

I am grateful to my husband, Dr. Duncan Ochieng Ouma, and our children for their unwavering support and cheering me on throughout the study period.

I am grateful to my parents, brothers and sisters for their immeasurable support throughout the study period.

I am indebted to my supervisors, Dr. Evans Otieno Omondi and Dr. Titus Okello Orwa for their availability, kind guidance, and support without which this thesis would not have been a success.

I am grateful to my classmates for their support and encouragement.

Dedication

This thesis is dedicated to God Almighty for giving me wisdom and good health.

Chapter 1

Introduction

1.1 Background to the Study

A fertilizer is combinations of nutrients that enables plants to grow (Voogt and Bar-Yosef, 2019). The essential elements of a fertilizer are nitrogen (N), phosphorus (P) and potassium (K), which provide crops with the nutrients required to achieve optimal production. Urea fertilizer is the major source of nitrogen, and is produced through the conversion atmospheric nitrogen using natural gas (Chena et al., 2012). Low adoption of improved land management practices is one of the main factors responsible for lagging agricultural productivity in many low-income countries (Hernandez and Torero, 2013). Although an increase in fertilizer use is not the only solution to this problem, countries that have increased their agricultural productivity have also considerably increased their use of fertilizer (Morris, 2007). Results by Morris (2007) also established that sub-Saharan African countries, which are generally characterized by low agricultural productivity, have far too low fertilizer application rates, averaging 10 kg/ha of nutrients of arable land, compared to 86 in South Asia, 118 in Latin America, 198 in an average middle-income country, and 288 kg/ha in a high income country. In terms of expenses, Newton (2019) found that fertilizer use constitutes about 20 percent of production cost to the farmer and for those dealing in the fertilizer imports their biggest buckets of spending are fertilizer purchase.

From the year 2002, prices of fertilizers in international markets have increased sharply for a variety of reasons such as normal cyclical patterns of investment in production capacity and higher shipping costs. The recent increases can be attributed partly to rising cost of oil for transport and natural gas for manufacturing urea.

In addition to this, lack of competition among suppliers and distributors within the country or region, poor dealer networks resulting in late or irregular delivery, high transportation costs due to lack of adequate infrastructure, uncertain policy environments and weak regulatory systems, a limited of market information, and limited access to finance among other factors have resulted into structural changes in the fertilizer market which have created a potential for increased volatility (Morris, 2007). This poses a serious threat to livelihoods in Sub Saharan Africa, which relies heavily on imported fertilizer which in turn slows down their uptake (Kelly et al., 2019).

Kim and Brorsen (2017) found that suppliers aggregate demand from the farmers across countries and place large orders on the global market mainly for fertilizers such as Urea DAP, CAN and NPK. Due to limited production capacity of the fertilizer industry and supply adjustments linled to global factors, there is a need to formulate models that can be used to analyse and predict fertilizer prices more efficiently to reduce the high risk on the side of supplier and proper planning and budgeting for the farmer. A variety of models to predict future prices have been formulated using historical data both on fertilizer prices and factors affecting fertilizer prices. These models include ARX, SARIMAX, ARIMA, combined SARIMA and Regression models, Bi-LSTM, Artificial Neural Networks (ANN) and time series and Dynamic models like the ARCH and GARCH models.

This study seeks to add to the existing body of knowledge by formulating a Hybrid SARIMA GARCH model that can be used in analysing fertilizer prices. The formulated model is applied to provide accurate and reliable fertilizer price predictions using historical data.

1.2 Statement of the Problem

Fertilizers play a significant role in agriculture-dependent Kenya. They are a major contributor to high agricultural fields for local and global markets. Costs associated with fertilizer purchase constitutes about 20 percent costs of inputs to the farmers. Volatility in fertilizer prices pose a huge risk to both farmers and suppliers. To manage fertilizer price volatility, a more efficient price risk management model is necessary. Future contracts have been tried on fertilizer, but have been unsuccessful in managing risks. The fertilizer cash market is also not active enough to support a future contracts as indicated by Kim and Brorsen (2017). Furthermore, fertilizer are only produced by a few private companies coupled with limited market information. It is therefore important to develop fertilizer management decision tools for farmers, extension agents and input distributors. Such models provide helpful information aimed at minimizing risks associated with fertilizer use.

Price forecasting models such as ARX,SARIMAX,ARIMA,combined SARIMA and Regression models,Bi-LSTM, Artificial Neural Networks (ANN) and time series and Dynamic models like the ARCH and GARCH models have been proposed and are considered as important tools in the analysis of time series data, specifically in forecasting volatility and have generally performed well. The stand alone models, however, have been criticized for their limitation in providing accurate price predictions as some failed to capture the true market scenarios. The combined models have always out performed the stand alone models (Xu et al., 2021).

While a lot of work has been done on fertilizer pricing, no attempts have been made in formulating a Hybrid SARIMA GARCH model. The current study proposes to formulate a Hybrid SARIMA GARCH model to analyze fertilizer prices and be applied to accurately and efficiently forecast retail fertilizer prices in Kenya.

1.3 Research Objectives

1.3.1 General Objective

The main objective of this study is to formulate, analyze and apply a Hybrid SARIMA GARCH model that can be used to forecast retail fertilizer prices in Kenya.

1.3.2 Specific Objectives

1. To formulate and analyze a Hybrid SARIMA-GARCH model and apply it to retail fertilizer price data.
2. To apply the hybrid model to forecast fertilizer prices and identify the best times for purchase.

1.4 Justification of the Study

Various factors contribute to volatility of fertilizer prices. These include prices of raw materials, increased dependency of global trade, shift in the global supply chain, change in demand and supply, ocean vessel transportation rates, foreign policies and exchange rates. The many factors coupled with longer supply chain also means that fertilizer retailers must project demand and purchase inventories six to nine months prior to the time that producers actually purchase and use the products. It is important therefore to have a model that reliably forecasts fertilizer prices.

Also, Investors are always willing to maximize their earnings and mitigate risk distribution from abnormal events through investment in the segmented market, but for the integrated market, this would be fruitless. However, some researchers argued that in emerging (even in developed) markets, when the competition rises then the efficiency of local markets also enhances which in return reduced the cost of capital and price instability and ultimately resulted in monetary development (Morris, 2007).

The result of this study is expected to guide the global sourcing teams of organizations dealing in importing fertilizer and those willing to invest to have a rough idea of the expected sourcing prices at a given time to enable them place competitive prices to the local market which will help them to stay afloat in their businesses and maximize on their profit margins. To the farmer this helps them to plan and budget appropriately to ensure an increased productivity.

1.5 Significance of the Study

This study is expected to guide the global sourcing teams of organizations dealing in importing fertilizer and potential investors to have good knowledge and understanding of the expected sourcing prices at a particular time to enable them place competitive prices to the local market which will help them stay afloat in their businesses and maximize on their profit margins. To the farmer this helps them to plan and budget appropriately to ensure an increased productivity.

1.6 Expected Output

In this study, we expect to achieve the following results:

1. A Hybrid SARIMA-GARCH model and apply it in predicting retail fertilizer price data.
2. Identify the best time to purchase fertilizer using the suggested Hybrid SARIMA GARCH model

1.7 Communication and Utilization of Results

The results will be disseminated through the Strathmore University website and Strathmore University library, authorized by the administrator of SU+@Strathmore university. Since there is limited market information on fertilizer prices, It's important to develop fertilizer management decision tools for farmers, extension agents and input distributors. Such models provide helpful information aimed at minimizing risks associated with fertilizer use.

This study is expected to guide the global sourcing teams of organizations dealing in importing fertilizer and potential investors to have good knowledge and understanding of the expected sourcing prices at a particular time to enable them place competitive prices to the local market which will help them stay afloat in their businesses and maximize on their profit margins. To the farmer this helps them to plan and budget appropriately to ensure an increased productivity.

Chapter 2

Literature Review

2.1 Introduction

This literature gives an overview of the models developed so far, the factors affecting fertilizer prices and their relationships, and compares their abilities to model and in turn give reliable and accurate forecast on fertilizer price time series data. A keen interest in studying fertilizer price volatility and forecasting began in the early 1990s. Fertilizer prices across the globe during the mid-1970s and throughout the 1980s, were highly volatile (Mtaita, 2003). Various models have been formulated and applied in forecasting studies such as Time Series models, machine learning models e.g Artificial Neural Networks and Hybrid Models.

2.2 Models

2.2.1 ARIMA Models

A study by Vroomen (1991), in the U.S created a model using a combination of time-series (ARIMA) and regression analysis. The study used time series analysis to generate forecast values for the wholesale prices of anhydrous ammonia, phosphoric acid, and potassium chloride. The forecasts were then incorporated into regression equations to forecast the retail prices of 14 major fertilizer mixtures and materials. The retail price forecasts were combined to generate a forecast of the index of fertilizer prices paid by farmers. Results showed that this method performed with reasonable accuracy for short-term forecasting purposes. Such models were found to frequently outperform behavioral models in short run forecasting and this class of models can also be easily updated therefore permitting the forecast user to benefit from the latest information available. A shortfall of this kind of model however, is its assumption of stationarity and that model parameters and error term is constant which is hardly the case in time series data.

Mishra et al. (2014) studied three major fertilizers in India: Nitrogen (N), Phosphorus (P) and Potassium (K) that play a key role in different stages of crop growth. The study used ARIMA method and found that ARIMA (1, 1, 1) model was best suited for estimation of nitrogen production data. They used fertilizer data from 1961 to 2002. The model forecast values obtained projected that production will increases to some extent in future.

Kim and Brorsen (2017) in their study, constructed various models to forecast the free-on-board (FOB) bulk price of granular urea traded in the U.S. New Orleans spot market. A variety of time-series methods and rolling window sizes were considered. They included ARG, ARX, ARMAX, SARMA, rolling window regression. However, inefficiency was observed in the ARG model with a window size of 36 and ARXG model with a window size 48, 60, 72, and 81 at the 5 percent significance level since forecast errors were autocorrelated. Based on the comparison among the constructed models using forecasting accuracy measures, the SARX for MAE and the ARX for RMSE with a window size of 48 months showed the lowest values indicating the window size of 48 months was the most appropriate model for reflecting urea market information.

Forecasts from the models with exogenous variables yielded more accurate forecasts than models without exogenous variables. All models showed MAE and RMSE values less than that of the no-change naïve model an indication that they were more superior to a no-change naive model. Upon comparing the constructed models and that of the Fertilizer Week company, even though the forecast accuracy measures indicate Fertilizer Week company output outperformed ARX and SARX with window size 48, the two MDM tests showed no statistical difference between Fertilizer Week company output and the two models. The ARX was chosen as the candidate for constructing a combination model with Fertilizer Week company results based on the forecast accuracy measures. Neither of Fertilizer Week results and ARX forecast encompassed the other and each contained unique information.The results implied that a combination model using both Fertilizer Week and the ARX rather than either alone would further give accurate forecasts. When the combination model was compared to Fertilizer Week, the combination model showed lower forecast accuracy measures and the MDM test based on squared error loss supported the fact that the composite model is statistically different and it showed improved forecasting performance at the 10 percent level. The combination forecast model was capable of explaining about half of the variability of percentage changes.

Galbraith (2010) used an error correction model to determine how anhydrous ammonia (AA) and urea prices at different locations in North America adjusted to changes in crop and input prices during two time periods 2002 to 2005 and 2006 to 2009. The study also considered measures of supply and demand shocks such as natural gas price futures as a measure of supply shock and corn futures as a measure of demand shock. The empirical results suggest that natural gas future prices had relatively more of an impact on nitrogen fertilizer prices during the 2002 to 2005 period. In addition, nitrogen prices tended to adjust more rapidly to increases in natural gas futures prices than decreases. In the 2006 to 2009 period corn futures prices had more of an impact on nitrogen prices. Nitrogen prices tended to respond immediately to decreases in corn futures prices. The nitrogen price responses to increasing corn futures prices were mixed. This model works well except that in most cases data on other variables is scarce and this may be limiting.

2.2.2 GARCH Models

Ott (2012) analyzed the drivers of the price volatility of fertilizers and their interplay with energy and food commodity market prices. Three issues were examined: the role of speculations for fertilizer price formation; the interaction among fertilizers, food, and energy prices, and fertilizer price volatility. First, the results showed the presence of speculative behaviour in fertilizer markets. However, the speculation on derivative markets was not the cause. Instead, the volume traded in the physical markets was the driver of volatility in fertilizer markets. It is likely that fertilizer derivative products may have been used as hedging tools and not as speculative ones. Second, the prices of food commodities were driving the fertilizer markets and not vice versa. In addition, higher food prices induced a higher demand for fertilizers, thus increasing food prices to higher levels. Third, the energy sector influenced the increase in fertilizer prices through the input cost channel. Various models (including generalized auto-regressive conditional heteroskedastic (GARCH), GJR and EGARCH were selected and estimated based on the Bayesian Information Criterion (BIC) and the regularity conditions of the quasi-maximum likelihood estimators to be consistent and asymptotically normal. The main result of the study is that crude oil prices exert a greater impact on fertilizer prices during the latter part of 2008. This result may suggest that as the volatility in global fertilizer prices has increased, vital energy prices and agricultural production are likely to be significantly affected. This may lead to future instability in agricultural commodity prices. Thus, it is important to understand the directional relationship between the volatility in energy and global fertilizer prices.

