# DETERMINATION OF SPOT WHEAT PRICES UNDER CLIMATE IMPACT USING COPULA APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İLAYDA KAYAPINAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ACTUARIAL SCIENCES

JULY 2022

Approval of the thesis:

**DETERMINATION OF SPOT WHEAT PRICES UNDER CLIMATE IMPACT USING COPULA APPROACH**

submitted by **İLAYDA KAYAPINAR** in partial fulfillment of the requirements for the degree of **Master of Science in Actuarial Sciences Department, Middle East Technical University** by,

Prof. Dr. A. Sevtap Kestel
Dean, Graduate School of **Applied Mathematics** ________________

Prof. Dr. A. Sevtap Kestel
Head of Department, **Actuarial Sciences** ________________

Prof. Dr. A. Sevtap Kestel
Supervisor, **Actuarial Sciences, METU** ________________

Dr. Bükre Yıldırım Külekci
Co-supervisor, **Mathematics, Technical University of Kaiserslautern** ________________

**Examining Committee Members:**

Prof. Dr. Ceylan Yozgatlıgil
Statistics Department, METU ________________

Prof. Dr. A. Sevtap Kestel
Actuarial Sciences Department, METU ________________

Assist. Prof. Dr. Uğur Karabey
Actuarial Sciences Department, Hacettepe University ________________

**Date:** ________________

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last Name: İLAYDA KAYAPINAR

Signature :

# ABSTRACT

DETERMINATION OF SPOT WHEAT PRICES UNDER CLIMATE IMPACT
USING COPULA APPROACH

Kayapınar, İlayda
M.S., Department of Actuarial Sciences
Supervisor : Prof. Dr. A. Sevtap Kestel
Co-Supervisor : Dr. Bükre Yıldırım Külekci

July 2022, 52 pages

Climate components have a significant impact on the supply of agricultural goods in two ways. Firstly, the climate condition influences the efficiency of agriculture and the volume of the harvested product. Secondly, the farmers harvesting their products would prefer to sell their goods in the season when the supply is limited at a higher price. Therefore, we can say that climate conditions can determine the price of agricultural products. Modeling the seasonal variables whose distributions are not the same and analyzing the dependence between agricultural products have great interest and importance in agricultural markets and risk theory.

The primary motivation is to forecast the spot wheat prices under the influence of the climate component. For this purpose, we employ time series analysis for Konya's monthly adjusted weighted average prices of wheat transactions, whose clearing is conducted together with Istanbul Settlement and Custody Bank Inc., and climate components. Afterward, the adjusted spot prices against inflation is remodeled by using t-copula under the influence of climatic parameters to improve the predictions. For this purpose, the best models are selected for the temperature, relative humidity, and precipitation, and the residuals derived from those models are used to determine the vine structure. Vine trees help us understand if there is a core climate component with the dependence structure with other variables. Then, the adjusted spot wheat prices against inflation are simulated with respect to the output of the vine copula structure.

The simulated adjusted wheat prices with t-copula give us a more accurate estimation than the predictions from the time-series analysis.

Keywords: ARIMA, SARIMA, Dependence, Vine copula

# ÖZ

COPULA YAKLAŞIMI İLE İKLİM ETKİSİ ALTINDA
SPOT BUĞDAY FİYATLARININ BELİRLENMESİ

Kayapınar, İlayda
Yüksek Lisans, Aktüerya Bilimleri Bölümü
Tez Yöneticisi : Prof. Dr. A. Sevtap Kestel
Ortak Tez Yöneticisi : Dr. Bükre Yıldırım Külekci

Temmuz 2022, 52 sayfa

İklim bileşenleri, tarımsal ürünlerin arzı üzerinde iki şekilde önemli bir etkiye sahiptir. İlk olarak, iklim koşulu tarımın verimliliğini ve hasat edilen ürünün hacmini etkiler. İkinci olarak, ürünlerini hasat eden çiftçiler, arzın sınırlı olduğu sezonda ürünlerini daha yüksek fiyattan satmayı tercih edeceklerdir. Dolayısıyla iklim koşullarının tarım ürünlerinin fiyatını belirlediği söylenebilir. Dağılımları aynı olmayan mevsimsel değişkenlerin modellenmesi ve tarımsal ürünler arasındaki bağımlılığın analiz edilmesi, tarım piyasaları ve risk teorisinde büyük ilgi ve öneme sahiptir.

Bu çalışmadaki ana motivasyon, iklim bileşeninin etkisi altında spot buğday fiyatlarını tahmin etmektir. Bu amaçla, İstanbul Takas ve Saklama Bankası A.Ş. ile takası yapılan buğday işlemlerine ait Konya'nın aylık düzeltilmiş ağırlıklı ortalama fiyatları ve iklim bileşenleri için zaman serisi analizi yapılmıştır. Daha sonra, tahminleri iyileştirmek için iklim parametrelerinin etkisi altında t-copula kullanılarak enflasyona karşı düzeltilmiş spot fiyatlar yeniden şekillendirilmektedir. Sıcaklık, bağıl nem ve yağış için en iyi modeller seçilir ve bu modellerden elde edilen artıklar asma yapısını belirlemek için kullanılır. Asma ağaçları, diğer değişkenlerle bağımlılık yapısına sahip çekirdek bir iklim bileşeni olup olmadığını anlamamıza yardımcı olur. Sonrasında, enflasyona karşı düzeltilmiş spot buğday fiyatları, asma yapısının çıktısına göre simüle edilir. Sonuç olarak, t-kopula ile simüle edilmiş düzeltilmiş buğday fiyatlarının bize zaman serisi analizindeki tahminlerden daha doğru bir tahmin verdiğini görüyoruz.

Anahtar Kelimeler: ARIMA, SARIMA, Bağımlılık, Asma kopula

*To My Family*

# ACKNOWLEDGMENTS

I would like to express my very great appreciation to my thesis supervisor, Prof. Dr. Ayşe Sevtap Selçuk Kestel, and Dr. Bükre Yıldırım Külekci for their patient guidance, enthusiastic encouragement, and valuable advice during the development and preparation of this thesis. I would like to thank them for my thesis process and every single day of the last three years.

Secondly, I would like to thank the Ministry of Agriculture for enabling us the datasets.

Also, I am thankful to my friends Berkay Selçuk, Cansu Sert, Hilal Ersoy, Burcu Ecem Karakaş and Cem Yavrum for their motivation and helps.

Last but most, I would like to thank my dear parents, Aybüke, Barış and Gümüş for their limitless support in my life.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ARIMA | Autoregressive Integrated Moving Average |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| SES | Simple Exponential Smoothing |
| DES | Double Exponential Smoothing |
| TES | Triple Exponential Smoothing |
| TMO | Turkish Grain Board (Toprak Mahsulleri Ofisi) |
| CME Group | Chicago Mercantile Exchange, Chicago Board of Trade, New York Mercantile Exchange, The Commodity Exchange |
| FAO | Food and Agriculture Organization |
| JB | Jarque Bera |
| EWR | Electronic Warehouse Receipt |
| AIC | Akaike Information Criteria |
| BIC | Bayesian Information Criterion |
| EBIC | Extended Bayesian Information Criterion |
| DIC | Deviance Information Criteria |
| ANFIS | Adaptive Network-Based Fuzzy Inference System |
| ME | Mean Error |
| MAE | Mean Absolute Error |
| MSE | Mean Square Error |
| MAPE | Mean Absolute Percentage Error |
| RMSE | Root Mean Square Error |
| GARCH | Generalized Auto-Regressive Conditional Heteroskedasticity |
| MSOA | Multi-stage Optimization Approach |
| ACF | Autocorrelation Function |
| PACF | Partial Autocorrelation Function |

# CHAPTER 1

# INTRODUCTION

Agricultural products are farmed livestock and harvested crops to provide food generally. The future of agricultural products is one of the most critical and discussed topics of the $21^{st}$ century because of the increasing world population, recorded high inflation in several countries, and severe climate change, which affects supply and demand relation and price formation of agricultural commodities directly. Increased demand with limited supply causes an escalation in prices or vice versa. Several reasons have an impact on the supply and demand structure of agricultural commodities. Apart from the impact of changing population, political and economic crises, and changing preferences and government regulations, agricultural product prices are affected by climate-related factors, which influence supply.

Wheat is one of the most strategic products. It is used for the production of necessary goods like bread. Since it is not a luxury good, people from different income segments are able to consume it. While the supply reaches its highest point during the harvest season, wheat is consumed in four seasons since it is a necessity good.

Agricultural products can also be converted into financial contracts and traded in financial markets. These contracts are traded in the spot or futures markets in the world. For example, many contracts in the Chicago Mercantile Exchange (CME) Group are traded in the futures markets. In Türkiye, on the other hand, local exchanges, which are the institutions that provide intermediary services for buying and selling electronic receipts whose underlying assets are agricultural commodities, were operating until the foundation of the Turkish Mercantile Exchange. The Electronic Warehouse Receipts (EWR) are the financial contracts in which underlying assets are the products stored in the licensed warehouses that the Ministry of Trade in Türkiye regulates. EWR recorded to Central Registry System stands for the standardized commodities with the same quantity, species, and subspecies. These features make EWR a financial contract. Like any agricultural contract, any factor that increases the cost or price of the agricultural product can directly affect the contract's price. For example, with the conflict between Ukraine and Russia, the prices of American Wheat Futures Contracts have reached their highest level since mid-2008 because these two countries are the leaders in wheat exports. According to the Food and Agriculture Organization (FAO) report, if the conflicts continue, prices will increase due to the global supply gap [5] [4].

Seasonality is one of the critical factors for the price formation of agricultural commodities. During the harvest season, an increase in supply causes a decrease in prices, whereas, during the non-harvest season, the prices increase. Thus, a farmer who harvests the products from the soil can store their product to sell it in a non-harvest season at a higher price to benefit from the low supply.

Price formation is also affected by storage conditions. The supply cannot last long without advanced storage techniques if the product is not durable. For this reason, supply cannot be accessible for all seasons, and volatility in supply cause volatility in commodity prices. On the other hand, if one agricultural commodity can be stored for long periods, supply can be distributed all seasons. Storage can help reduce the impact of seasonality in pricing agricultural commodities. As wheat is a storable product, the pricing formation cannot be the same as other necessity products, which are unstorable.