A study by Chena et al. (2012) used both the autoregressive distributed lag (ARDL) model and the generalized autoregressive conditional heteroskedasticity (GARCH) model to evaluate the effects of oil and global fertilizer prices, and to model the volatility in global fertilizer and crude oil prices. The data used in this study was the weekly global fertilizer supply prices that were obtained from the Fertilizer Market Bulletin (hereafter FMB) weekly fertilizer report, while the weekly Dubai crude oil prices are obtained from the database in the Bureau of Energy during the period 2003-2008. He used various volatility models including GARCH, EGARCH, and GJR models to investigate the relationship between crude oil price and six global fertilizer prices. The empirical results from ARDL show that most fertilizer prices were significantly affected by the crude oil price while the volatility of global fertilizer prices and crude oil price from March to December 2008 are higher than in other periods. Chena et al. (2012) findings were in agreement with the report by Huang et al. (2009).

The study by Sanyal et al. (2015) considered the impact of energy price variability on global fertilizer price, application of alternative volatility models. They focused on the mean and volatility effects of oil and natural gas prices on both the mean and volatility changes in fertilizer prices. Results showed that the volatility effects of oil and natural gas prices had significant effects on fertilizer prices. Both symmetric models [GARCH (1, 1)] and asymmetric models [GJR (1,1)] were used to model volatility in fertilizer prices and to evaluate the effects of the volatility over different time periods using Bai-Perron structural break tests. The results show that changes in oil and natural gas prices increased fertilizer prices after the crisis period, during June 2007 to June 2008. Both the ARCH and GARCH had significant effects on fertilizer prices, suggesting that the volatility effects of oil and natural gas prices on fertilizer prices were also significant. The period of data used in this study was from December 1993 to January 2012. The crude oil spot prices, oil future prices, natural gas spot, and natural gas future prices were obtained from the Energy Information Administration (EIA) while the fertilizer prices were obtained from the World Bank Global Economic Monitor Database.

A study by Etienne et al. (2016) used a vector error correction model (VECM) to examine the short- and long-run price dynamics between the natural gas, fertilizer, and corn markets and to evaluate the inter-market volatility spillover. The study used a similar Multivariate Generalized Autoregressive Conditional Heteroskedasticy (MGARCH) framework as in Gardebroek and Hernandez (2013), Ait Sidhoum and Serra (2016). Evidence of price and volatility interplay between ammonia and corn markets and mild transmission from natural gas to the other two markets was found. A relationship between fertilizer and corn existed and was strong in both their long-run price relationship (cointegration) and in volatility spillovers (in both directions). Volatility spillovers from natural gas prices to corn prices for the sub-sample of 2006-2014 vanish, corresponding to the boom in shale gas production in the US. The shale production-driven oversupply of natural gas in the US kept natural gas prices at low levels but this lower price and reduced volatility has not yet transmitted to the other markets.

2.2.3 Machine Learning Models

Shui et al. (2022) created a prediction model using data from China's agricultural material price big data management platform, and a total of 570 records of urea price data from 2014 to 2021 were collected. In this study, bi directional LSTM and the prediction results of the

LSTM model was found to reflect the trend of price data as a whole, but there was still a large error compared with the real value. By contrast, the A-LSTM model and the Bi-LSTM model have higher accuracy in the prediction results, moreover, the prediction curve based on the A-Bi-LSTM model was found to fit the price curve more accurately. This showed that the model in their work could more accurately predict the changing trend of fertilizer price data, and make up for the shortcomings of other models that only utilize unilateral information by extracting the forward and backward laws of time series and making use of the attention mechanism to assign appropriate weights to each feature. Consequently, the proposed model had better feasibility in actual data prediction. However this method may give misleading results because in order to work properly, machine learning models require good data and for this, a little feature engineering is usually done. The objective behind feature engineering is to design more powerful models that exploit different patterns in the data. In so doing, it is somewhat challenging to capture the true market scenarios which ultimately leads to unreliable results making such models relatively inefficient.

The study by Bonilla Cedrez et al. (2020) used Random Forest Regression and Thin plate spline algorithms to fit models. Random Forest has the benefit of flexibility for fitting potentially irregular surfaces resulting from complex interactions. Fertilizer price data for eighteen countries in West and East Africa from two data sources namely Africa data and the Living Standards Measurement Study-Integrated Surveys on Agriculture were compiled. The study found a considerable amount of spatial variation in urea prices in many countries. They did this by fitting models of prices as a function of longitude, latitude, and additional predictor variables that capture aspects of market access, demand and environmental conditions In most countries the spatial models fit the data reasonably well and outperformed the Null model of no spatial variation. They also found that in countries where there were subsidies fertilizer prices were relatively lower. One model was used to make prediction for all the sampled countries and did not perform well but a model for each country performed well as it captured the unique factors for the countries under study. Machine learning models work well with large volumes of data which in most cases is scarce in the fertilizer industry. Also, if the data quality is not good the results of any model may be misleading.

Sabu and Kumar (2020) used time series models and machine learning models to predict monthly prices of arecanut in Kerala. The models included SARIMA, Holt-Winter's Seasonal method and LSTM neural network and their performance evaluated based on the RMSE value on the arecanut dataset. Prices from 2007 to 2017 of arecanut were used. LSTM neural network model was found to be the best model that fits the data with biggest challenge being adequate data to formulate an LSTM Model. These models are capable of producing forecasts with relatively good performance. LSTM model was selected as it has the ability to capture the non-linear dependence of the data points that makes it favourable for time-series forecasting.

2.2.4 Regression Model

Schnitkey (2016) estimated annual anhydrous ammonia prices using a structural model. The study predicted prices with natural gas and corn price as explanatory variables that were obtained from the Energy Information Service (EIA) and National Agriculture Statistics Survey (NASS). Anhydrous ammonia price data was acquired from USDA Economic Research Service (ERS). The model estimation was done using ordinary least squares (OLS) regression. Each of the explanatory variables was found to have positive effects on the price of fertilizer, indicating that as natural gas and corn prices increase, fertilizer prices will also increase. The equation had an adjusted R^2 of 0.88. The results were limited due to the projection of only ammonia. The study emphasizes the rise of natural gas prices until 2006, when a new technology called fracking was introduced in the oil and gas industry. Since then, there have been increases in the supply of natural gas. Anhydrous ammonia prices were highly correlated with natural gas prices until this change. Despite low input prices of natural gas, fertilizer prices remain constant. Schnitkey suggests that farmers could plant more acres into soybeans to put more pressure on the producers of anhydrous ammonia. OLS method assumes linearity making it not fit for formulating models intended for volatile markets.

2.2.5 VAR Model

Newton (2019) used a VAR model to conduct his research on the best time to purchase fertilizer as it included individualistic nature of the markets as variables. The study factored in the regions that included the Corn Belt and Southern Plains regions. While other factors like Natural gas and crude oil had an impact on the price of fertilizer, the study noted no much variablility in terms of time of purchase of fertilizer prices for the two regions and for different fertilizers. The model requires behavioral variables to make the model sufficient which limits its use in making decisions for both the farmer and the fertilizer dealers.

Uçak et al. (2022) analysed the effect of volatility in fertilizer prices on selected agricultural products by using the Diebold-Yilmaz connectedness approach, which is based on time-varying parameter (TVP) vector auto-regression (VAR). The findings showed that the spread of volatility and the interconnectedness between these variables increased in times of crisis and that the risk pass-through was due to fertiliser prices. However, empirical results showed that the price volatility of phosphate rock and urea was highly correlated to the volatility of other products. Further to this, it was found that sugar, soybean and cotton were the agricultural products most vulnerable to the effects of external shocks.

2.2.6 SARIMA GARCH Models

In their study, Bhardwaj et al. (2014) used the Box Jenkins Autoregressive integrated moving average (ARIMA) and Generalized autoregressive conditional heteroscedastic (GARCH) models and applied it to model and forecast the spot prices of Gram at Delhi market. Augmented Dickey Fuller (ADF) test was used for testing the stationarity of the series. ARCH-LM test was used for testing the volatility. The study found that ARIMA model cannot capture the volatility present in the data set whereas GARCH model had successfully captured the volatility. Root Mean square error (RMSE), Mean absolute error (MAE) and Mean absolute prediction error (MAPE) were computed. The GARCH (1,1) was found to be a better model in forecasting spot price of Gram. The values for RMSE, MAE and MAPE obtained were smaller than those in ARIMA (0,1,1) model. The AIC and BIC values from GARCH model were smaller than that from ARIMA model. Therefore, it shows that GARCH is a better model than ARIMA for estimating daily price of Gram.

2.2.7 Hybrid SARIMA GARCH Models

Shetty et al. (2018) used three models namely, GARCH, SARIMA and SARIMA-GARCH to model the gold prices. They compared the accuracy between the models based on error statistics such as MAPE, MAE and RMSE. The forecasts results by SARIMA-GARCH were found to be better since the RMSE, MAE and MAPE were lower than those produced by ARIMA and GARCH. He concluded that in the case of the monthly selling prices of 1 Troy ounce Indian gold, the hybrid model of SARIMA-GARCH can be an effective way to improve the forecasting accuracy. The data used in the study is the monthly selling price of 1 Troy ounce Indian gold from the period 1st January 2000 to 1st December 2017

Kaur et al. (2022) applied autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroscedasticity (GARCH), exponential GARCH (EGARCH) and threshold GARCH (TGARCH) model along with other estimation procedures on the daily closing prices of Nifty 50 from Jan 1, 2009 to Dec 30,2019. The study identified ARIMA(2,1,2)-EGARCH(1,1,1) as best models to predict the closing prices of Nifty 50. Their findings indicated that the static forecast provided better results as compared to the dynamic forecast.

2.2.8 Conclusions

Various studies have been conducted in a bid to develop models that accurately forecast fertilizer prices (Galbraith, 2010; Kim and Brorsen, 2017; Ott, 2012; Shui et al., 2022). Stand alone models have been critisized for failing to capture the true market conditions by capturing only the unilateral information. Others like the A-Bi-LSTM model have been developed and had better accuracy in actual data prediction. Their limitation however is the fact that availability of data is limited and therefore a little feature engineering is usually done.

The objective behind feature engineering is to design more powerful models that exploit different patterns in the data which may fail to capture the true market information and as a result may lead to unreliable results making such models relatively inefficient. Combined models, however, have been praised for yielding better performing models. Specifically, time series models have been hailed for yielding better performing models. The models that have been formulated so far have factored in variables including Natural gas, transport,volumes traded, crude oil prices, corn prices, ethanol, market concentration and regions.

2.2.9 Current Research

This thesis, seeks to formulate a Hybrid SARIMA GARCH model and will add to the existing body of knowledge a model that works by factoring in the port through the fertilizer is transported in the model. The SARIMA and GARCH processes have been applied in various fields such as mobile communication networks, climatology, tourism , and economics to mention a few. No study has yet investigated the port factor using this framework to study the behaviour of fertilizer price. This is important as most of the fertilizer used is imported and understanding whether ports are key in the pricing of fertilizer will be important information for fertilizer dealers. We shall formulate the model and apply the model to determine the best times to purchase fertilizer from the global market at a lower cost and identify the best port for a given season.

Chapter 3

Methodology

3.1 Introduction

This chapter discusses the data set used, SARIMA, GARCH and Hybrid SARIMA-GARCH methods that are used to model the data, provide statistical inference and parameter estimations. Goodness of fit test is conducted to determine the best model for this data. The first sections tackles the development of a SARIMA model, GARCH Model, and finally a Hybrid SARIMA GARCH model. The aim is to formulate a Hybrid SARIMA GARCH model that accurately and efficiently forecasts fertilizer prices in Kenya. The last section addresses evaluation of the model performance.

3.2 Data

To carry out this study, fertilizer price time series from two ports, one from Yuzhny and another from Arab gulf, are used to perform the task. For both ports the data comprises of 996 observations sourced from Profercy nitrogen. Weekly fertilizer price data from 2004 to 2018, is used for the development of models. Data from 2017 to 2018 is used for the validation of the developed models. The study uses SARIMA modelling to find optimal mean equation. The residuals are added in the symmetric or asymmetric GARCH models that is identified to measure the variability of the series. The data Analysis is conducted using the R Software R version 4.2.3 (2023-03-15 ucrt). In this study, several steps are presented: i) dealing with missing values ii) the choice of hypothesis tests used to verify stationarity and normality of the data: as well as autocorrelations and heteroskedasticity of the residuals from the selected model; iii) the best model is selected using goodness of fit information criteria;

and iv) the forecasting performance of the coupled Hybrid SARIMA-GARCH model. The Figure 3.1 illustrates the steps we shall employ in the work as used by (Pandey et al., 2019).

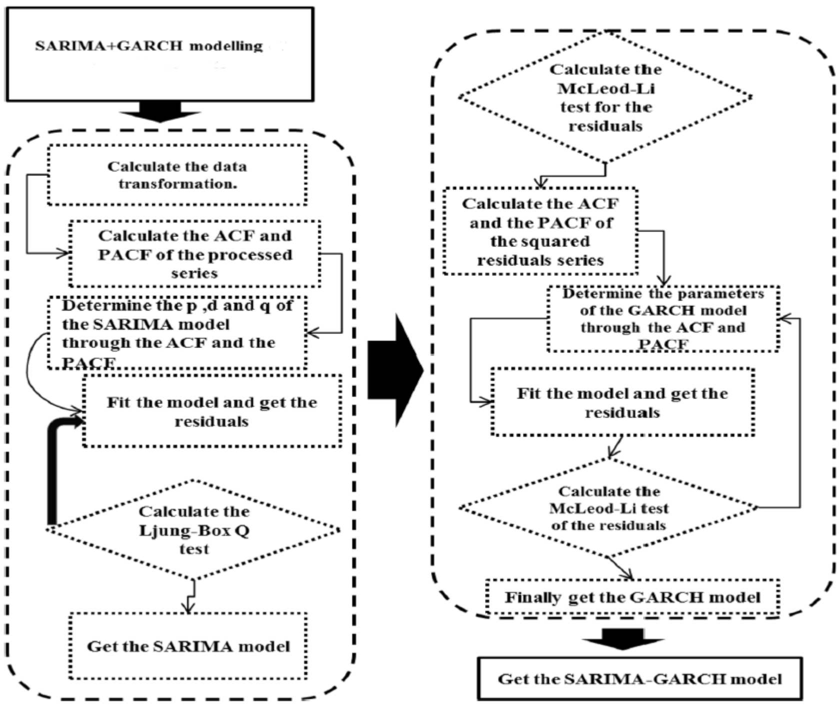


Figure 3.1: Model formulation work flow

3.3 Data Analysis

Time series data of fertilizer prices consists of successively generated observations over time such that data is ordered with respect to time and successive observations may be dependent. The underlying process to the data may indicate long-term behavior such as trend and or cyclical fluctuations that depict seasonality. Using different models to capture such underlying processes enables forecasts of the likely observation at a time point in future to be made. Based on the literature reviewed, the preferred classes of models that will be explored are SARIMA, GARCH, and the Hybrid SARIMA-GARCH Model.

3.4 Model Formulation

3.4.1 SARIMA Model

Seasonal ARIMA models are an extension of ARIMA models when seasonal terms are included. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. The SARIMA time series model may be expressed as a combination of both non-seasonal and seasonal parameters as shown below:

$$Y_t = (1 - B^d)(1 - B^S)^D e_t, \tag{3.1}$$

where $(1 - B^d)$ is the differencing operator of order d and $(1 - B^S)^D$ is the seasonal differencing of order D, S represents the seasonal length, B is the backshift operator, Y_t is the observed series at time t and e_t is the independent identically distributed (I.I.D) error with zero mean and standard deviation (SD) equal to 1. Graphically, the ACF and PACF can be quicker methods to be used in identifying the proper order of SARIMA models and to check fitted model accuracy. A study by (Modarres and Ouarda, 2012), indicates that model adequacy is ensured by conducting diagnostic tests done by checking the randomness of residuals, and independent identically distributed (I.I.D) error terms has zero mean and unit variance at 95 percent confidence limits (± 1.96).

A preliminary Box-Jenkins analysis is conducted by making a run sequence plot of the initial fertilizer time series data as the starting point in determining an appropriate model. This entails the 4 iterative stages of identification, estimation, diagnostic checking and forecasting.

Preliminary procedures include making the data a stationary series, done by differencing. This is followed by identifying seasonality (we decompose the data to clearly see the components of the data) and conducting seasonal differencing as well since seasonality is present in the data. Auto arima function in R software is be used to select the best model or we may consider Auto correlation and partial autocorrelation functions of the time series to determine which Auto Regressive and or Moving Average component to be used in the model and also factoring the seasonal component. Partial Auto correlation function measures the correlation between an observation k periods ago and the current observation, after controlling for observations at intermediate lags (all lags<k) and is useful for telling the maximum order of the AR. The Auto Correlation Function plot is useful for telling the order of the Moving Average process. We then used available methods that include maximum likelihood, the method of least squares, or Yule-Walker equations for the parameter estimation. This stage of model identification and estimation also involves the use of the Akaike Information Criterion (AIC) of selecting the best model from a set of models. The chosen model has the lowest AIC. The diagnostic checks entail the analysis of the residuals where we plot the standardized residuals, the autocorrelation function of the residuals, and the p-values for Ljung-Box Q statistic. At this stage, the assumptions of the ARIMA model are checked, e.g. the hypothesis of errors being independently and normally distributed (Kokkinen and Wouters, 2016). The forecasting stage will entail using the estimated model to generate forecasts and their confidence limits.

3.4.2 Hybrid SARIMA GARCH Model

The generalized auto regressive conditional heteroscedasticity (GARCH) family models were developed by Engle and Bollerslev (1986), and are considered the best in capturing volatility clustering and predicting volatilities in the future. GARCH aims to minimize errors in forecasting by accounting for errors in prior forecasting and enhancing the accuracy of ongoing predictions. The GARCH model will be applied to residuals of the SARIMA model after verifying by the suggested test of the square of residuals in this case the ARCH LM test. There is a possibility of time-dependent variance even in well-fitted SARIMA models, which can be addressed by using a GARCH model. As a hybrid, the SARIMA–GARCH methodology means that behavior of time series is modeled using a SARIMA model and conditional variance using GARCH models. Engle (1982) introduced ARCH models, which were later modified by Bollerslev (1986), by introducing lagged conditional variance to improve the smoothing of series. The GARCH (A, B) model is defined by Modarres and

Ouarda (2012) as :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{A} \alpha_A \varepsilon_{t-A}^2 + \sum_{j=1}^{B} \beta_B \sigma_{t-B}^2, \tag{3.2}$$

where $\alpha_0 > 0$,$\alpha_i \geqslant 0$, i=1,2,......A and $\beta_j \geqslant 0$, j=1,2,......B and $\sum_{i=1}^{A} + \sum_{j=1}^{B} < 1$ are the parameters of the GARCH (A, B) model. Where σ_t^2 represents the conditional variance of the fertilizer prize at time t, α_0 is the constant term or the unconditional variance, α_A is the coefficient of the lagged squared residuals or the auto regressive term, β_B is the coefficient of the lagged conditional variances or the moving average term, ε is the error term. The GARCH model assumes that the volatility of asset returns varies over time and is influenced by its own past values and the unanticipated shocks to the return. The auto regressive and moving average terms capture the persistence and mean reversion of volatility, respectively, while the constant term represents the long-run average level of volatility. The proposed hybrid SARIMA–GARCH is one in which the variance of the error term of the SARIMA model follows a GARCH process. It is designed by combining SARIMA equation and GARCH equation. The SARIMA–GARCH model was obtained as

$$\left.\begin{aligned} Y_t &= (1-B^d)(1-B^S)^D e_t \\ \sigma_t^2 &= \sum_{i=1}^{A} \alpha_A \varepsilon_{t-A}^2 + \sum_{j=1}^{B} \beta_B \sigma_{t-B}^2. \end{aligned}\right\} \tag{3.3}$$

3.5 Data Transformation

We transformed the fertilizer price time series using differencing to make non stationary data stationary.

3.6 Correlation Test of Sequence

We applied various methods for correlation test of sequences. Auto correlation function (ACF) For a time series, u_t the autocorrelation coefficient of k-order lag can be estimated as below

$$r_k = \frac{\sum_{t=k+1}^{T}(u_t - \bar{u})(u_{t-k} - \bar{u})}{\sum_{t=1}^{T}(u_t - \bar{u})}, \tag{3.4}$$

where r_k is the k-order auto correlation coefficient of u, which is the mean of time series u_t. The auto correlation coefficient can represent correlation degree between the adjacent data. Partial auto correlation coefficient (PACF) Partial auto correlation coefficient refers to the conditional relevance between u_t and u_{t-k} after adjusting for all other terms of shorter lags i.e. $u_{t-1} \cdots u_{t-k-1}$. The partial auto correlation coefficient $\rho_{k,k}$ of k-order lag can be estimated as below when k is greater than 1

$$\rho_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} \rho_{k-1}, j r_{(k-j)}}{1 - \sum_{j=1}^{k-1} \rho_{k-1}, j r_{(k)}}. \tag{3.5}$$

The Ljung–Box LB (Q) test was used to check autocorrelation in a time series at various lags by adjusting the degree of freedom. The LB (Q) statistic is given as

$$\rho_{LB} = T(T+2) \sum_{j=1}^{p} \frac{r_j^2}{T-j}, \tag{3.6}$$

where r_j represents the squared autocorrelation values for residual time series, T is the size of time series and p represents the autocorrelation numbers of residuals included in the statistic. Before ARCH tests, it is suggested to test the null hypothesis of ARCH effect. Based on LB (Q) test results, the decision to apply ARCH test can be made.

3.7 Tests for Heteroscedasticity Using Auto-Regressive Conditional Heteroskedasticity (ARCH LM Test)

The ARCH LM test is recommended to check for ARCH effect in the residual time series. The test involves regressing the squared residuals on their lags and checking if they are significant and is expressed as:

$$\sigma_t^2 = E\left[(x_t - \bar{x}_t)^2\right] = E\left[x_t^2\right] - \bar{x}_t^2, \tag{3.7}$$

where σ_t^2 is the conditional variance and $\bar{x}_t$ is the conditional mean.

Where the time series does not have a significant mean, then the conditional variance is expressed as:

$$\sigma_t^2 = E\left[(x_t - \bar{x}_t)^2\right] = E\left[x_t^2\right] = E\left[y_t\right] \approx x_t^2. \tag{3.8}$$

3.8 Evaluation of Model Performance

Multi-criteria decision analysis (set of statistical indices) approach will be used for assessing the performance of the developed models. The criteria decision analysis will use methods that include indices based on absolute error, relative error and dimensionless metrics. The performances of the data-driven models will be evaluated to compare their predictive accuracies based on the following statistical indices as suggested by (Modarres and Ouarda, 2012).

The mean absolute error (MAE) is given by,

$$MAE = \frac{1}{n}\sum(|\rho_i - \rho_p|), \tag{3.9}$$

and the root mean square error (RMSE) is

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(\frac{\rho_i - \rho_p}{n}\right)^2}. \tag{3.10}$$

where, ρ_i is the observed time series, ρ_p is the predicted time series, $\bar{\rho}$ is the mean of the observed time series, and n is the number of observations.

Chapter 4

Results and Interpretation

4.1 Introduction

This chapter discusses the descriptive statistics, diagnostic tests, results from the models fitted, and the interpretation of the findings. The significance of modelling the auto regressive integrated moving average (ARIMA) models in fertilizer price modeling is primarily for their use in determining whether there is a nonlinear mechanism in the processes that generate the data that can be well explained by GARCH class models. The generalized auto regressive conditional Heteroskedastic model is a nonlinear model derived from a residual series of an ARIMA model. In order to check and model the behavior of fertilizer price data in the study areas, the weekly and monthly average fertilizer price data sets of Yuzhny and Arab Gulf for the period 01/01/2004 to 31/04/2018 were used. In this scenario, the time series(s) were decomposed, the seasonal effects were estimated and removed from the series, producing the series free of seasonality. This is the fundamental technique of fitting an appropriate ARIMA model to time series. The series(s) looks to be quite stable over time after differencing. Using Box and Jenkins' methods, the orders p and q of the ARIMA models were determined and calculated for both series. The fitted Seasonal ARIMA model results are shown in Table 4.3 and Information criteria were used to choose the models.

4.2 Descriptive Statistics for the Ports of Yuzhny and Gulf

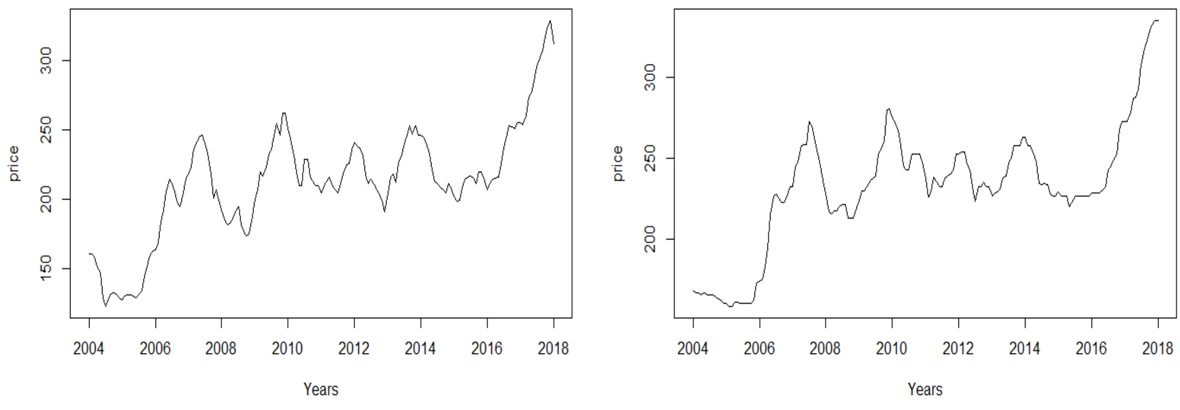
The descriptive statistics displayed in Table 4.1 indicate a mild skewness and kurtosis. The values for skewness are between (-2,+2) and kurtosis between (-7,+7) which are within the acceptable range of normality (Demir, 2022) and therefore we can work with the data without transforming to remove skewness.