In Türkiye, there are regulations for price formation similar to other countries. In the case of a sudden price increase, Turkish Grain Board (TGB) takes action by selling or buying products from the market to balance the prices. TGB takes similar action to what Central Banks do to control inflation. For instance, in the case of a bullish environment for a specific product, TGB announces to sell its stock to the market. This action increases the supply of the product in the market, so prices decrease. The buyers of the product sold by TGB cannot resell the product to the market to profit. Hence, this mechanism controls the prices. For instance, on the 26th of March, 2022, TGB announced that the local purchasing price for bread wheat is between 216.45-224.21 USD (3.210-3.325 TL) per ton, while the weighted average price is between 323.66-451.11 USD (4.800-6.690 TL), according to the bulletin of Konya Mercantile Exchange [1] [2].

## 1.1 Literature Survey

The models in literature taking time effect into account in climate variables are numerous. The composition of time series models in line with capturing the dependence structure using copula to forecast wheat spot prices in Türkiye is not countable. The remarkable literature supporting the choice of the models and implementation of the proposed is summarized in two aspects: time series and copula modeling.

Jain and Mallick [27] compare the performance of ARIMA and Exponential Smoothing (ETS) model, and forecast weather parameters such as rainfall, relative humidity, wind speed, and air temperature. Their method consists of two main steps. After understanding the outliers, they select the best model with respect to Akaike information Criteria (AIC) and Bayesian Information Criterion (BIC) to forecast the future values.

Tektaş [38] works on wind speed, average temperature, and air pressure datasets to forecast the weather of Göztepe, Istanbul. He employs ARIMA and Adaptive Network-Based Fuzzy Inference System (ANFIS) and compares the performance of these two models. According to his results, the ANFIS model is the best model for MAE, RMSE, and $R^2$ criteria.

In another study by Sarraf et al., monthly temperature and relative humidity datasets are used for time series analysis [32]. As a result of their research, ARIMA(0,0,1)(0,1,1) and ARIMA(0,0,1)(2,1,2) are selected as best models for monthly temperature and relative humidity datasets.

On the other hand, in the literature, there are many studies for forecasting production and the prices of agricultural commodities. Iqbal et al.[26] work on forecasting wheat cultivated area and the production amount for up to 2022 in Pakistan. They use the Box-Jenkins methodology for forecasting and conclude that ARIMA(1,1,1) and ARIMA(2,1,2) are the best models for forecasting wheat area and production amount, respectively. The estimated production amount is found as 29.774,8 thousand tons in 2022.

Paul [8] compared the performance of GARCH and ARIMA models for forecasting spot wheat prices of Gram in the Delhi Market in his work. According to his results, since the ARIMA model could not capture the volatility in price, the GARCH model is suggested as the best model.

In the study of Haofei [22], the food prices are forecasted for China using MSOA, ARIMA and the BP models. According to their evaluation, the MSOA model is the best in terms of accuracy.

Fattah et al. forecast a company's food demand with Box-Jenkins methodology. They select ARIMA(1, 0, 1) as the best model [17].

Copula models are used in many studies for understanding dependence relations between random variables. Thanks to its flexibility in modeling multivariate distributions, copula is used in either cross-sectional or time series analyses. Univariate marginals can be linked to the multivariate distribution with the help of the copula model.

Sklar [35] introduced to copulas first to understand the dependence structure across random variables at which n-univariate marginal distribution and n-dimensional copula can be generated by n-dimensional joint distribution.

Copulas are used in many areas to capture dependence structure. In Actuarial Science, the implementation varies from non-life to life insurance problems such as estimating joint-life mortality and competing risks-multiple decrement theory [18]. For instance, Evkaya et al. [16] constructed an index-based insurance design for Türkiye. They model the weather yield with weather index variables and calculate the premium and compensation using linear regression equations. Also, copula functions are used in many topics like bivariate option pricing, risk evaluation techniques, and diversifying and hedging assets in finance [11], [19].

Patton [30] reviews and discusses the literature on time series copula models. In this study, he recommends using copulas to catch the cross-sectional series dependent on the past. Also, Smith et al. [36] use copulas for continuous-valued time series to model serial dependence.

They use a D-vine copula for longitudinal data [37]. Also, they show that vine copula models could be used in either cross-sectional or time-series datasets.

Silva et al. estimate the parameters of bivariate copula functions and three marginal distributions with the Bayesian approach. After they compare the model selection criteria such as AIC, BIC, EBIC and DIC, they show that the Deviance Information Criteria (DIC) is useful[34].

Cong and Braddy [12] model the joint distribution of temperature and precipitation with copula. They compare the power of the different copula models. The conclusion is that student t-copula is the best copula for determining the interdependence relationship between precipitation and rainfall for Scania, Sweden. In another study, spatial dependence is investigated by using R-vine copula models for different locations. They use daily mean temperature data for 54 different locations and develop an R-vine model. They also evaluate the model performance between spatial R-vine copula and the Gaussian spatial model [15]. Also, copula analysis is widely used in empirical finance. For instance, Gronwald [20] studies the interdependence relationship between European carbon, commodity, and financial markets using different copula models. They found that student t-copula performs better than alternative copula models.

To the best of our knowledge, this thesis is the first study of estimating adjusted wheat spot prices under climate impact using copula approach in the local area of Türkiye. We show how the effect of the climatic conditions specific to each region or city affects the spot wheat prices in that region. The residuals from the suggested models are used to investigate the interdependence structure of the temperature, relative humidity, precipitation, and wheat prices. Then, the parameters obtained from the copula analysis are used to estimate wheat prices.

## 1.2 Aim of the Study

The main goal of this work is to emphasize and show the impact of local climate components on the local spot wheat prices. Konya is selected as a representative for this work due to its importance in the production in the Turkish wheat market. Konya, one of the biggest cities in terms of area with steppe climate conditions, is the leading city in terms of production and wheat cultivation. Even though Konya holds tremendous importance, there is not enough study investigating the impact of climate components on local spot wheat prices, so this study can fill this gap in the literature.

This work contributes to the literature on the investigation of the impact of the interdependence structure of climate components on the forecast of local spot wheat prices for Türkiye. Local agricultural production is directly dependent on the local climate factors, so the price is as well. The prices used in this work are the prices of EWR, which are defined as the standardized electronic receipts that are regulated by the clearing house of Türkiye. These features of EWR make it a financial contract. Because these contracts cover the transaction within Konya, it is also a local contract that contains the impact of the local supply and demand factors. Since the supply of agricultural goods is highly dependent on climatic factors, in this thesis, we show the impact of local climatic factors on the financial contract whose price structure depends on the local factors. At the beginning of this study, it is planned to conduct spatial analysis for this work to understand the neighborhood impact. Due to the difficulties in obtaining wheat price data sets for the neighboring cities, this implementation did not come true.

The temperature, relative humidity, and precipitation are the main climate components that affect wheat production, thus, its price. Extremely high temperatures and drought have significant impacts on the number of crops harvested. For these reasons, the leading three climate components are used for the analyses.

In the first part of the thesis, suitable models such as ARIMA, SARIMA, and Triple Exponential Smoothing (TES) are used to model temperature, relative humidity, precipitation, and spot wheat prices to forecast their future realizations. Afterward, the differences between the predicted and actual values, i.e., residuals for climate components, are drawn in vine structure in order to understand the dependence structure of these variables. In other words, vine trees help us to understand if there is a core climate component that affects other climate components or not. Then, bivariate t-copula is applied to the adjusted wheat prices and prescribed climate component, and simulations are made for predictions based on the best fitting copula density. Lastly, the predictions coming from the time series model and copula-based model are compared with respect to the mean absolute error value to see if the copula-based prediction, which contains the impact of climate component, optimized our predictions or not.

The implementation is made by temperature, rainfall, and relative humidity datasets obtained from the Ministry of Agriculture-Turkish State Meteorological Service. These datasets consist of monthly temperature values, Precipitation ($m^2$), and Relative Humidity (%) realizations of Konya. It covers the monthly average values between January 2007 and July 2021. Konya is chosen as a representative for this analysis because it is at the first rank in all cities regarding wheat production. In 2021, %9.4 percent of the total supply belonged to Konya [3]. Konya region is the leading supplier of wheat.

Spot wheat for Bread Contract prices is obtained from the website of Konya Mercantile Exchange. The spot wheat price is the weighted average price of the transactions executed and recorded by the Konya Mercantile Exchange, and it consists of monthly prices between 2007 and 2021. This dataset has monthly wheat prices in terms of the kilogram.

The outcomes of this study can be used for further studies on premium estimation of agricultural insurance for several products and also for other regions in Türkiye.

This thesis is organized as follows: Chapter 2 presents the preliminaries on methods implemented. In Chapter 3, we proceed with the time series analysis for climate components and wheat prices, and then we demonstrate the copula analysis and its evaluation of performance. Finally, Chapter 5 gives some concluding findings.

# CHAPTER 2

# PRELIMINARIES

## 2.1 Univariate Time Series Process

A univariate time series is the series of observations of a specific component in time order. The time series can show the increasing and decreasing trends and systematic fluctuation called seasonality or outliers.

Stationary series are the series that have the constant mean, variance, and autocorrelation over time. If the mean, variance, and autocorrelation of a series do not vary over time, the series is defined as stationary series.

The autoregressive model of order $p$, AR($p$), is the composed of the current value of a time series depending on the previous values and the error term. AR($p$) process is defined by:

$$X_t = a + \sum_{n=1}^{p} \varphi_i X_{t-i} + \epsilon_t \tag{2.1}$$

where $\varphi_1, \ldots, \varphi_p$ are the parameters, $a$ is a constant number and $\epsilon_t$ is the white noise error term which is independent and identically distributed.

In the Moving Average model of order $q$, MA($q$), the dependent variable is affected by the current and previous random shocks and is expressed as:

$$X_t = \mu + \sum_{n=1}^{q} \beta_i \epsilon_{t-p} + \epsilon_t \tag{2.2}$$

where $\mu$ is the mean of the series, $\beta_1, \ldots, \beta_p$ are the parameters of the model and $\epsilon_t$ is the white noise error term.

ARMA($p$,$q$) process can be defined as the combination of AR and MA models. In other words, dependent variable is affected by either previous values or shocks.

$$X_t = a + \sum_{n=1}^{p} \varphi_i X_{t-i} + \sum_{n=1}^{q} \beta_i \epsilon_{t-i} + \epsilon_t \tag{2.3}$$

where $\varphi_1, \ldots, \varphi_p, \beta_1, \ldots, \beta_p$ are the parameters of the model, $a$ is a constant number and $\epsilon_t$ is the white noise error term.

### 2.1.1 Non-Stationary Time Series Process

The existence of a trend creates non-stationarity in time series. In a data set, non-stationarity can be observed in two ways: the mean or variance of random samples selected from the data set is not stationary.

First, non-stationarity in the mean can be observed either deterministic, i.e., it has the explicit function form, or stochastic trend, which is not predictable and has "random walk" movement. Deterministic and stochastic non-stationary series can be detected with the KPSS test.

Second, non-stationarity can be observed due to variance instability. In this case, we can apply variance stabilizing transformations if the dataset consists of positive values. The other reason of the non-stationarity can be seasonality.

#### a. Autoregressive Integrated Moving Average Model

Autoregressive Integrated Moving Average, ARIMA($p$,$d$,$q$), is used for analyzing the non-stationary datasets [9]. It is defined as a combination of three statistical models: AR($p$), I($d$) and MA($q$). ARIMA models are distinguished from ARMA models due to integrated terms.

Box-Jenkins methodology for model selection requires a 4-step procedure. The first step is to apply data processing to assure stationarity, and the second is the model identification using ACF and PACF or software packages to find the best ARIMA model. The final model selection is made with respect to performance indicators, minimum AIC or BIC values, whose details are given below.

AIC is a measure of the goodness-of-fit for the fitted model and expressed as:

$$\text{AIC} = -2\log(L) + 2k \tag{2.4}$$

where, $L$ stands for the maximized value of the likelihood function for the estimated model while $k$ is the number of parameters in the statistical model. The component 2k is a penalty for large number parameters in the model. The model whose AIC value is smaller is selected as better model.

BIC is a criterion for model selection among a class of parametric models. It is given as

$$\text{BIC} = -2\log(L) + k\log(n) \tag{2.5}$$

where, $k$ is the number of the parameters in the model, $n$ is the sample size. The model whose BIC value is smaller are selected as better model.

After the model selection, diagnostic checks should be completed. To do so, the validity of Gauss Markov assumptions is checked for the selected model. In other words, residuals should have zero mean, and there should not be any autocorrelation across residuals. Also, if the residual of the selected model is white noise, forecasts can be obtained.

### b. Seasonal Time Series Models

The time series repeating a regular pattern, usually within a one-year period, is called a seasonal time series model. Seasonal series are observed primarily in economics because of the business cycle. Climate-related variables such as temperature follow a seasonal process. The temperature is high in the summer seasons, while the temperature is low in the winter. According to Hylleberg [24], there are three types of seasonality in time series. Time series can show deterministic and stochastic seasonality. The series, which have deterministic seasonality, can be analyzed with seasonal dummies, while those following stochastic seasonality can be analyzed with SARIMA models.

Deterministic seasonality can be analyzed with seasonal dummy variable regression or trigonometric series and is defined as:

$$X_t = \sum_{n=1}^{s} \theta_n D_{nt} + u_t \tag{2.6}$$

where, $D_{nt}$ indicates the dummy variable for $n^{th}$ observation in $s$ observation and $D_{it} = 1$ if the observation at time $t$ is in $n^{th}$ observation and 0 otherwise. $\theta$'s are the seasonal factors.

On the other hand, the stochastic seasonality can be analyzed with the SARIMA models. Models have seasonal and non-seasonal parts. That is, ARIMA$(p, d, q)(P, D, Q)_s$ has the non-seasonal part with $p$, $d$, $q$ parameters and the seasonal part with $P$, $D$, $Q$ where $s$ stands for the number of observation within a year. For instance, for monthly data sets, $s$ is taken as 12.

The best model is selected by measuring forecasting accuracy using the measures such as Mean Error (ME) and MAE.

## 2.2 Exponential Smoothing

As a widely-used and simple forecasting method, exponential smoothing is suitable for discrete time-series data. It is a powerful forecasting technique, even though it has simplicity in terms of computational efficiency. The idea of exponential smoothing is that forecasted values are the weighted averages of past observations.

Exponential Smoothing is developed by Brown [10], Holt [23] and Winters [39] in 1950s. Brown [10] works on the Simple Exponential Smoothing (SES) for the time series with no trend and seasonality. In their work, the forecasts depend on the weighted average of previous observations, and earlier observations have less weight and less impact on future forecasting. In other words, weights decline exponentially as the observations get older [25].

In SES, the forecast $\hat{y}_{t+1}$ is the weighted average combination of most recent observation $y_t$

with weight $\alpha$, and the most recent forecast $y_t$ with a weight of $(1-\alpha)$. It is expressed as:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{2.7}$$

where, $\alpha$ is the smoothing parameter between 0 and 1, $y_t$ is the observation at time $t$, $\hat{y}_t$ is the forecast value for time period $t$ and $\hat{y}_{t+1}$ is the forecast value for $t+1$.

To include the impact of the trend to the simple exponential smoothing, equation 2.7 is extended to Double Exponential Smoothing (DES). This method consists of smoothing parameter ($\alpha$) and trend coefficient, such that

$$\hat{Y}_t(h) = S_t + hT_t \tag{2.8}$$

where, $S_t$ is the current level and $T$ is the trend while $h$ is the coefficient of trend at which:

$$\begin{aligned} S_t &= \alpha Y_t + (1-\alpha)(S_{t-1} + T_{t-1}) \\ T_t &= \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \end{aligned} \tag{2.9}$$

where, $\alpha$ and $\beta$ are smoothing parameter and trend coefficients, respectively.

Equation 2.9 shows that $S_t$ is the weighted average combination of the value of $Y_t$ and the sum of the last smoothed value and the trend of the previous period $T_{t-1}$. The second line in Equation 2.9 updates the trend as a weighted average of the difference between the current level and one-step ahead level and the estimated trend term of one-step ahead.

Furthermore, Holt-Winters exponential smoothing technique is used to analyze time series that have both trend and seasonal irregularities. To do so, three smoothing parameters $\alpha$, $\beta$, and $\gamma$ are needed for level, trend, and seasonal variations, respectively. These equations are shown below:

$$\begin{aligned} S_t &= \alpha \frac{Y_{t-1}}{I_{t-s}} + (1-\alpha)(S_{t-1} + T_{t-1}) \\ T_t &= \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1} \\ I_t &= \gamma \frac{Y_t}{S_t} + (1-\gamma)I_{t-s} \end{aligned} \tag{2.10}$$

where, $S$ is the smoothed observation, $T$ is the trend factor, $I$ is seasonal index, $t$ denotes time period, $\alpha$,$\beta$,$\gamma$ values are determined by using minimized mean square error (MSE) value. Based on these three equations, $k$-step ahead forecast is expressed as:

$$\hat{Y}_t(k) = (S_t + hT_t)I_{t+k-s} \tag{2.11}$$

## 2.3 Correlation Coefficients and Copulas

Pearson's correlation coefficient is a measure of linear correlation between two variables. $X_1$ and $X_2$ is represented as:

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sigma_{x_1} . \sigma_{x_2}} \tag{2.12}$$

where $Cov(X_1, X_2) = E[(X_1 - E(X_1))(X_2 - E(X_2))]$, $\sigma_{x_i}$ is the standard deviation of $x_i$ for $i = 1, 2$, and $-1 < \rho(X_1, X_2) < 1$.

Correlation coefficients $1$ and $-1$ correspond to perfect positive and negative linear correlation, respectively, while $0$ implies the independence of these two variables [31].

When the random variables are correlated with each other, copula is a beneficial tool to model interdependence. According to Pfaff, [31], the linear correlation coefficient gives us the correct measure if and only if these variables are jointly elliptically distributed. Moreover, the definition of the linear correlation coefficient is valid for pairs of variables with finite variance. Thus, different measures such as the copula should be used to measure the dependence relationship between random variables.

A copula is the distribution function in the d-dimensional space of a d-element random vector with standard uniformly distributed marginal functions. With the copula functions, random variables whose distributions are different from each other can be employed, and the dependence structure of these distributions can be detected. In the literature, there are many copula families. With the help of the copula, the closed form of the joint probability distribution function can be obtained. For instance, Gaussian copula, one of the copula families, is used when the components are normally distributed. Another copula family, Archimedean copulas, consists of the copulas like Clayton, Gumbel, and Joe. Archimedean copulas provide high-dimensional modeling with one parameter. T-copula, however, is used to model extreme circumstances, namely tail dependency, due to its flexibility.

The joint cumulative distribution function of two continuous random variables can be shown as [35]

$$T(X, Y) = C[F(X), H(Y)] \tag{2.13}$$

where, $F(X)$ is the marginal distribution of $X = x_1, \dots, x_d$ and $H(Y)$ is the marginal distribution of $Y = y_1, ..., y_d$. In this equation, $C$ stands for the copula function, built over uniform marginals. Thus, the cumulative distribution function of two random variables can be found by determining the dependence structure $C$ and the marginal distributions of random variables. This equation can be extended for the d-element random vector [31].