Table 4.1: Descriptive statistics port of Yuzhny and Gulf

Measures	**Yuzhny**	**Gulf**
Mean	294.6	308.8
median	265.0	280.0
maximum	825.0	817.5
Minimum	122.5	158.5
Skewness	-0.05	-0.08
Kurtosis	3.60	3.61

4.3 Time Series Plot for Yuzhny & Gulf

A trend and seasonality exist in the fertilizer price data hence data is non-stationary as in Figures 4.1(a) and 4.1(b). This is also confirmed by (See Figures A.3 and A.4) in Appendix B which represent the decomposition of the observed time series. Decomposition of the observed series into three components for both ports, namely trend, seasonal and random components. Here estimated random component is obtained by eliminating the estimated trend and seasonal components from observed time series.



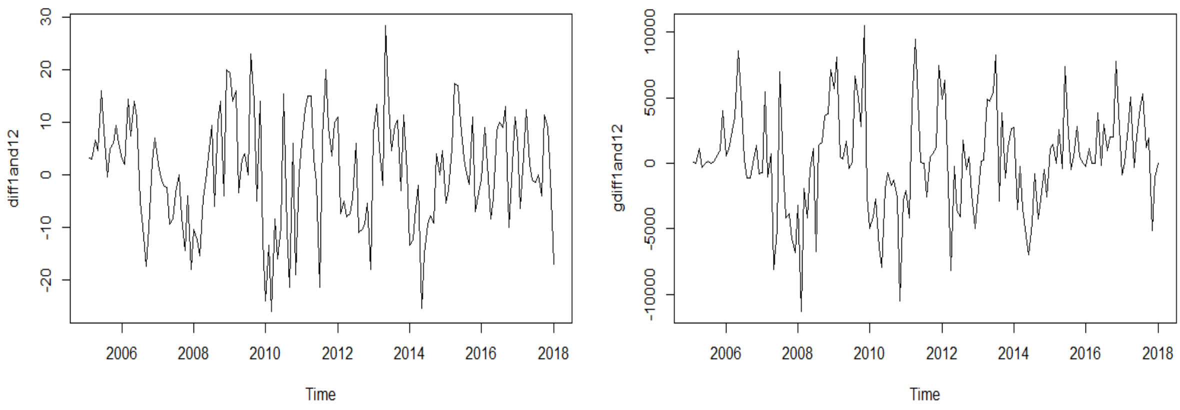
(a) (b)

Figure 4.1: Panel (a) shows the Time series plot for Yuzhny. Panel (b) shows the Time series plot for Gulf.

The raw fertilizer price time series auto correlation function (ACF) and partial auto correlation function (PACF) for both ports show a progressively fading pattern that echoes the series' non-stationarity for both the ports (see Figures A.1 and A.2) in Appendix A.

4.4 Data Transformation

The data is transformed by taking the seasonal and first differences for both the Yuzhny and Gulf data sets. The transformed data run sequence plots exhibits stationarity as shown in Figures 4.2(a) and 4.2(b) respectively. The variance appears to be relatively stable.



(a) (b)

Figure 4.2: Panel (a) shows the transformed Yuzhny data. Panel (b) shows the transformed Gulf data.

4.4.1 Stationary Test Results

The transformed series exhibits stationarity evidenced by the Augmented Dickey-Fuller test (ADF test) that yields a p-values of 0.01 for both the ports which is less than the threshold of 0.05 and hence there is sufficient evidence to conclude that the transformed time series is stationary. (see Table 4.2). Other stationarity tests were done as in Table 4.2 and confirmed that transformed data through differencing was stationary. Since the data was skewed as in Table 4.1. We attempted to reduce skewness using available methods like taking the square, square root, Log, reciprocal, cube, and sin but these techniques were unsuccessful in reducing the skewness so we worked with the data as is with the assumption that the skewness was negligible (we assumed normality) as the skewed value before attempting to reduce it was the lowest. We attempted to estimate a SARIMA model using transformed data (by squaring data) but the models yielded were not the best since they had way too big AIC values (see Table 4.10).

Table 4.2: Stationarity Test Results

Test	Differenced Yuzhny Data	Differenced Gulf Data
Augmented Dickey-Fuller Test	-4.0529(0.01)	-4.163(0.01)
KPSS Test	0.052432(0.1)	0.09475(0.1)
Phillips-Perron Unit Root Test	-92.029(0.01)	-88.644(0.01)

4.5 Model Estimation for ARIMA (p,d,q)(P,D,Q)12 Models

To eliminate any potential researcher bias, the best models were chosen using the auto ARIMA algorithm. The auto arima function in R suggested a number of models, and the program selected the best and most parsimonious model based on four criteria: the Akaike information criterion (AIC), loglikelihood criteria, the Ljung Box statistic, and the Augmented Dickey Fuller Unit Root Test of the residuals of the models. According to diagnostic examination, (see Figures A.5 and A.6 in Appendix C). The residuals appear to follow a normal distribution going by the Normal Q-Q plots, the residuals are white noise. The p-values in the Yuzhny port diagnostics are not good and suggest that we need to fit the residuals to a GARCH model while those p-values of the Gulf port indicate that it a good model. The most suitable SARIMA models for the observed series are SARIMA (2, 1, 0) (2, 1, 0)[12] for the Gulf port and SARIMA (1, 1, 0) (2, 1, 0)[12] for the Yuzhny port as shown in Table 4.3. The RMSE (5.601546 and 7.843295) for both models are the lowest for all the SARIMA models fit. We see a very strong absence of correlation in the Gulf port given the Ljung Box Test p-value(0.9508) while that of the Yuzhny port is just 0.1661 which may warrant us to move to GARCH model.

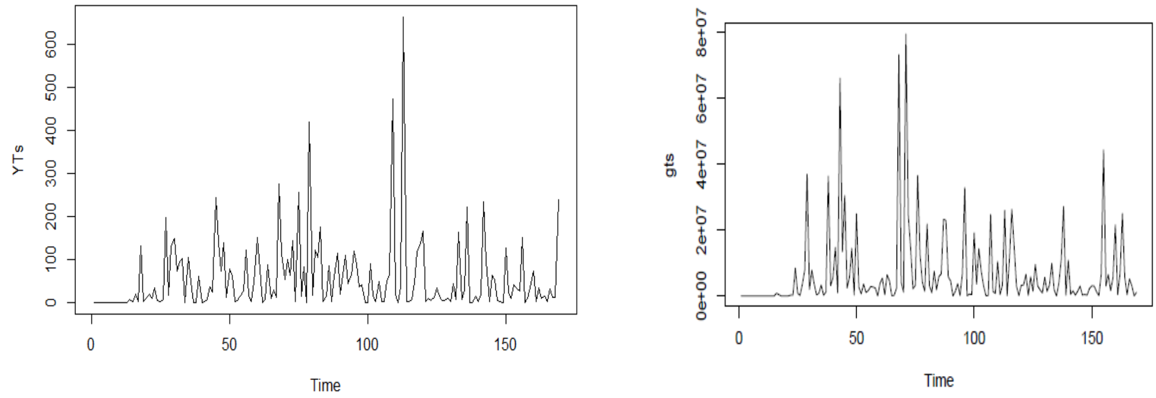
4.5.1 The Most Effective SARIMA Models, Evaluated

Table 4.3: Estimated SARIMA models

Parameter	**Gulf ARIMA(2,1,0)(2,1,0)[12]**	**Yuzhny ARIMA(1,1,0)(2,1,0)[12]**
Intercept	0	0
AR(1)	0.3960(0.0791)	0.4226(0.0759)
AR(2)	0.1419(0.0794)	-
SAR(1)	-0.7009(0.0742)	-0.6760(0.0829)
SAR(2)	-0.3530(0.0740)	-0.2648(0.0829)
AIC	997.52	1008.86
BIC	1012.77	1020.71
log likelihood	-493.76	-500.43
RMSE	5.601546	7.843295
Ljung Box Test p-value	0.9508	0.1661

4.5.2 Plots of Squared Residuals, their ACF and PACF.

The plots in Figures 4.3(a) and 4.3(b) show that the squared residuals have non constant variance (Heteroskedastic). This may have been due to world food prices that increased dramatically in 2007 and the first and second quarter of 2008, creating a global crisis and causing political and economic instability and social unrest in both poor and developed nations. A visual inspection of the squared residuals via the ACF and PACF shows that some lags fall slightly outside the confidence band (see figures 4.4 and 4.5). This confirms the non constant variance and therefore necessitates the generalized auto regressive conditional Heteroskedastic model (GARCH).



(a) (b)

Figure 4.3: Panel (a) shows the Yuzhny squared residuals. Panel (b) shows the Gulf squared residuals.

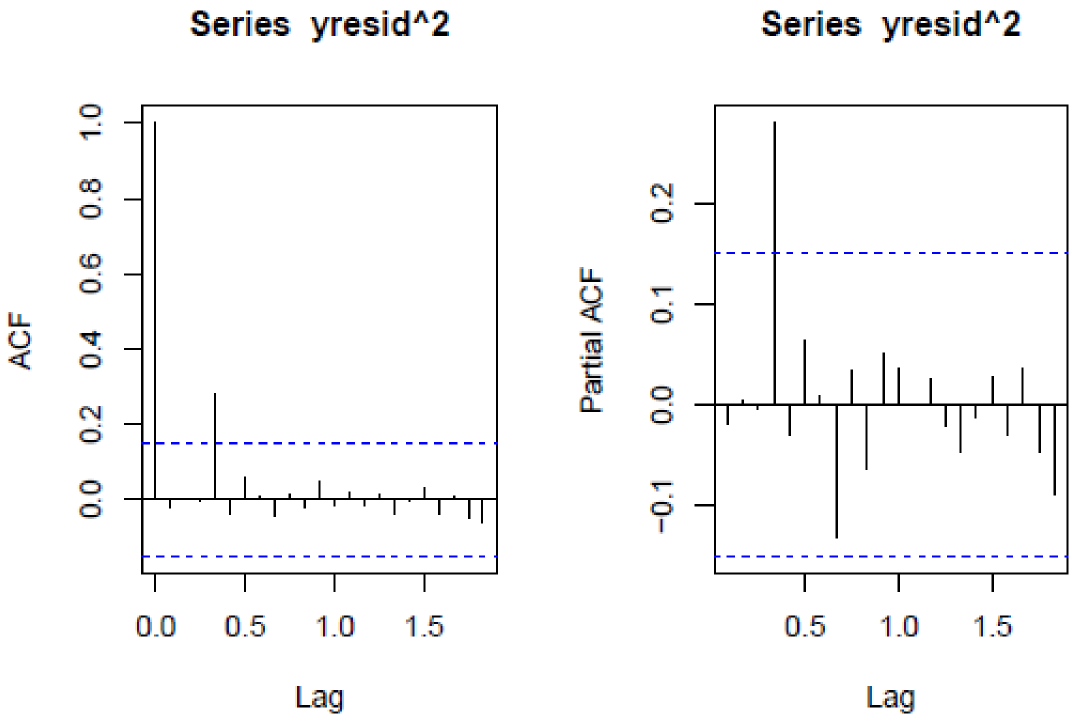


Figure 4.4: Plot of ACF and PACF Yuzhny squared residuals

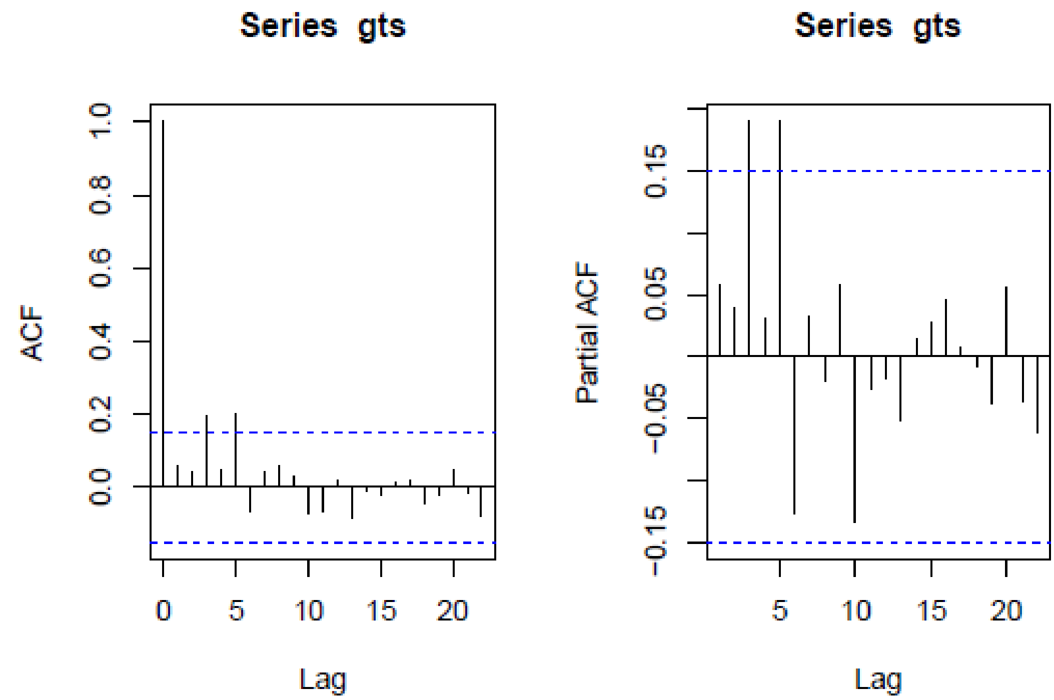


Figure 4.5: Plot of ACF and PACF Gulf squared residuals

4.6 Estimation of GARCH Models on the SARIMA Residuals

The Hybrid SARIMA-GARCH model is one in which a GARCH process governs the variance of the SARIMA model's error component. The Hybrid SARIMA-GARCH model combines two models: (1) the SARIMA model, which represents a time series' mean behavior, and (2) the GARCH model, which models the variance behaviour (ARCH effect) of the residuals. Appropriate GARCH models were constructed using the squared residual series from the fitted SARIMA models.