For bivariate case, the copula function can be expressed as the combination of two density functions:

$$C(x) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_0^{x_1} \int_0^{x_2} c(x) dx \tag{2.14}$$

$$C(x) = \frac{f(F_1^{-1}(x_1), F_2^{-1}(x_2))}{f_1(F_1^{-1}(x_1)) f_2(F_2^{-1}(x_2))}, \tag{2.15}$$

or equivalently,

$$c(F_1(x_1), F_2(x_2)) = \frac{f(x_1, x_2)}{f_1(x_1) f_2(x_2)}. \tag{2.16}$$

In general,

$$F(x_1, ..., x_d) = P[F_1(x_1), F_2(x_2), ..., F_d(x_d)] = C(F_1(x_1), ..., F_1(x_1)) \tag{2.17}$$

$$C(F_d(x_d), ..., F_d(x_d)) = \frac{f(x_1, ..., x_d)}{f_1(x_1)...f_d(x_d)}. \tag{2.18}$$

In the multivariate case, it is advantageous to model bivariate forms connected to each other by specific copula formations called Vine-copulas. When we use vine copula high dimensional probability distributions, instead of multi-dimensional copula, we use bivariate copulas as building blocks in order to create multivariate copulas [13]. To do so, conditional probability functions are obtained from the probability density functions with d-variables first, and then bivariate copulas are created from such conditional probabilities. For example, assuming that two variables case, the representation is as follows:

$$f(x_1, x_2) = f_{1|2}(x_1|x_2).f(x_2) = f_{2|1}(x_2|x_1).f(x_1) \tag{2.19}$$

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2))f_1(x_1)f_2(x_2). \tag{2.20}$$

Thus, $f_{1|2}(x_1|x_2)$ can be shown as:

$$f_{1|2}(x_1|x_2) = c_{1,2}(F_1(x_1), F_2(x_2)f_1(x_1). \tag{2.21}$$

For 3-dimensional space, probability density function with three variables can be shown as bivariate copula and marginal probability distribution function:

$$f(x_1, x_2, x_3) = f_{1|2,3}(x_1|x_2, x_3).f_{2|3}(x_2|x_3).f_3(x_3) \tag{2.22}$$

Equation 2.22 leads to the following Equations 2.23 and 2.24:

$$f_{2|3}(x_2|x_3) = c_{2,3}(F_2(x_2), F_3(x_3)f_2(x_2) \tag{2.23}$$

$$f_{1|2,3}(x_1|x_2, x_3) = c_{1,2|3}(F_{1|3}(x_1|x_3), (F_{2|3}(x_2|x_3))f_{1|3}(x_1|x_3) \tag{2.24}$$

Also, we can write $f_{1|3}(x_1|x_3)$ as

$$f_{1|3}(x_1|x_3) = c_{1,3}(F_1(x_1), F_3(x_3)f_1(x_1). \tag{2.25}$$

Finally, we can express the PDF with three random variables as the PDF of each random variable and copula functions as pairs. The pairs which are expressed as

$$\begin{aligned} f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3)c_{2,3}(F_2(x_2), F_3(x_3))c_{1,3}(F_1(x_1), F_3(x_3)) \\ c_{1,2|3}(F_{1|3}(x_1|x_3)(F_{2|3}(x_2|x_3))). \end{aligned} \tag{2.26}$$

Such kind of pair copula constructions can be extended to d-random variables. The factorization is visualized as regular vine trees. The parameters of the above equations are either two-dimensional or conditional two-dimensional. These kinds of vines are called regular vines. For $d = 4$, there are three layers of tree representation.

Allen [7] proposes regular vines (R-Vines) as generalized treed as the flexible tools in high dimensional modelling. Figure 2.1 shows two examples of vine structures which is published in the work of Allen [7].

In Figure 2.1, regular and non-regular vine trees on four nodes are shown, respectively. We see that the edges in the first tree is the node of the second tree for the regular vine and each edge in the first tree has only one node for the second tree. On the other hand, we see that one edge has two nodes for the non-regular vine.

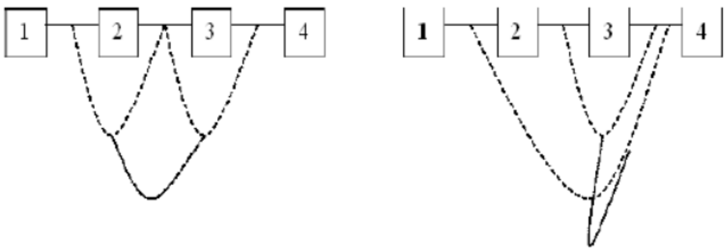



Figure 2.1: Regular and Non-regular Vines

### 2.3.1 Regular Vine (R-Vine)

Kurowicka and Cook [7] states that a regular vine V on n elements with $E(V) = E_1 \cup \cdots \cup E_{n-1}$ corresponding to the set of edges of V if it follows the properties below:

i. There should be N-1 edges for N nodes: $V = T_1, \ldots, T_{n-1}$

ii. $T_1$ is a connected tree with nodes $N_1 = 1, \ldots, n$, plus edges $E_1$; for i =2,. . . ,n-1, $T_i$ is a tree with nodes $N_i = E_{i-1}$,

iii. Proximity condition: Every edge has an impact on the next layer's joint density.

Regular Vines are divided as Canonical Vine (C-Vine) or Drawable Vine (D-Vine), including star trees and line trees [14]. In the C-Vine, if there is a central impact in each layer, it is visualized with a star schema.

If each tree of a regular vine has a unique node of degree $n - 1$, then it is called a C-Vine, while if all nodes in any tree have degrees no more than 2, it is called a D-vine [6].

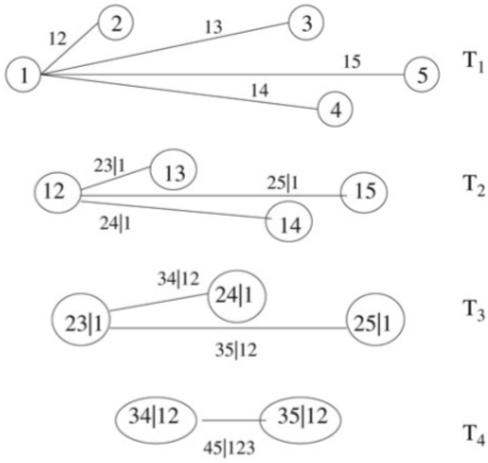



Figure 2.2: D-Vine

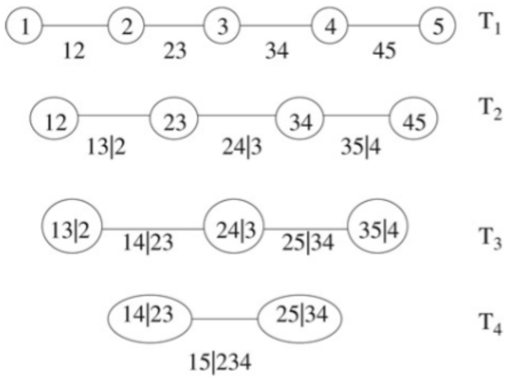



Figure 2.3: C-Vine

Since star schema is used to model dependence structure across variables, C-Vine is useful when one factor is regarded as the vital factor. On the other hand, when no central factor influences the dependency, D-copula is beneficial.

There are two methods to determine the unknown parameters of the copula [31]. The first method, which is a parametric procedure, is to use two-step estimation, which is proposed by Joe [28] and Xu [40] and Shih and Louis [33]. In this way, firstly, unknown parameters for the models of the marginal distributions are estimated. Then, pseudo-uniform variables are extracted from the inverse distribution functions. The likelihood maximization is applied by using these variables. In the second method, the semi-parametric procedure, empirical distribution functions are used to obtain the pseudo-uniform variables, and then these variables are used for maximizing the pseudo-likelihood. Tools of numerical optimization techniques are used to determine the parameters of copula [29].

# CHAPTER 3

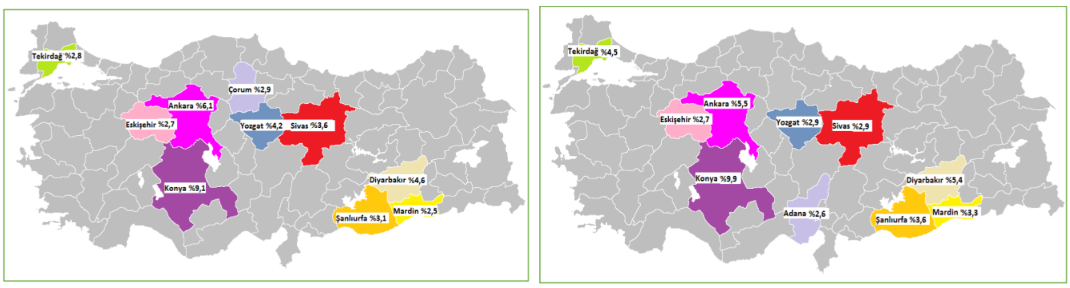
# TIME SERIES ANALYSES

In line with the proposed approach, the climate variables affecting the spot wheat prices are firstly analyzed using time series models explained in Chapter 2. After presenting the data descriptions, each variable is analyzed separately to find the best model representing their future forecast.

## 3.1 Data and Descriptives

The data set contains Konya's monthly average temperature (Celsius), rainfall ($m^2$), and relative humidity (percent) from January 2007 to July 2021. Since Konya alone produces 9.9% of all wheat production in Türkiye and ranks first in Türkiye with 9.1% of wheat planting area, the climate data is collected specifically to this region.

Figure 3.1 from the Ministry of Agriculture [3] shows that wheat planting and production information as of 2021. Figures 3.1a and 3.1b show that Konya is the leading city in wheat planting and production, followed by Ankara and Diyarbakır, respectively.




(a) Wheat planting. (b) Wheat production.

Figure 3.1: Wheat mapping of Türkiye.

Table 3.1 illustrates that among 175 monthly periods, the average rainfall is $28.05mm^2$, and the median is $19.80mm^2$, while these values are closer to the temperature and relative humidity datasets. Table 3.1 temperature and relative humidity distribution are more symmetrical than the rainfall dataset. Coefficient of variation (CV) shows that level of dispersion around the mean is the highest for precipitation and lowest for adjusted wheat prices. For wheat prices, we see that spot wheat prices are between 0.4394 and 4.2831, with a mean value of 1.0305. Since the price increase is dramatically high, especially in 2021, we see a high maximum value. After adjusting the price, we see the mean and median in adjusted wheat prices become closer.

The correlation matrix in Table 3.2 shows a negative correlation between relative humidity and temperature. The reason for the negativity is the increase in temperature which leads to a decrease in relative humidity (%). Therefore, the air becomes drier or vice versa. Also, in the Konya region, we see the existence of steppe climate conditions. This means that the temperature is higher in the summer period with drought. Since Konya's precipitation is almost zero in summer, there is a negative correlation between temperature and precipitation. Table 3.2 shows a strong negative linear correlation between temperature and relative humidity.

Jarque Bera (JB) test is applied to understand whether the time series is normally distributed or not. The results show that none of the variables is normally distributed ($p < 0.01$).

Table 3.1: Descriptive statistics for all climate components and wheat prices

| | Temperature C° | Precipitation ($mm^2$) | Rel Humidity (%) | Spot Wheat (TL) | Adjusted Wheat (TL) |
|---|---|---|---|---|---|
| Mean | 12.59 | 28.06 | 59.52 | 1.03 | 2.19 |
| Median | 12.80 | 19.80 | 59.50 | 0.89 | 2.16 |
| Minimum | -4.80 | 0.00 | 29.20 | 0.44 | 1.78 |
| Maximum | 27.70 | 116.80 | 95.80 | 4.28 | 4.28 |
| Stddev | 8.71 | 25.80 | 17.27 | 0.59 | 0.28 |
| C.V. | 0.69 | 0.92 | 0.29 | 0.57 | 0.13 |
| Skewness | -0.03 | 1.04 | 0.09 | 2.11 | 3.17 |
| Ex. Kurtosis | -1.26 | 0.63 | -1.14 | 5.86 | 17.61 |
| Interquartile range | 15.5 | 37.20 | 30.20 | 0.48 | 0.28 |
| JB test statistics | 11.67 | 34.66 | 9.77 | 80.93 | 54.50 |
| JB p-values | 0.0029 | 0.0030 | 0.0076 | 0.0024 | 0.0014 |

Table 3.2: Correlation matrices for climate components and wheat prices

| | Temperature C° | Precipitation ($mm^2$) | Relative Humidity (%) | Wheat Price (TL) | Adj. Wheat Price (TL) |
|---|---|---|---|---|---|
| Temperature | 1 | -0.3836 | -0.9054 | 0.0079 | -0.1658 |
| Precipitation | | 1 | 0.5202 | -0.0658 | 0.1009 |
| Rel. Humidity | | | 1 | -0.0928 | 0.1727 |
| Wheat Price | | | | 1 | 0.6378 |
| Adj. Wheat Price | | | | | 1 |

The time series modeling of each variable is performed as the next. The observations, autocorrelation function (ACF) and partial Autocorrelation functions (PACF) are plotted to see the influential components in the dataset. Software packages in RSTUDIO are used to find candidate models after the dataset is divided into train and test sets.

## 3.2 Temperature

The dataset shows the monthly temperature of Konya between 2007-2021. The time series plot, ACF and PACF are utilized to see the serial dependence.

Firstly, the time series plots and the result of the KPSS test show that the temperature variable is stationary. Furthermore, the same time series plot of the variable displays the regular pattern. This fluctuation indicates the existence of seasonality as can be seen in Figure 3.2. Figure 3.2, ACF and PACF functions of temperature clearly show the seasonal variation. While the temperature is low in the winter period between December and February, it is high in the summer period. The wheat product is harvested in the spring, especially in May.

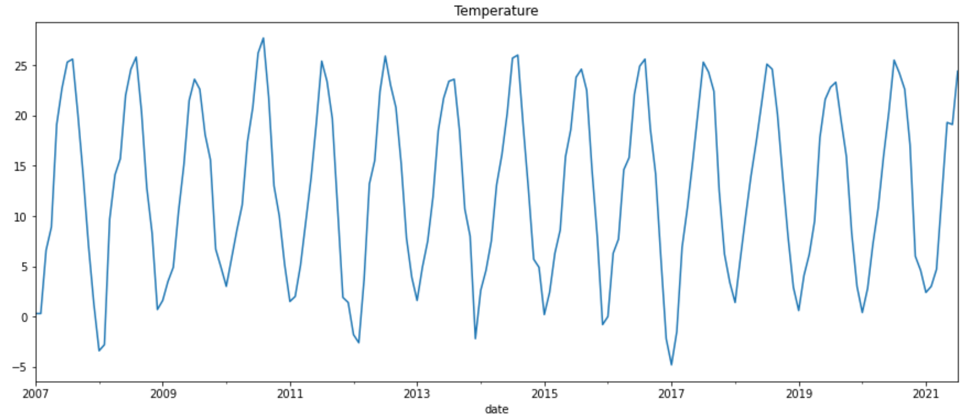



Figure 3.2: Time series of monthly average temperature

Figure 3.3, shows that in the ACF plot of the monthly temperatures are correlated among themselves in the summer and winter seasons.

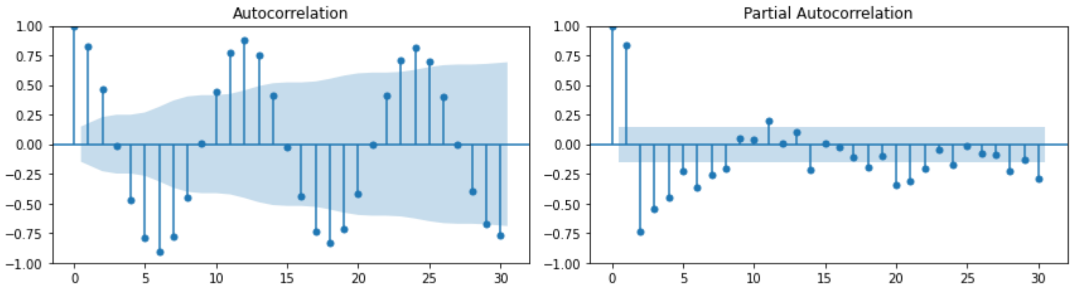



Figure 3.3: ACF and PACF plots of temperature

Time series plots show a constant mean and variance of the pattern using KPSS test and Hansen Canova tests. The results of the KPSS test show that the temperature variable is stationary with a p-value equal to 0.10. So it is validating that the series has constant mean and variance. A variable can follow a seasonal pattern because of two main reasons: deterministic seasonal or stochastic seasonal patterns. Hansen-Canova test, which its null hypothesis assumes a deterministic seasonality, is applied to the temperature dataset. Table 3.3 exposes the test statistic which assures the deterministic seasonality where p-values for all dummy variables exceed 5%.

Table 3.3: Canova-Hansen test of seasonal stability for temperature

| Statistic | p-value | Season |
|---|---|---|
| L1 = 0.0641 | 0.86846 | 1 |
| L2 = 0.1950 | 0.31442 | 2 |
| L3 = 0.0636 | 0.87116 | 3 |
| L4 = 0.0505 | 0.94285 | 4 |
| L5 = 0.1483 | 0.45026 | 5 |
| L6 = 0.2866 | 0.16151 | 6 |
| L7 = 0.0870 | 0.73287 | 7 |
| L8 = 0.0889 | 0.72167 | 8 |
| L9 = 0.1268 | 0.53254 | 9 |
| L10 = 0.2595 | 0.19561 | 10 |
| L11 = 0.0391 | 0.98470 | 11 |
| L12 = 0.0814 | 0.76498 | 12 |

Table 3.4: Model output of ARIMA(1,0,0)(0,0,1)[12] for temperature

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| Const | 1.91983 | 0.2832 | 6.777 | 0.0000 |
| $\phi_1$ | 0.3481 | 0.0776 | 4.485 | 0.0000 |
| $\Theta_1$ | −0.4813 | 0.1194 | −4.030 | 0.0001 |
| dm1 | −1.65533 | 0.3220 | −5.140 | 0.0000 |
| dm2 | 0.5379 | 0.3785 | 1.421 | 0.1553 |
| dm3 | 4.9646 | 0.3971 | 12.50 | 0.0000 |
| dm4 | 9.7500 | 0.3990 | 24.43 | 0.0000 |
| dm5 | 14.2591 | 0.4012 | 35.53 | 0.0000 |
| dm6 | 19.0691 | 0.4016 | 47.47 | 0.0000 |
| dm7 | 22.9992 | 0.4005 | 57.42 | 0.0000 |
| dm8 | 22.7679 | 0.3990 | 57.06 | 0.0000 |
| dm10 | 11.3494 | 0.3771 | 30.09 | 0.0000 |
| dm11 | 4.9558 | 0.3250 | 15.25 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 12.48 | S.D. dependent var | 8.82 |
| Mean of innovations | 0.00 | S.D. of innovations | 1.63 |
| $R^2$ | 0.96 | Adjusted $R^2$ | 0.96 |
| Log-likelihood | −279.08 | Akaike criterion | 588.17 |
| Schwarz criterion | 632.82 | Hannan–Quinn | 606.32 |

In the analyses, the data set is divided into train and test sets with 80%-20% partitions resulting in 140 observations in the training set. Then, the model selections are made using the train set, and the performance of the selected model is tested on the test set. The time series with deterministic seasonality can be analyzed with seasonal dummy variables or Fourier terms.

For this purpose, we include 11 seasonal dummy variables as independent variables for the analyses.

Since the seasonal dummies are significant, the temperature data set is analyzed with ARIMA model with 11 seasonal dummies. Among all plausible models, ARIMA(1,0,0)(2,0,0)[12] (with seasonal dummies), ARIMA(1,0,0)(0,0,1)[12] (with seasonal dummies), and ARIMA (1,0,1)(1,1,1)[12] (without seasonal dummies) are selected as a candidate models based on AIC, AICc and BIC. In order to check the significance of the models, we look at the last estimated parameters of each component. Which reduced the choice to ARIMA(1,0,0)(2,0,0)[12] and ARIMA(1,0,0)(0,0,1)[12] whose parameters are significant. Also, $R^2$ values are almost the same. However, ARIMA(1,0,0)(0,0,1)[12] has the lower AIC value than the ARIMA (1,0,0)(2,0,0)[12]. On the other hand, the parameters of ARIMA(1,0,1)(1,1,1)[12] are not significant. The output of each candidate models are shown in the Appendix A.1 and Appendix A.2. Finally, ARIMA(1,0,0)(0,0,1)[12] with seasonal dummies is the best model since AIC value is lower and the parameters are significant whose summary is given in the table 3.4. According to the results, the p-values of all parameters are less than 0.05 except for the dummy variable for February.

Forecasting results show that ARIMA(1,0,0)(0,0,1)[12] fits well with the test data as it gives the lowest AIC value. Figure 3.4 shows that forecast results and test datasets follow the same movement across seasons. After every peak in the summer season, forecasts and the actual temperature decreased around zero each winter. The forecast evaluation of the temperature dataset shows that the mean error of the model is 0.4170.