4.6.1 Testing for ARCH Effects in the Squared Residuals of Estimated SARIMA Models

Before fitting the GARCH model, the ARCH LM test was employed to see if ARCH effects were evident in the squared residuals. The ARCH LM test findings are shown in Table 4.4, where they fail to reject the Gulf port's hypothesis that there is no ARCH effect in the squared residuals. Accordingly, GARCH modeling is required for the Yuzhny port.

4.6.2 ARCH Effect Test Results on the Residuals of Gulf and Yuzhny SARIMA Models

The squared residuals for Yuzhny port pass the ARCH LM test as shown in Table 4.4.

Table 4.4: ARCH effect test for the Gulf and Yuzhny SARIMA residual and squared residuals

	Test	**Statistic**	**p-value**
residuals	ARCH LM-test(Gulf)	14.58	0.2652
	ARCH LM-test (Yuzhny)	16.999	0.1496
squared residuals	ARCH LM-test(Gulf)	20.353	0.0607
	ARCH LM-test(Yuzhny)	25.063	0.01453

4.6.3 GARCH Model Fitting for Squared Residuals of SARIMA Gulf

The port of Gulf's residuals and squared residuals failed the ARCH effects test (see Table 4.4). It is obvious that we cannot continue with the fitting of a hybrid SARIMA GARCH model for this port. To see whether it could capture any information, GARCH modeling was also evaluated for the Gulf port, but our efforts were unsuccessful since the program showed non-convergence errors, which made it evident that the Hybrid SARIMA GARCH model was not a suitable fit for the data. Consequently, we draw the conclusion that the SARIMA model can adequately represent the time series generation process for this port. This result is consistent with a research by Mishra et al. (2014) who looked at three primary fertilizers used in India: nitrogen (N), phosphorus (P), and potassium (K), all of which are important throughout various phases of crop development. The ARIMA approach was utilized in the investigation, and it was discovered that the ARIMA (1, 1) model was best suitable for estimating the data on nitrogen production. They made use of data on fertilizer from 1961 to 2002. The model's predicted values indicated that production will rise somewhat in the future. The results of this study were trustworthy since fertilizer prices have steadily climbed over time, showing that models in the ARIMA class can likewise perform a good job of modeling fertilizer prices.

4.6.4 Forecasts

According to our study, the forecast for the port of Gulf indicates that in 2018 the prices dropped and remained relatively steady which agrees which the actual occurrence at the time and also as determined by Galbraith (2010) and hence our model could pass for a reliable model. The model was used to generate a forecast as shown in Figure 4.6.

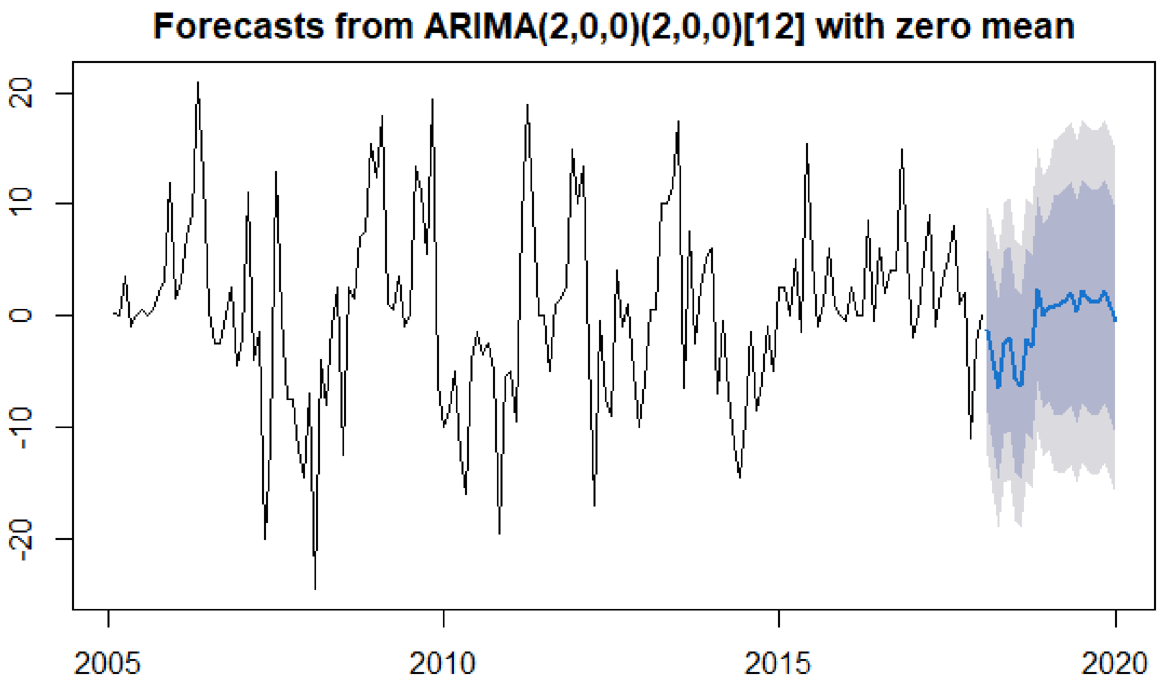


Figure 4.6: SARIMA forecast plot for Gulf port

4.6.5 GARCH Model Fitting for Squared Residuals of SARIMA for Yuzhny

On the Yuzhny SARIMA model residuals, a number of GARCH models were estimated (see Table 4.5). The model with the lowest AIC (10.360) was chosen as the best one in this case sGARCH(1,1) with the distribution specified as Student's t distribution (sstd).

Table 4.5: Extensions of GARCH models fitted for Yuzhny port

Model	Distribution	AIC	Significance
eGARCH(1,1)	sstd	10.366	ω & β_1 significant
eGARCH (1,1)	snorm	10.966	ω significant
gjr-GARCH(1,1)	sstd	10.375	ω, β_1, γ_1 significant
gjr-GARCH(1,1)	snorm	11.000	ω significant
sGARCH(1,1)	sstd	10.360	ω & β_1 significant
sGARCH(1,1)	snorm	10.952	ω significant

4.6.6 The Results of the Estimated sGARCH(1,1) Model for Yuzhny on the SARIMA Residuals

The SARIMA (1, 1, 0) (2, 1, 0)[12] model's residuals are fitted using the R-studio program's Generalized Auto Regressive Conditional Heteroskedasticity (GARCH(1,1)) function and it yielded the lowest AIC value of (10.360). With the exception of the α_1 , all coefficients are statistically significant (see Table4.6) . Within standardized residuals, there is no correlation. The residuals of the GARCH model after executing model diagnostics show a more or less white noise, suggesting no correlation as in Figures A.11 and A.12 in appendix F. The Ljung box test on both the residuals and squared residuals give p-values = 0.6744 and p-value=0.6712 respectively indicating that the residuals are I.I.D. which is also an assumption for the Ljung box test. The Ljung box test also assumes that the data is drawn from a Gaussian distribution which was the assumption in our case after every attempt to transform our data yielded bad models. According to Burns (2002), the widely accepted number of lags for an Ljung box test is 15. In our study we considered the first seven lags. An ARCH effect test using the ARCH LM test revealed no arch effects on the residuals at lag[1] but present in other lags, indicating that the model had insufficiently captured them and while it could be considered a good model it is not the best.We however see a better model compared to SARIMA model in terms of AIC value (see Tables 4.10).

Table 4.6: Results of GARCH(1,1) model Residuals of SARIMA

Parameter	Estimate	Std. Error	t value	p-value
μ	54.61697	0.106499	6.0801e+02	0.000
ω	31.49469	0.051799	3.2943	0.001
α_1	0.0000	0.000009	8.8000e-05	0.999930
β_1	0.99623	0.000382	2.6062e+03	0.0000

4.6.7 The Results of the Estimated Hybrid SARIMA GARCH Model for Yuzhny

The Hybrid SARIMA-GARCH model is a combination of two models: (1) ARIMA model that take into account the mean behavior of a time series and (2) GARCH model which is used to model the variance behavior (ARCH effect). The results of the combined model SARIMA(1, 1, 0) (2, 1, 0)[12] GARCH(1,1) is given in Table 4.7.

Table 4.7: Results of Hybrid SARIMA-GARCH model

	Yuzhny			
Parameter	Estimate	Std. Error	t value	p-value
ar1	0.6466	0.146592	4.41076	0.00001
ma1	-0.226327	0.192786	-1.17398	0.24040
ω	4.237273	4.822484	0.87865	0.37959
α_1	0.064715	0.043495	1.48785	0.13679
β_1	0.894368	0.067107	13.32754	0.000

Given the the results in Table 4.7, we can write the equation as

$$\left.\begin{aligned} Y_t &= 0.6466y_{t-1} + 0.226327w_{t-1}, \\ \sigma_t^2 &= 4.237 + 0.065\varepsilon_{t-1}^2 + 0.894\sigma_{t-1}^2 \end{aligned}\right\} \quad (4.1)$$

4.6.8 ARCH Effect Tests on the Residuals of the Hybrid SARIMA GARCH Model

The diagnostics carried out on the residuals of the Hybrid SARIMA GARCH model indicate that the ARCH effects at all lags have been sufficiently captured (see Table 4.8). At lags 3, 5, and 7 the p-values yielded fail to reject the null hypothesis that there are no ARCH effects.

Table 4.8: Weighted ARCH LM Tests

	Weighted ARCH LM Tests results	
ARCH Lag	Statistic	P-value
3	0.6681	0.4137
5	2.7315	0.3312
7	3.5934	0.4090

4.6.9 Heteroscedasticity Test for Residuals of SARIMA-GARCH Model

The results of the Box-pierce and L-Jung Box test shows that ARCH effect absence in the residual series since residual and squared residuals of fitted SARIMA-GARCH model does not suffer from serial correlation (see Table 4.9). This shows that the model has satisfactorily captured the arch effects and is therefore a good model. The results of the residual diagnostics are also captured in Figures A.13 in Appendix G. The residuals follow a normal distribution and the ACF of the standard residuals and the squared standard squared residuals show a more or less white noise and this could pass for a good model.

Table 4.9: Box test results on residuals of Hybrid SARIMA GARCH model

Test	Statistic	p-value
Box-Pierce	0.074592	0.7848
Ljung Box	0.076168	0.7826
For squared residuals		
Box-Pierce	0.60181	0.4379
Ljung Box	0.61453	0.4331

4.6.10 Fitted GARCH(1,1) Model on Squared Residuals of SARIMA

For each port, a variety of models were estimated during the course of the study, and the best models as determined by the AIC values are displayed in Table 4.10. We saw in the literature that many researchers used various methodologies in their study. In model formation, stand-alone models performed well, but this work was frequently enhanced by extending the analysis, using hybrid models, and also taking into account factors that may provide the models more nuance. We said that the models with the lowest AIC values would be chosen as the best. This study is consistent with the findings of earlier research (Shetty et al., 2018) and (Pandey et al., 2019) on employing hybrid models to improve model performance, as seen in Table 4.10. This is so because the hybrid model's AIC values are the lowest.

Table 4.10: AIC values for the best models estimated

SARIMA Gulf transformed	nontransformed Gulf	SARIMA Yuz	GARCH	Hybrid Yuz
2930.23	997.52	1008.86	10.360	7.4389

4.6.11 Forecasting Results for Yuzhny port for the models fitted

In terms of performance, the models' MASE, MAE, and RMSE values were calculated, and the model with the lowest values of these metrics was chosen as the best model. We used the listed evaluation metrics since we had already taken into account the effect of their performance due to outliers which was addressed by checking for outliers at the initial stages of our work and found no significant outliers. This was also confirmed by the data working well without conducting other transformation methods. The Hybrid SARIMA GARCH model was superior to the stand alone SARIMA model, as evidenced by the findings shown in Table 4.11 which showed that it recorded the lowest values of the evaluation metrics.