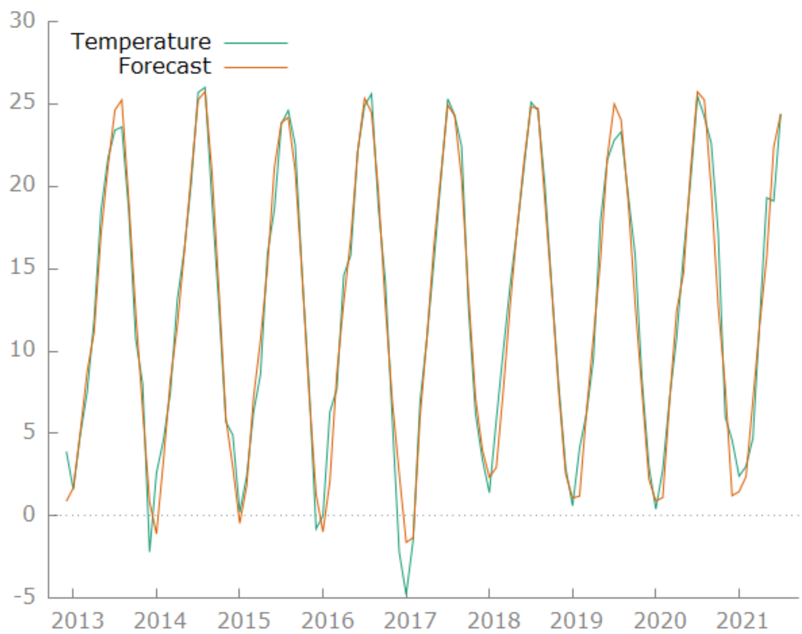


Figure 3.4: ARIMA(1,0,0)(0,0,1)[12] with seasonal dummies

We can use either visual ways or tests to check the normality of residuals. For a visual perspective, residuals are normally distributed since points are clustered in the reference line

Table 3.5: Autocorrelation and normality tests for ARIMA(1,0,0)(0,0,1)[12]

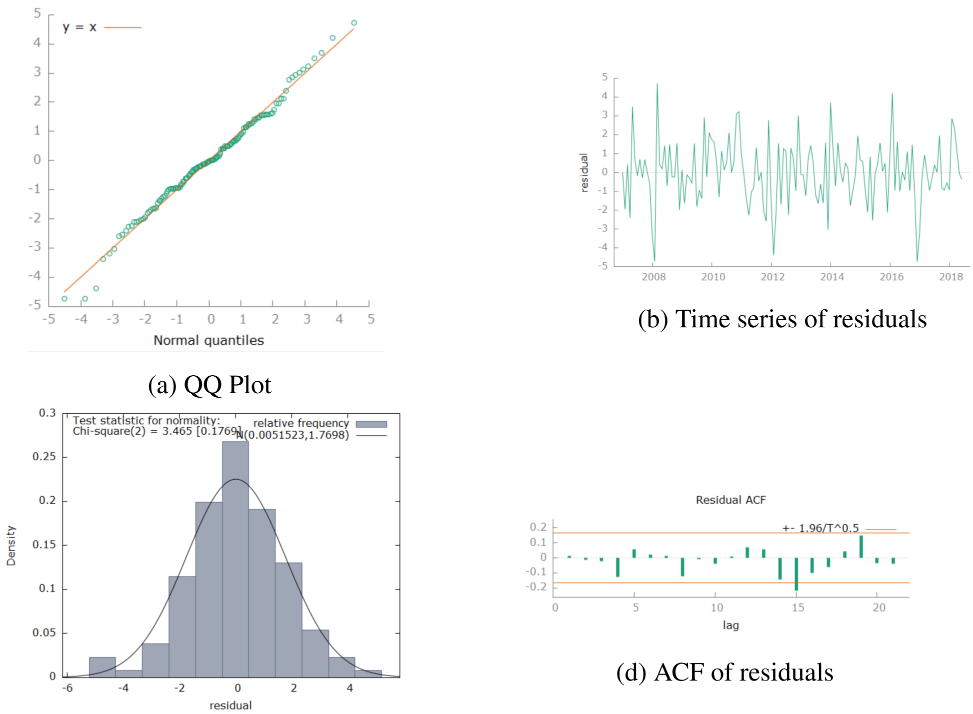
| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 0.00 | Jarque-Bera (JB): | 1.92 |
| Prob(Q): | 0.80 | Prob(JB) | 0.38 |

in the QQ plot of Figure 3.5. Also, Jarque-Bera test results show that the null hypothesis is accepted with $p = 0.91$, so the residuals are normally distributed as well. Moreover, the Ljung-Box test is on the residuals to test if the residuals are white noise or not. The results in Table 3.5 show that the null hypothesis is accepted with a p-value of 0.80, so we can conclude that the residuals are white noise and follow the normal distribution. This resulting variable is implemented as input to the copula analyses.

The evaluation results of the selected model is presented Table 3.6. Root mean squared error, which expresses the square root of variances of the errors, is found as 1.80. On the other hand, the Mean Absolute Percentage Error (MAPE) value is calculated as 22.30 since the residuals, which are close to zero, increase the MAPE. Therefore, we can say that the forecasts are successfully captured.

Table 3.6: Forecast evaluation results of ARIMA(1,0,0)(0,0,1)[12] for temperature

| | |
|---|---|
| Mean Error | 0.4170 |
| Root Mean Squared Error | 1.80 |
| Mean Absolute Error | 1.35 |
| Mean Percentage Error | 2.05 |
| Mean Absolute Percentage Error | 22.30 |




(a) QQ Plot

(b) Time series of residuals

(c) Histogram of temperature

(d) ACF of residuals

Figure 3.5: Diagnostic check of ARIMA(1,0,0)(0,0,1)[12]

## 3.3 Relative Humidity

Monthly average relative humidity follows a seasonal pattern around a constant mean. KPSS test is applied to this dataset and validated that the series is stationary with a p-value of 0.10. Also, a seasonal pattern is detected in Figures 3.6 and 3.7. The relative humidity increases in the winter period and decreases in the summer. Table 3.7 shows that the time series has stochastic seasonality similar to the temperature data set.

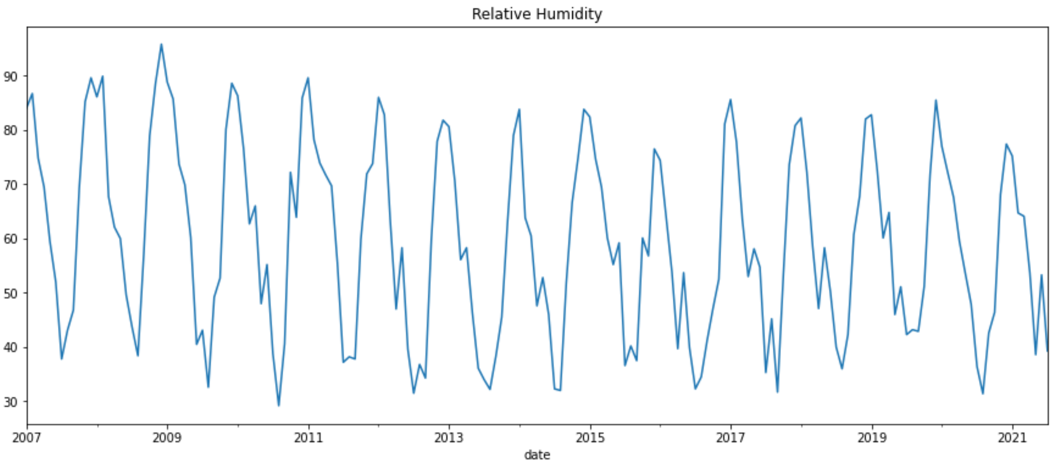



Figure 3.6: Time series of monthly relative humidity (%)

Autocorrelation function of relative humidity in the Figure 3.7 shows that the series is stationary and seasonal. KPSS test result shows that that the series is stationary.

Table 3.7: Canova-Hansen test for seasonal stability for relative humidity

| Statistic | p-value | Month |
|---|---|---|
| L1 = 0.7787 | 0.00362 *** | 1 |
| L2 = 0.8804 | 0.00138 *** | 2 |
| L3 = 0.4746 | 0.04205 ** | 3 |
| L4 = 0.5026 | 0.03418 ** | 4 |
| L5 = 0.4274 | 0.05946 * | 5 |
| L6 = 0.0895 | 0.71842 | 6 |
| L7 = 0.2640 | 0.18948 | 7 |
| L8 = 0.0974 | 0.67481 | 8 |
| L9 = 0.2978 | 0.14922 | 9 |
| L10 = 0.5421 | 0.02541 ** | 10 |
| L11 = 0.4951 | 0.03614 ** | 11 |
| L12 = 0.4781 | 0.04097 ** | 12 |

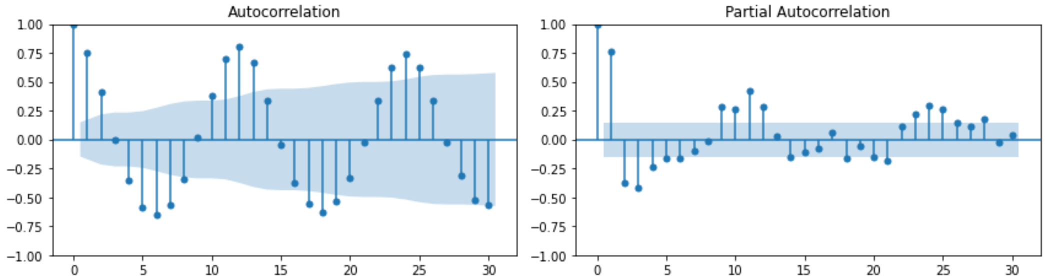



Figure 3.7: ACF and PACF functions

As a result, we select ARIMA(1,0,1)(1,1,1)[12] and ARIMA(0,1,1)(0,1,1)[12] as candidate models based on AIC. However, the model results of ARIMA(1,0,1)(1,1,1)[12] in the Table A.3 show that $\Phi_1$ parameter is insignificant since the p-value of the relevant parameter is greater than 0.50. Thus, this model cannot be one of the candidate models.

Table 3.8: Model output of ARIMA(0,1,1)(0,1,1)[12] for relative humidity

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| Const | 0.0166 | 0.0603 | 0.2756 | 0.8260 |
| $\theta_1$ | $-0.7298$ | 0.0725 | $-10.06$ | 0.0000 |
| $\Theta_1$ | $-0.7731$ | 0.0756 | $-10.23$ | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | $-0.0889$ | S.D. dependent var | 11.38845 |
| Mean of innovations | $-0.1176$ | S.D. of innovations | 6.9436 |
| $R^2$ | 0.8485 | Adjusted $R^2$ | 0.847296 |
| Log-likelihood | $-432.1568$ | Akaike criterion | 872.3135 |
| Schwarz criterion | 883.6903 | Hannan–Quinn | 876.7390 |

Since ARIMA(0,1,1)(0,1,1)[12] has lowest AIC value among other candidate model, it is selected as the best model. The model description of alternative models is in the Appendix part.

Figure 3.8 shows the actual and forecasts of the selected model. It is clear that the relative humidity decreases from the winter to the summer seasons. The estimated values and actual values move in the same direction. Table 3.10 indicates that on average, the sum of absolute errors is 4,62 $mm^2$ and %9 for relative humidity.

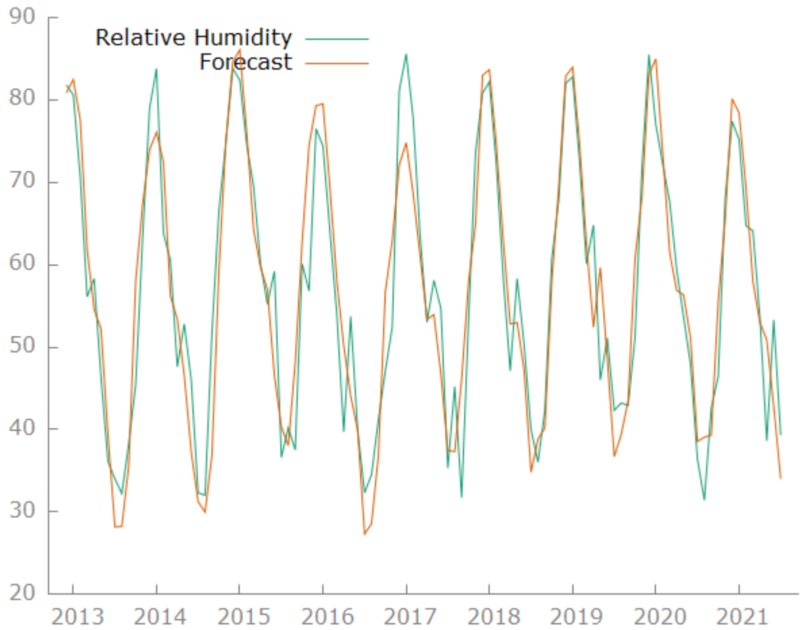



Figure 3.8: Forecasting relative humidity

Also, the Ljung-Box and Jarque-Bera tests are applied to the best model for the diagnostic checks. The results shown in Table 3.9 indicate that the residuals are white noise and normally distributed.

Table 3.9: Diagnostic checks for ARIMA(0,1,1)(0,1,1)[12]

| | | | |
|---|---|---|---|
| Ljung-Box (L1) (Q): | 6.0759 | Jarque-Bera (JB): | 1.6062 |
| Prob(Q): | 0.6387 | Prob(JB): | 0.4479 |

Table 3.10: Forecast evaluation of relative humidity for ARIMA(0,1,1)(0,1,1)[12]

| | |
|---|---|
| Mean Error | −0.4696 |
| Mean Squared Error | 5.8313 |
| Mean Absolute Error | 4.6186 |
| Mean Percentage Error | −1.4118 |
| Mean Absolute Percentage Error | 9.0014 |

## 3.4 Precipitation

Unlike the other two sections, TES is applied to the precipitation dataset in addition to time series models in this section. Since the candidate ARIMA models could not catch the predictions well and the $R^2$ of the candidate ARIMA models are low, we employ the Triple Exponential Smoothing as an alternative model to the time series models for only the precipitation.

The time series plot shows that the precipitation amount is lower for the summer season than in winter. For 175 observations, the minimum precipitation amount is 0, while the maximum is 116.8. KPSS test is applied to the data set, and it is validated that the time series is stationary. In Figure 3.10, there is no indication of non-stationarity.

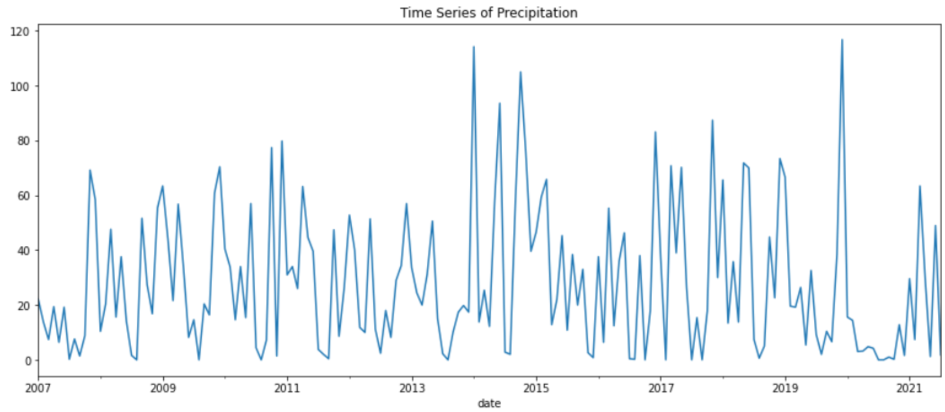



Figure 3.9: Time series of precipitation

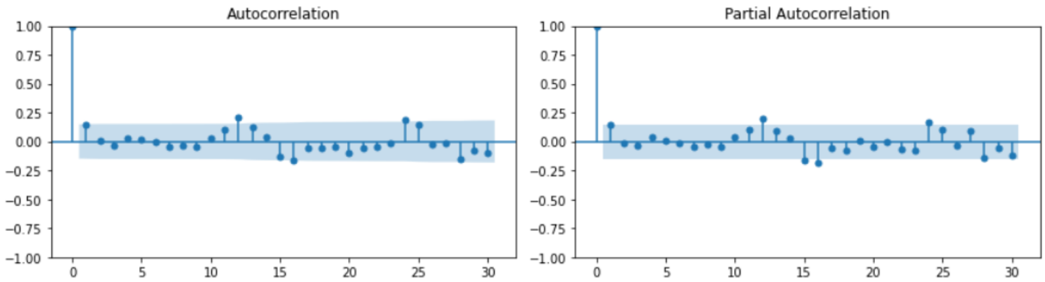



Figure 3.10: ACF and PACF of precipitation

Firstly, Holt-Winters is applied to the precipitation dataset as it detects the impact of seasonality and trend. Since default smoothing level, trend, and seasonality parameters are not give reliable results; hyperparametric optimization is applied to the dataset. According to the results of the optimization best mean absolute error value is equal to 19.79 with the 0.1, 0.5, 0.1 best alpha, best beta and best gamma values, respectively. Then, the final Triple Exponential Smoothing model is conducted with the best parameters. The forecast results are shown in Figure A in Appendix. The predictions, test set values and model results are shown in the Figure 3.11 and in the Appendix part. Also, the JB test is applied to the residuals obtained from precipitation set. The results show that the errors are not normally distributed.

Secondly, candidate models are identified using the ACF, PACF, and software packages so that the ARIMA model's parameters are determined. As a result, ARIMA(1,0,1), SARIMA(0,0,1)(0,1,1)[12], SARIMA(0,0,0)(0,1,1)[12], SARIMA(1,1,1)(1,1,1)[12], SARIMA(0,1,1)(0,1,1)[12] models are selected as candidate models.

The parameters of SARIMA(0,1,1)(0,1,1)[12] model are significant. The model has 16.5% $R^2$ value with 1177.009 AIC value. The parameters of SARIMA(0,0,0)(0,1,1)[12] are also

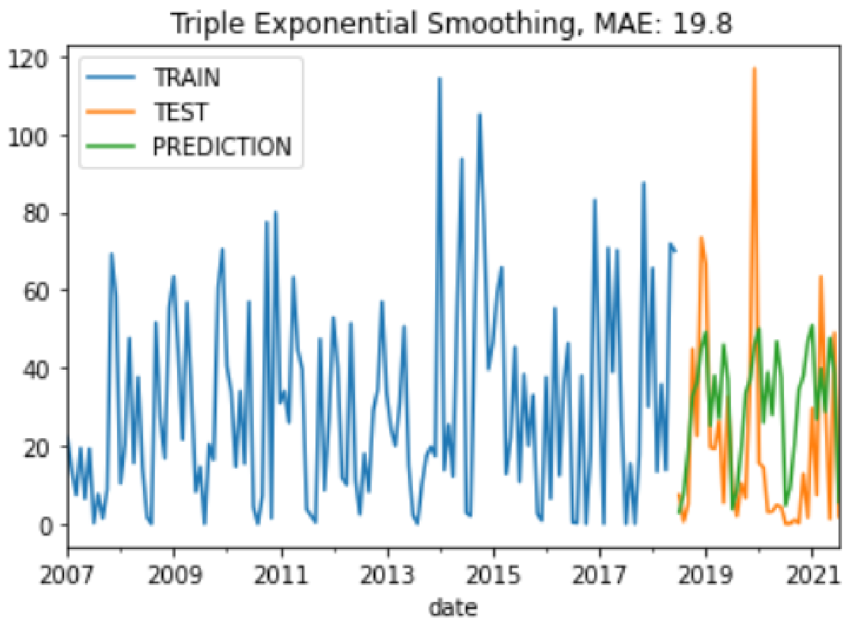



Figure 3.11: Triple exponential smoothing

Table 3.11: Triple exponential smoothing model results

| Dependent Variable | Precipitation | No. of Observations | 140 |
|---|---|---|---|
| Model | Exponential Smoothing | SSE | 88267.98 |
| Optimized | True | AIC | 923.601 |
| Trend | Additive | BIC | 970.437 |
| Seasonal | Additive | AICc | 929.349 |
| Seasonal Periods | 12 | | |
| Box-Cox | False | | |
| Box-Cox Coefficient | None | | |

significant. It has the 17.5% $R^2$ value with 1176.407 AIC value. On the other hand, ARIMA (1,0,1), SARIMA(0,0,1)(0,1,1)[12] and SARIMA(1,1,1)(1,1,1)[12], we see that $\theta_1$ and $\Phi_1$ parameter are insignificant so these models cannot be candidate models.

Among these models, ARIMA(0,0,0)(0,1,1)[12] is selected as best model because its AIC value is lower than the AIC value of SARIMA(0,1,1)(0,1,1)[12]. The details of the candidate models are shown in the Appendix part.