Table 4.11: Forecasting Results for Yuzhny port for the models fitted

	SARIMA	**SARIMA-GARCH**
MASE	0.3876	0.6885
MAE	6.1646	5.4797
RMSE	7.8433	7.5802

The performance indices as shown in Tables 4.11, and 4.10 indicate that SARIMA–GARCH models fitted to fertilizer price time series performed best. The findings of this study are consistent with those of Shetty et al. (2018) and Modarres and Ouarda (2012). They both used Hybrid SARIMA GARCH models alongside other methods in their work and found that Hybrid SARIMA GARCH models outperformed the ARIMA,GARCH and other stand alone models. It can be concluded that in the case of the fertilizer prices, the hybrid model of SARIMA-GARCH can be an effective way to improve the forecasting accuracy and adds to the existing information that the port through which fertilizer is imported through has an effect on the price imposed locally.

Chapter 5

Discussions, Conclusions and Recommendations

5.1 Introduction

The objective of this research was to create a hybrid SARIMA GARCH model for fertilizer price prediction in Kenya. In addition, the study was designed to determine whether the model could be used to select the optimum port for fertilizer import at various periods and to determine whether the port through which fertilizer is imported affects the price charged. To conduct this study, we cleansed the data and transformed it using sophisticated algorithms to guarantee a valid result. Then, the GARCH, SARIMA, and Hybrid SARIMA GARCH techniques were used to analyze fertilizer price trends.

5.2 Discussions

In our work, we formulated and applied a mathematical model of fertilizer prices in two ports namely Yuzhny port and Gulf port. We fitted various models namely SARIMA, GARCH and Hybrid SARIMA-GARCH and used the AIC values to pick the best model. Other models that have used similar techniques were investigated by (Shetty et al. (2018) and Pandey et al. (2019) among others. We also determined whether the port through which fertilizer is imported influences the price imposed. In comparing our model to others, we strictly considered the methodology used to undertake the work. For instance, Shetty et al. (2018) in his work, employed the hybrid model of the linear seasonal autoregressive moving average (SARIMA) and the non-linear generalized autoregressive conditional heteroscedasticity (GARCH) in modeling and forecasting the Indian gold price. The goodness of fit of the model is measured using Akaike information criteria (AIC), while the forecasting performance is assessed using root mean square error (RMSE), mean absolute Error (MAE) and mean absolute percentage error (MAPE). He estimated SARIMA model which scored an RMSE(140.7048), MAE(136.4942) and MAPE (10.79058). The GARCH model was also estimated and upon evaluation the RMSE(103.552183), MAE(99.4379) and MAPE (7.8532). The Hybrid SARIMA-GARCH scored RMSE(53.05311), MAE (58.6986) and MAPE (4.14353). The study concluded that Hybrid SARIMA-GARCH is a more appropriate model forecasting Indian gold price. Our work in Yuzhny port, followed the same trend in terms of performance in that after fitting our SARIMA model, we kept improving it using GARCH and finally the hybrid SARIMA-GARCH was the best. We also evaluated our models using the same evaluation metrics and the Hybrid SARIMA GARCH had the least RMSE (7.5802). Other studies that are comparable to ours are the studies conducted by Kumar et al. (2018) where of all the models that were fit, hybrid SARIMA GARCH model appeared to be superior to them in terms of performance.

Other studies by Pandey et al. (2019) on improving rainfall prediction accuracy used data that was highly skewed and therefore this necessitated transformation. In our work, the data was mildly skewed and we worked with it in its original form (except for the differencing) and still got better results than with the transformed data. While data transformation seems to be a very effective way to reduce variance and improve the SARIMA model's execution, in some special circumstances it may not be possible. In our work, attempts to transform data from Gulf yielded a poor SARIMA model contrary to what was found in similar studies like that undertaken by Pandey et al. (2019) where the findings suggested that the performance of SARIMA models can be enhanced by using appropriate transformation (Box-Cox) along with GARCH model of residuals of highly skewed rainfall time series from two climatic environments that were under study. More specifically, it was observed that the two ports fitted unique models that is a SARIMA model for the gulf port, and Hybrid SARIMA GARCH model for the Yuzhny port a clear indication that the port through which fertilizer is imported has an effect on the price that is established since each port generated a unique model. Pandey et al. (2019) found that for Agartala, a station of monthly rainfall time series, SARIMA (0,1,1)(0,1,1)12-GARCH(1,2) with coefficient of determination $R^2 = 0.72$ and RMSE = 25.22. Following the transformation, the most optimal model was determined to be SARIMA(0,1,1)(0,1,1)12-GARCH(2,4), exhibiting an R^2 value of 0.87 and an RMSE of 0.672. Specifically for the monthly rainfall at Jodhpur station, the preferred model was SARIMA(0,1,1)(0,1,1)12-GARCH(1,2), with an R^2 value of 0.68 and an RMSE of 16.75. However, upon applying the Box-Cox transformation, the model that yielded the best results became SARIMA(0,1,1)(0,1,1)12-GARCH(1,2), with an improved R^2 value of 0.79 and an RMSE of 1.917. These performance indices demonstrate that the hybrid SARIMA-GARCH models, fitted to the transformed time series rainfall data, outperformed other models in both humid and arid regions. Consequently, it can be concluded that Hybrid SARIMA-GARCH models remain superior and have the potential for further enhancement through appropriate data transformations, following careful evaluation.

In this study, the Hybrid SARIMA GARCH model, as opposed to the SARIMA model, was shown to better match the data from the Yuzhny port while the SARIMA method matched the data from the port of Gulf best as it recorded the least AIC(7.4389) as compared to the GARCH(10.360) and SARIMA model(1008.86) for the port of Yuzhny. This can be attributed to the fact that the volatility of conditional heteroscedasticity was accounted for in the work. The findings of our study are consistent with the previous researches conducted that hybrid models can be used to improve forecasting accuracy.

According to this study, nonlinear variance features have distinct effects on the price series for fertilizer at the ports of Yuzhny and the Arab Gulf. Even though the SARIMA model appears to be suitable for the mean and seasonal behavior of fertilizer time series as seen at Gulf port, the current analysis demonstrates that it cannot account for seasonal heteroscedasticity. We discovered that the SARIMA model by itself was unable to adequately represent the seasonal change of variance at Yuzhny port. Furthermore, it was demonstrated that both seasonal and non-seasonal differencing failed to eradicate heteroscedasticity from the Yuzhny port data. We see a better model when we fit the GARCH model to the residuals of the because GARCH models are better at capturing volatility.When we apply the GARCH model to the generated SARIMA residuals for the port of Yuzhny, we see an improved model. This is the case because, as shown in Table 4.10, the GARCH model records a lower AIC value than the SARIMA model alone. It is true that every port is different in terms of the numerous aspects that affect its operations, and a thorough study of these characteristics may help develop focused strategies, which in turn may help increase the use of fertilizer in Sub-Saharan nations, particularly Kenya.

5.3 Conclusions

When heteroscedasticity is taken into consideration, the hybrid SARIMA-GARCH model performed better than the SARIMA model alone in Yuzhny port while SARIMA fits the gulf data best. In contrast to Gulf port, which did not suit the hybrid SARIMA GARCH model well, this port's time series generation was relatively well captured by the SARIMA model alone. We conclude that Hybrid models are superior to stand alone models and stand high chances of yielding better and reliable models.

We also conclude that the port through which fertilizer is imported has an effect on the price that is established since each port generated a unique model. Each port is distinct, and many causes have varying effects on them. It therefore proved challenging to recommend a specific port to be used at a given time because of the varying effects.

5.4 Recommendations

It proved difficult to determine the best port, but armed with this knowledge, we advise that this study be expanded in the future to include more factors and become more comprehensive, in order to enable fertilizer dealers to better understand the best port through which fertilizer can be imported so that they are able to set competitive prices for their products.

In terms of policy making, Morris (2007) established that sub-Saharan African countries, which are generally characterized by low agricultural productivity, have far too low fertilizer application rates, averaging 10 kg/ha of nutrients of arable land, compared to 86 in South

Asia, 118 in Latin America, 198 in an average middle-income country, and 288 kg/ha in a high income country, this work will serve to guide them on the policy frameworks to develop for the various ports to make the ports more attractive to importers since we have established that indeed the port influences the price imposed. The government of Kenya could also collaborate with stakeholders to enable local production as this could boost the country's economy and reduce expenditure in the long run.

5.5 Limitations of the Study

Limited data was a problem when doing this research, particularly when the theoretical approaches that were supposed to alter the data failed to improve it. The future research might employ more volume and variables to make the model more complete and dependable. Although transformations seem to be a very effective way to reduce variance and improve the SARIMA model's execution, in some special circumstances, depending on the data, the requirement for transformation needs to be carefully assessed in order to create models that are reliable.

References

Ait Sidhoum, A. and Serra, T. (2016). Volatility spillovers in the spanish food marketing chain: the case of tomato. *Agribusiness*, 32(1):45–63.

Bhardwaj, S., Paul, R. K., Singh, D., and Singh, K. (2014). An empirical investigation of arima and garch models in agricultural price forecasting. *Economic Affairs*, 59(3):415.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bonilla Cedrez, C., Chamberlin, J., Guo, Z., and Hijmans, R. J. (2020). Spatial variation in fertilizer prices in sub-saharan africa. *PloS One*, 15(1):e0227764.

Burns, P. (2002). Robustness of the ljung-box test and its rank equivalent. *Available at SSRN 443560*.

Chena, P.-Y., Changb, C.-L., Chena, C.-C., and McAleerc, M. (2012). Modelling the effects of oil prices on global fertilizer prices and volatility. *Journal of Risk and Financial Management*, 5(1):78–114.

Demir, S. (2022). Comparison of normality tests in terms of sample sizes under different skewness and kurtosis coefficients. *International Journal of Assessment Tools in Education*, 9(2):397–409.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.

Etienne, X. L., Trujillo-Barrera, A., and Wiggins, S. (2016). Price and volatility transmissions between natural gas, fertilizer, and corn markets. *Agricultural Finance Review*.

Galbraith, C. (2010). An examination of factors influencing fertilizer price adjustment. Technical report, Denver, Colorado.

Gardebroek, C. and Hernandez, M. A. (2013). Do energy prices stimulate food price volatility? examining volatility transmission between us oil, ethanol and corn markets. *Energy Economics*, 40:119–129.

Hernandez, M. A. and Torero, M. (2013). Market concentration and pricing behavior in the fertilizer industry: a global approach. *Agricultural Economics*, 44(6):723–734.

Huang, W.-Y., McBride, W., and Vasavada, U. (2009). Recent volatility in us fertilizer prices: causes and consequences. Technical report, Research in Agricultural and Applied Economics.

Kaur, P., Singla, R., et al. (2022). Modelling and forecasting nifty 50 using hybrid arima" garch model. *The Review of Finance and Banking*, 14(1):7–20.

Kelly, V., Crawford, E., and Ricker-Gilbert, J. (2019). The new generation of african fertilizer subsidies: panacea or pandora's box? *Gates Open Research*, 3(214):214.

Kim, S.-W. and Brorsen, B. W. (2017). Forecasting urea prices. *Applied Economics*, 49(49):4970–4981.

Kokkinen, A. and Wouters, H. (2016). Euro area and european union gdp flash estimates at 30 days. Technical report, Eurostat.

Kumar, V., Singh, N., Singh, D. K., and Mohanty, S. (2018). Short-term electricity price forecasting using hybrid sarima and gjr-garch model. In *Networking Communication and Data Knowledge Engineering: Volume 1*, pages 299–310. Springer.

Mishra, P., Sahu, P., Uday, J., et al. (2014). Arima modeling technique in analyzing and forecasting fertilizer statistics in india. *Trends in Biosciences Journal*, 7(2):170–176.

Modarres, R. and Ouarda, T. B. (2012). Generalized autoregressive conditional heteroscedasticity modelling of hydrologic time series. *Hydrological Processes*, 27.

Morris, M. L. (2007). *Fertilizer use in African agriculture: Lessons learned and good practice guidelines*. World Bank Publications.

Mtaita, T. A. (2003). Chapter 3 - food. In HAZELTINE, B. and BULL, C., editors, *Field Guide to Appropriate Technology*, pages 277–480. Academic Press, San Diego.

Newton, M. J. (2019). *Develop a fertilizer price forecasting model to assist with farm management decisions*. PhD thesis.

Ott, H. (2012). Fertilizer markets and their interplay with commodity and food prices. *Report for the European Commision Join Research Centre, Brussels*.

Pandey, P., Tripura, H., and Pandey, V. (2019). Improving prediction accuracy of rainfall time series by hybrid sarima–garch modeling. *Natural Resources Research*, 28(3):1125–1138.

Sabu, K. M. and Kumar, T. M. (2020). Predictive analytics in agriculture: Forecasting prices of arecanuts in kerala. *Procedia Computer Science*, 171:699–708. Third International Conference on Computing and Network Communications (CoCoNet'19).

Sanyal, P., Malczynski, L. A., and Kaplan, P. (2015). Impact of energy price variability on global fertilizer price: application of alternative volatility models. *Sustainable Agriculture Research*, 4(526-2016-37964).

Schnitkey, G. D. (2016). Anhydrous ammonia, corn, and natural gas prices over time. *Farmdoc Daily*, 6.

Shetty, D. K., Summithra, and Ismael (2018). Hybrid sarima-garch model for forecasting indian gold price. *International Journal of Multidisciplinary*, 3:263–269.

Shui, J., Song, L., Zhou, L., Wang, T., and Xiong, J. (2022). Prediction of fertilizer price based on bidirectional lstm. In *5th International Conference on Computer Information Science and Application Technology (CISAT 2022)*, volume 12451, pages 140–146. SPIE.

Uçak, H., Ari, Y., and Yelgen, E. (2022). The volatility connectedness among fertilisers and agricultural crop prices: Evidence from selected main agricultural products. *Agricultural Economics*, 68(9):348–360.

Voogt, W. and Bar-Yosef, B. (2019). Chapter 10 - water and nutrient management and crops response to nutrient solution recycling in soilless growing systems in greenhouses. In Raviv, M., Lieth, J. H., and Bar-Tal, A., editors, *Soilless Culture (Second Edition)*, pages 425–507. Elsevier, Boston, second edition edition.

Vroomen, H. (1991). A short-run forecasting model for nitrogen fertilizer prices. *Agribusiness*, 7(6):515–522.

Xu, C., Chen, X., and Zhang, L. (2021). Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models. *Journal of Environmental Management*, 295:113085.

Appendix A

Additional results

A.1 Raw Time Series ACF and PACF plots

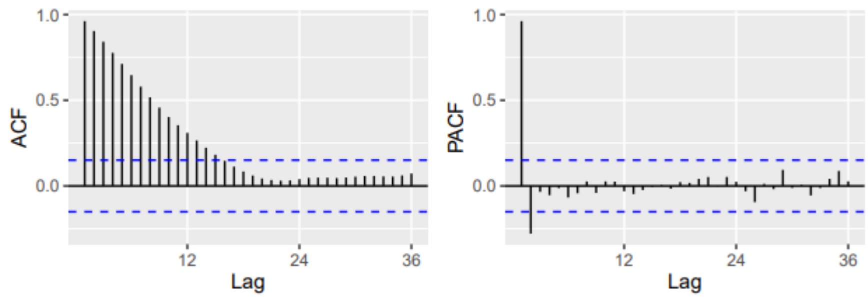


Figure A.1: Raw time series ACF and PACF plot Yuzhny

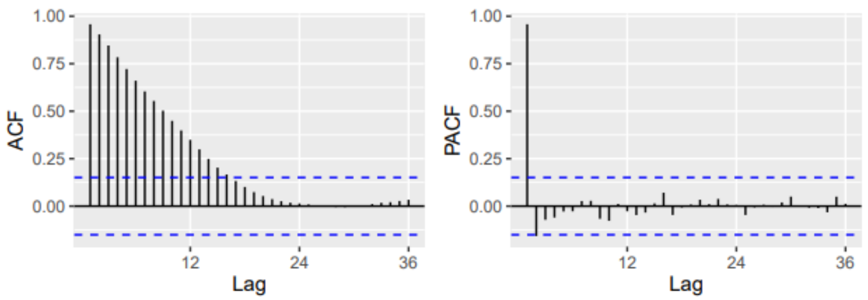


Figure A.2: Raw time series ACF and PACF plot Gulf

A.2 Decomposed Time Series Plots

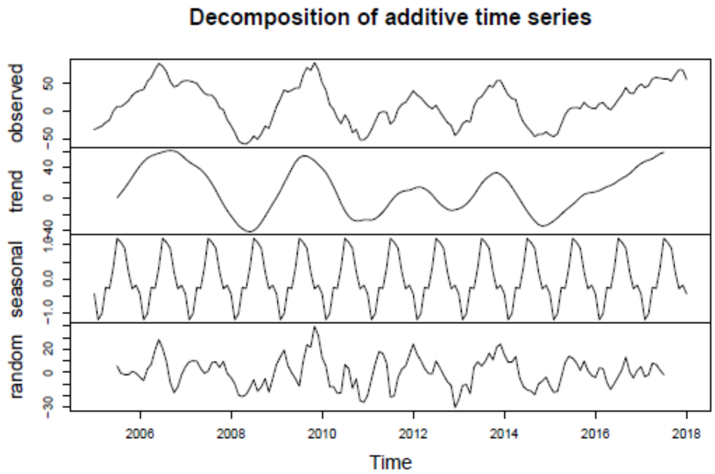


Figure A.3: Differenced time series plot Yuzhny

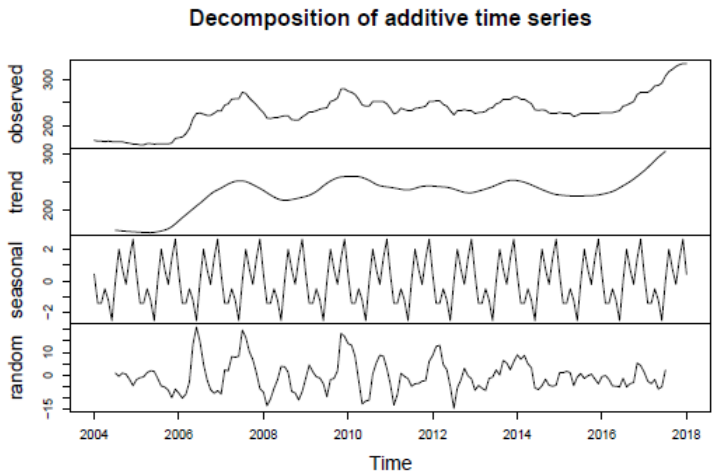


Figure A.4: Differenced time series plot Gulf

A.3 Residual Diagnostics

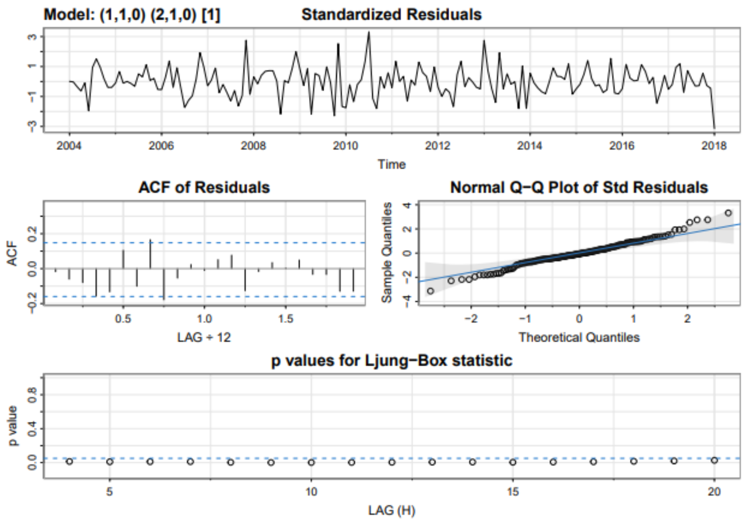


Figure A.5: Yuzhny diagnostics

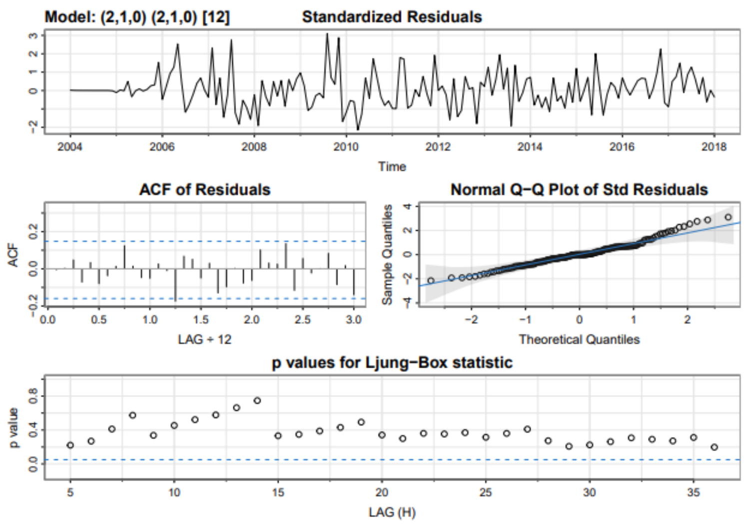


Figure A.6: Gulf diagnostics

A.4 Residual Diagnostics

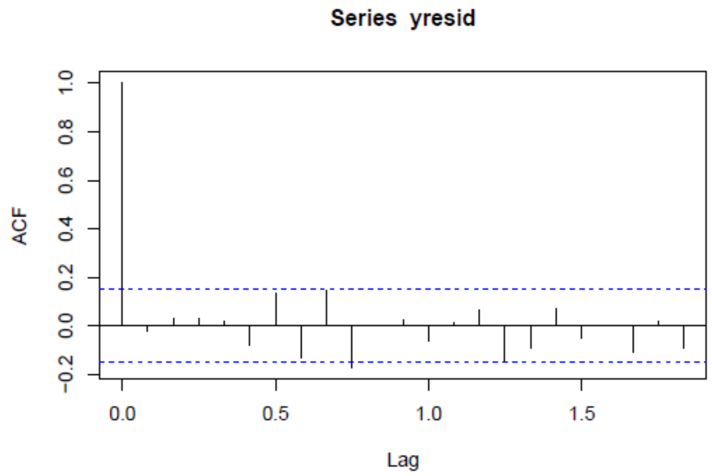


Figure A.7: ACF of yuzhny SARIMA residuals

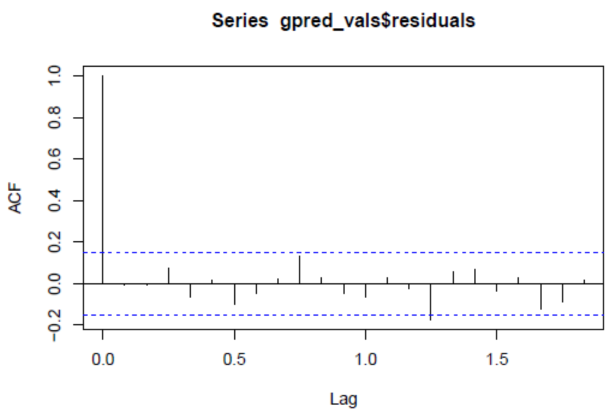


Figure A.8: ACF of gulf SARIMA residuals

A.5 Histograms for squared residuals

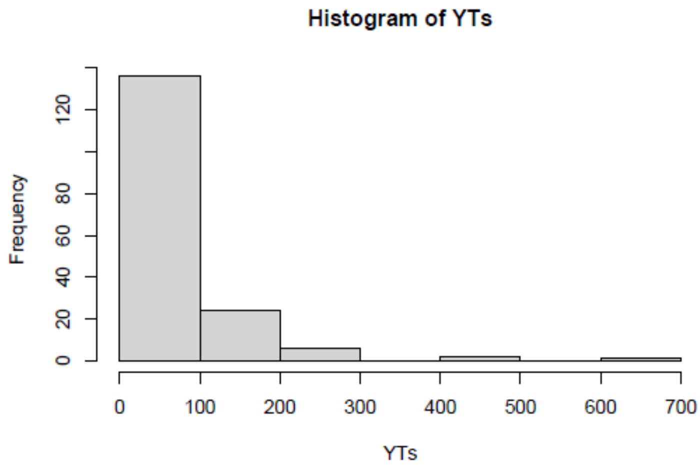


Figure A.9: Histogram of squared residuals Yuzhny

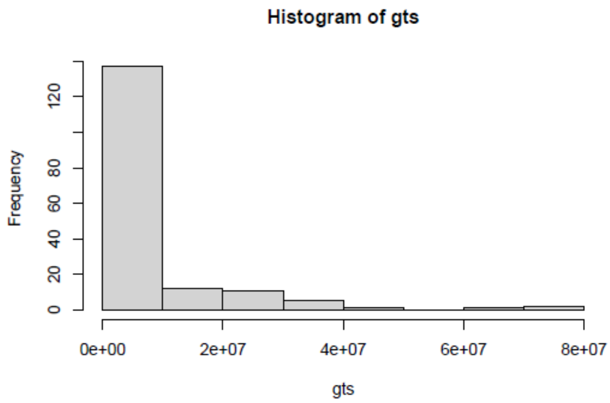


Figure A.10: Histogram of squared residuals gulf

A.6 ACF plots for GARCH model residuals

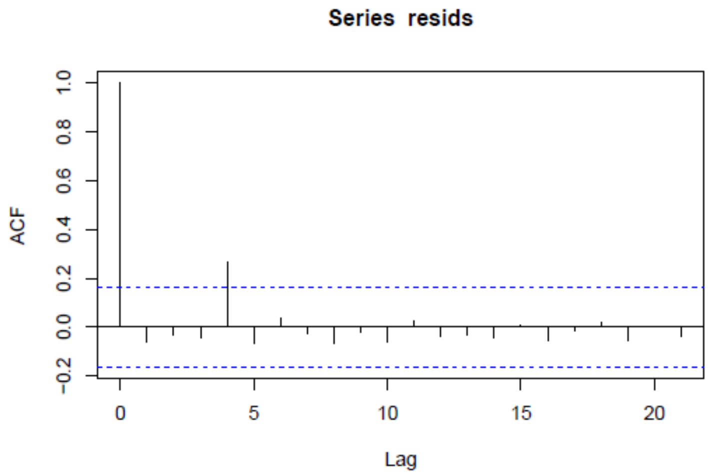


Figure A.11: ACF plot GARCH residuals

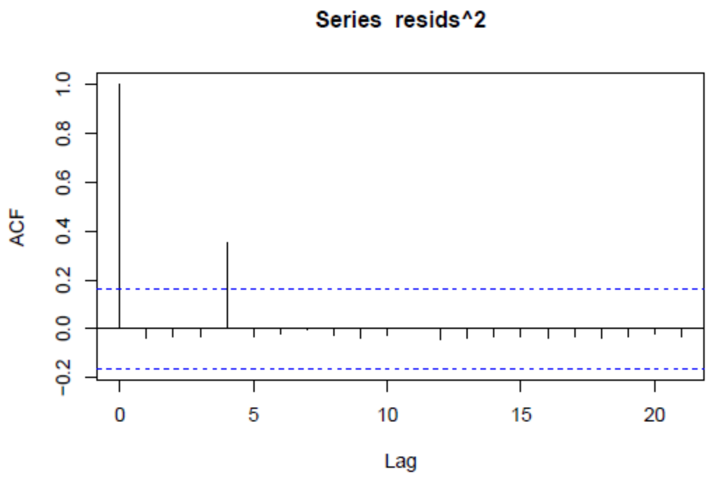


Figure A.12: ACF plot GARCH squared residuals

A.7 Residual Diagnostics

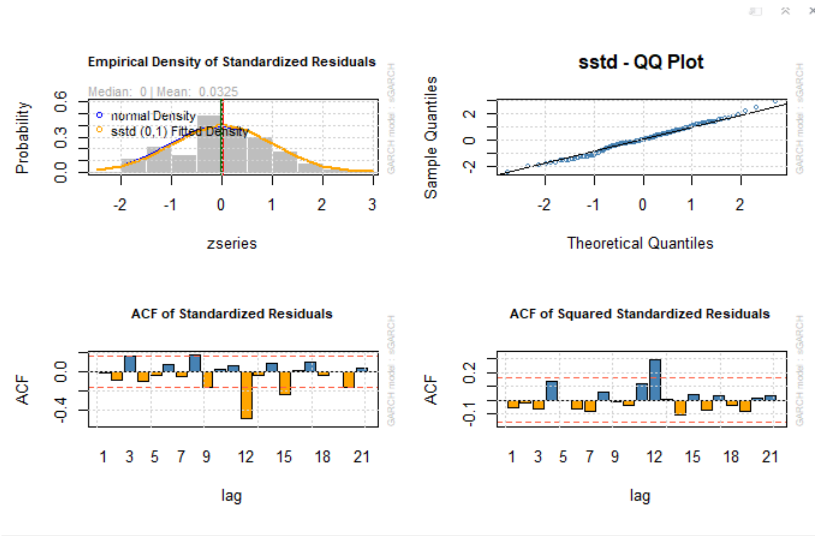


Figure A.13: Hybrid diagnostics

Appendix B

Ethical approval

Strathmore
UNIVERSITY

27th March 2023

Ms Okello Elizabeth,
elizabeth.okello@strathmore.edu

Dear Ms Okello,

RE: Application of Hybrid Seasonal ARIMA-GARCH Model in Modelling and Forecasting Fertilizer Prices in Kenya

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU- master's** research proposal. Your application reference number is **SU-ISERC1611/23.** The approval period is from **27th March 2023 to 26th March 2024.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, and MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 48 hours of notification
iv. Any changes, anticipated or otherwise, that may increase the risks or affect the safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 48 hours
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-ISERC

Cc: Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C

Similarity index

Ouriginal
by Turnitin

Document Information

Analyzed document	113285.pdf (D158927147)
Submitted	2023-02-17 18:25:00
Submitted by	
Submitter email	Elizabeth.Okello@strathmore.edu
Similarity	4%
Analysis address	library.strath@analysis.urkund.com

Sources included in the report

SA **MScThesis.docx**
Document MScThesis.docx (D113866033)

SA **Plagrism 1 st doc.docx**
Document Plagrism 1 st doc.docx (D157432877)

SA **Goswami Viniya.pdf**
Document Goswami Viniya.pdf (D110761392)

SA **GARCH.docx**
Document GARCH.docx (D147761211)

SA **Anushaka_Thesis.docx**
Document Anushaka_Thesis.docx (D113883253)

SA **S. POYYAMOZHI - Mathematics - Final Thesis.pdf**
Document S. POYYAMOZHI - Mathematics - Final Thesis.pdf (D141421335)

SA **THESIS FINAL.pdf**
Document THESIS FINAL.pdf (D55409630)

Appendix D

R code

```
profercy<-read.csv("C:/Users/spectre/OneDrive/Documents/time series
content/profercy.csv",header=T,skip=3)
profercy[,c("Week.No.","X")]<-list(NULL)#to drop the empty columns
head(profercy)#To see the first top rows
yuz<-profercy.complete[,c("Month.Year","P.UREA.Yuzhny.FOB")]
yuzts<-ts(yuz[,2],start=c(2004,1),end=2018,frequency = 12)#TO write into a ts
object
print(kurtosis(yuzts))
print(skewness(yuzts))
sd(yuzts)
shapiro.test(yuzts)#To test for normality
adf.test(yuzts, k=2)#we use the adf test to check for stationarity
yuz_diff12 = diff(yuzts,12)#difference to make it stationary and assign to
variable name
par(mfrow=c(1,2))
acf(yuz_diff12, 48)#To plot the acf of seasonal differenced data
pacf(yuz_diff12, 48)#To plot the pacf of seasonal differenced data
myts.train <- window(diff1and12, end=c(2016,12))
myts.test <- window(diff1and12, start=2017)
yuz1_auto<-auto.arima(myts.train,
          stepwise = F,
          approximation = F,
          seasonal = T,
          allowdrift = F
          )#Use auto arima to automatically select the best model
summary(yuz1_auto)
fit1_yuz<-sarima(yuzts, 1,1,0,2,1,0,S=1,details = T,Model = T)
autoplot(residuals(yuz1_auto)) # plot the residuals
yuz1_auto_residuals <- residuals(yuz1_auto)
ArchTest(yuz1_auto_residuals)
ArchTest(yuz1_auto_residuals^2)
re<-yuz1_auto_residuals^2
yts<-ts(re)#write to a ts object and assign to a variable name
ytss<-as.numeric(re)
YTs<-ts(ytss)
s<-rep(1:169)
plot.ts(YTs)#To plot the ts object
hist(YTs)#To check for the distributionof the skewness
hist(yuz1_auto_residuals)
adf.test(YTs)#To test for stationarity of the residuals
#plot(abs(YTs))
qqnorm(YTs, main = "Squared resids -QQ Plot", col = "blue")
qqline(YTs)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0), include.mean = T),
           variance.model=list(model="eGARCH",garchOrder= c(1,1)),
          distribution.model = "sstd",)
yfit <- ugarchfit(spec, data=YTs,solver = "hybrid")
show(yfit)
myfit<-ugarchforecast(yfit, n.head=20)
show(myfit)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0), include.mean = T),
           variance.model=list(model="eGARCH",garchOrder= c(1,1)),
          distribution.model = "snorm",)
yfit1 <- ugarchfit(spec, data=YTs,solver = "hybrid")
show(yfit1)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0), include.mean = T),
           variance.model=list(model="eGARCH",garchOrder= c(2,0)),
          distribution.model = "snorm",)
yfit7 <- ugarchfit(spec, data=YTs,solver = "hybrid")
show(yfit7)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0), include.mean = T),
           variance.model=list(model="gjrGARCH",garchOrder= c(1,0)),
          distribution.model = "sstd",)
yfit_2 <- ugarchfit(spec, YTs,solver = "hybrid")
show(yfit_2)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0), include.mean = T),
yfit_8 <- ugarchfit(spec, YTs,solver = "hybrid")
show(yfit_8)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0),
include.mean = T),

variance.model=list(model="gjrGARCH",garchOrder=
c(1,1)),
          distribution.model = "snorm",)
yfit_3 <- ugarchfit(spec, YTs,solver = "hybrid")
show(yfit_3)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0),
include.mean = T),

variance.model=list(model="sGARCH",garchOrder=
c(1,1)),
          distribution.model = "snorm",)
yfit_5 <- ugarchfit(spec, YTs,solver = "hybrid")
show(yfit_5)
spec <- ugarchspec(mean.model = list(armaOrder = c(0,0),
include.mean = T),

variance.model=list(model="sGARCH",garchOrder=
c(1,1)),
          distribution.model = "sstd",)
yfit_4 <- ugarchfit(spec, YTs,solver = "hybrid")
show(yfit_4)#To showthe plots
infocriteria(yfit_4)
resids<-residuals(yfit_4)#To assign the residuals of the
yfit_4 to a variable name
acf(resids)
acf(resids^2)
#yfit_4@fit
#yfit_4@model
cond_mean <- ts(as.numeric(fitted(yfit_4)))
cond_var <- ts(as.numeric(sigma(yfit_4)^2))
yfore<- ugarchforecast(yfit_4,n.ahead = 12)
yfore
yfit4forecast<-ugarchforecast(yfit_4, n.ahead = 12)
yfit4forecast
specs <-
  ugarchspec(
    variance.model = list(
      model = "sGARCH",
      garchOrder = c(1, 1),
      submodel = NULL,
      external.regressors = NULL,
      variance.targeting = FALSE ),
    mean.model = list(armaOrder = c(1,1,0,2,1,0),
include.mean = FALSE),
    distribution.model = "sstd"
fit <- ugarchfit(specs, myts.train,
          solver = "hybrid")
fit_forecast <- ugarchforecast(fit, n.ahead =
length(myts.test))
point_forecast <- fit_forecast@forecast$seriesFor
lower_pi <- fit_forecast@forecast$lower[,1]
upper_pi <- fit_forecast@forecast$upper[,1]
fit_rmse <- sqrt(mean((myts.test - point_forecast)^2))
mae <- mean(abs(myts.test-point_forecast))
mean_abs_error <- mean(abs(myts.test - point_forecast))
mean_abs_deviation <- mean(abs(diff(myts.test)))
mase <- mean_abs_error/mean_abs_deviation
gulf<-
profercy.complete[,c("Month.Year","P.UREA.Arab.Gulf.FOB
")]
summary(gulf)
```

```
        variance.model=list(model="gjrGARCH",garchOrder= c(1,1)),
       distribution.model = "sstd",)

IQR <- IQR(gulf_overall_outlier$P.UREA.Arab.Gulf.FOB)
no_outliers <- subset(gulf_overall_outlier,
gulf_overall_outlier$P.UREA.Arab.Gulf.FOB > (Q1 - 1.5*IQR) &
gulf_overall_outlier$P.UREA.Arab.Gulf.FOB < (Q3 + 1.5*IQR))
dim(no_outliers)
View(no_outliers)
View(gulf_overall_outlier$P.UREA.Arab.Gulf.FOB)
cleangulfts<-
ts(gulf_overall_outlier$P.UREA.Arab.Gulf.FOB,start=c(2004,1),end=2018,freque
ncy = 12)
shapiro.test(cleangulfts)
skewness(cleangulfts)
describe(sqrt(gulfts))
#head(guroot)
describe(log(gulfts))
describe(sin(gulfts))
describe(1/gulfts)
describe(gulfts^4)
adf.test(gulfts, k=2)#we use the adf test to check for stationarity
shapiro.test(gulfts^2)
gudiff12 = diff(gulfts,12)
gdiff12 = diff(gulfts^2,12)#  difference to make the data stationary
acf2(gdiff12, 48)
acf(gudiff12, 48)
gudiff1and12 = diff(gudiff12, 1)
plot(gudiff1and12)
gdiff1and12 = diff(gdiff12, 1)
plot(gdiff1and12)
par(mfrow=c(1,2))
acf(gdiff1and12, 48)
pacf(gdiff1and12, 48)
adf.test(gdiff1and12)
kpss.test(gdiff1and12)
pp.test(gdiff1and12)
gugulf_auto<-auto.arima(gudiff1and12,
          stepwise = F,
          approximation = F,
          seasonal = T,
          allowdrift = F )
summary(gugulf_auto)
gure<-residuals(gulf_auto)
Box.test(gure)
ArchTest(gure)
ArchTest(gure^2)
autoplot(residuals(gulf_auto)) # plot the residuals
#plot(gulf_auto) # inspect the roots
ggAcf(residuals(gulf_auto),lag.max=36) # inspect the ACF
ggPacf(residuals(gulf_auto),lag.max=36) # inspect the PACF
```

```
gulfts<-ts(gulf[,2],start=c(2004,1),end=2018,frequency =
12)
plot(gulfts,xlab="Years",ylab="price")#to plot the ts object
ggtsdisplay(gulfts)
dec<-decompose(gulfts)#to decompose the data
plot(dec)
shapiro.test(gulfts)#To test for normality
print(kurtosis(gulfts))
print(skewness(gulfts))
print(sd(gulfts))
gulf<-
profercy.complete[,c("Month.Year","P.UREA.Arab.Gulf.FOB
")]
gulf_overall_outlier <- gulf
dim(gulf_overall_outlier)
Q1 <- quantile(gulf_overall_outlier$P.UREA.Arab.Gulf.FOB,
.25)
Q3 <- quantile(gulf_overall_outlier$P.UREA.Arab.Gulf.FOB,
.75)
```