Since ARIMA(0,0,0)(0,1,1)[12] has the lowest AIC value, this model is chosen as the best model across the alternative SARIMA models. Summary statistics and the diagnostic checks are shown in Table 4.13 and in Figure 3.13. Both plots and summary statistics show that the fitted values are not normally distributed, and the model is autocorrelated. Formally, the JB test is applied to the residuals of the SARIMA model, and the null hypothesis is rejected with $p = 0.01$.

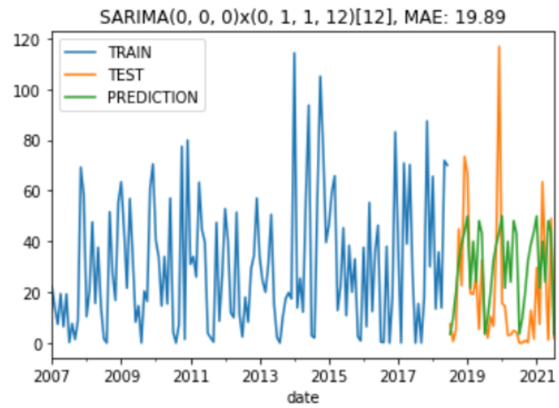
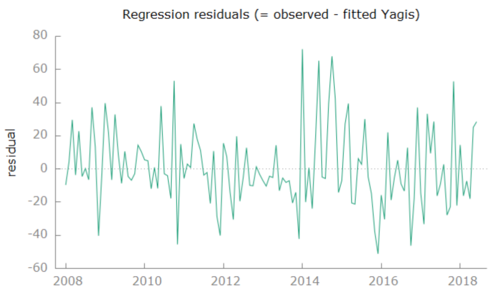



Figure 3.12: SARIMA model

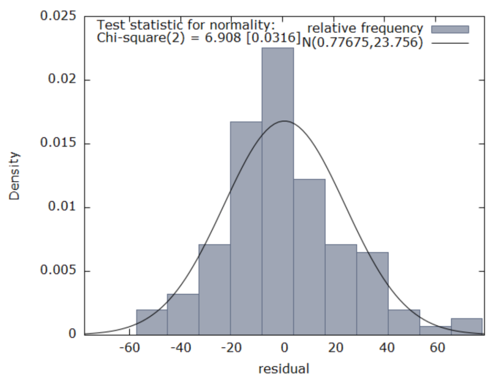
Table 3.12: Model output of SARIMA(0,0,0)(0,1,1)[12]

| | Coefficients | Standart Error | z | p-value |
|---|---|---|---|---|
| Const | 0.998855 | 0.634506 | 1.574 | 0.1154 |
| $\Theta_1$ | $-0.869266$ | 0.119568 | $-7.270$ | 0.0000 |

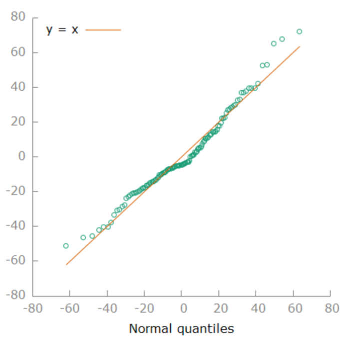
| | | | |
|---|---|---|---|
| Mean dependent var | 1.484127 | S.D. dependent var | 32.09498 |
| Mean of innovations | 0.776751 | S.D. of innovations | 23.57986 |
| $R^2$ | 0.175042 | Adjusted $R^2$ | 0.175042 |
| Log-likelihood | $-585.2034$ | Akaike criterion | 1176.407 |
| Schwarz criterion | 1184.916 | Hannan–Quinn | 1179.864 |



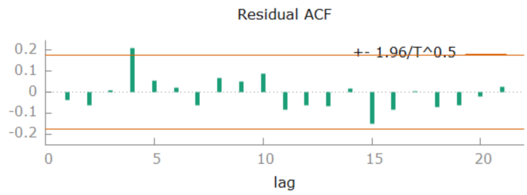

(a) Time series of residuals




(b) Histogram of residuals




(c) QQ plot




(d) ACF of residuals

Figure 3.13: Diagnostic checks of SARIMA(0,0,0)(0,1,1)[12]

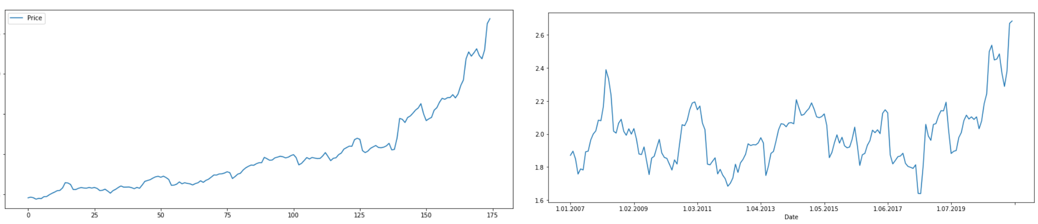
Ultimately, we can say that all candidate models' residuals are not normally distributed. But AIC value of Triple Exponential Smoothing is the lowest. Also, we see that the mean absolute error value (19.89) of the candidate SARIMA model is greater than the value of the TES model (19.80), so the residuals obtained from Holt-Winters are used for the copula analyses.

## 3.5 Wheat Prices

In this part, monthly wheat prices are modeled and forecasted. This section has two parts. In the first part, time series analyses are applied for the wheat prices, which are taken from Konya Mercantile Exchange's website. The second part involves the analysis of adjusted prices against inflation.

The data set covers 2007-2021 and represents monthly wheat prices per ton in Turkish Lira for Konya. The time series plots and ACF-PACF functions of spot wheat prices are also shown to indicate the market's price movement.

Spot wheat prices are highly sensitive to inflation. As shown in Figure 3.14a, it has an upward trend. On the other hand, the second time series in Figure 3.14b shows the adjusted prices against inflation. The change in Consumer Price Index data set is obtained from the Central Bank of Türkiye's website, and the prices are adjusted for the inflation rate. While the gradual increase in prices is determined in 3.14a, prominent fluctuations exist until September 2021. Descriptive statistics of spot wheat prices and adjusted wheat prices in Table 3.1 show that after adjusting prices against inflation, the mean and median of wheat prices get closer, and the standard deviation decreases. Also, after the adjustment, the coefficient of variation lowers, which means there is a lower level of dispersion around the mean.




(a) Time series of wheat prices (TL)

(b) Adjusted wheat prices

Figure 3.14: Time series

ACF plots for both spot price and adjusted price show slow decay which is the indication of non-stationarity in the figures 3.15 and 3.16.

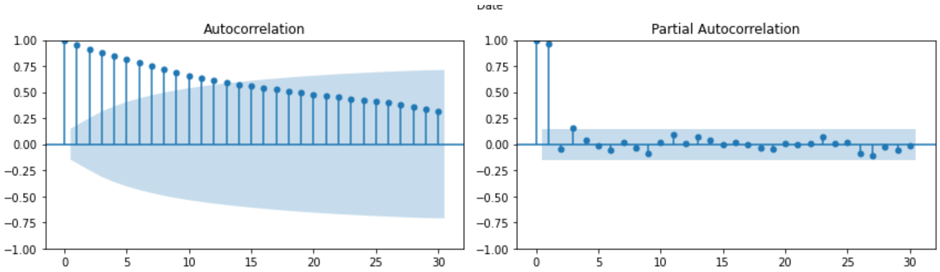



Figure 3.15: ACF and PACF of wheat prices

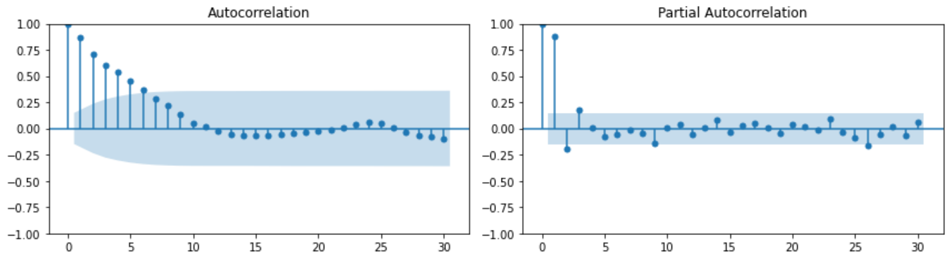
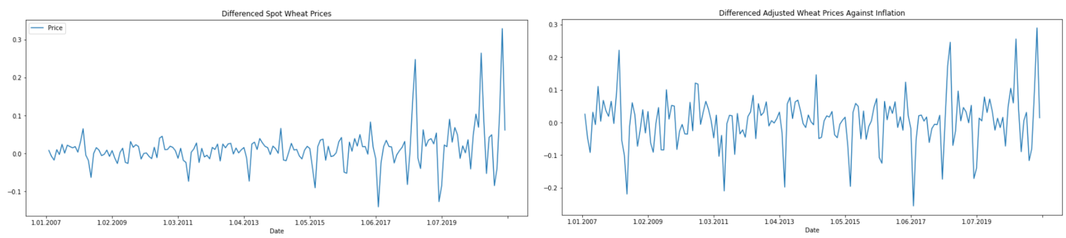



Figure 3.16: ACF and PACF of adjusted wheat Prices

It is revealed that non-stationary data sets have a unit root problem. The KPSS test is applied to test stationarity. The result shows that null hypothesis is rejected with a p-value of 0.01, so the prices have a stochastic trend because of the unit root problem.

We take the first differences of the data set to eliminate the trend effect. Figure 3.17a shows the differenced spot prices of wheat, while Figure 3.17b demonstrates the differences of the adjusted prices against inflation. Then KPSS test is applied again for two time series. According to the results of the KPSS test, differenced spot prices has still unit root problem. On the other hand, the results of the KPSS test indicate that the null hypothesis is not rejected so adjusted prices become stationary.




(a) Differenced wheat prices (b) Differenced adjusted wheat prices

Figure 3.17: Wheat prices

By the help of ACF and PACF plots of the differenced dataset, ARIMA(2,1,2) can be suggested as a candidate model.

The alternative model is found with the Box-Cox transformation. In order to stabilize the variance of non-stationary time series, the variance transformation is applied to the dataset. The type of transformation is determined with the lambda value. We use the BOXCOX()

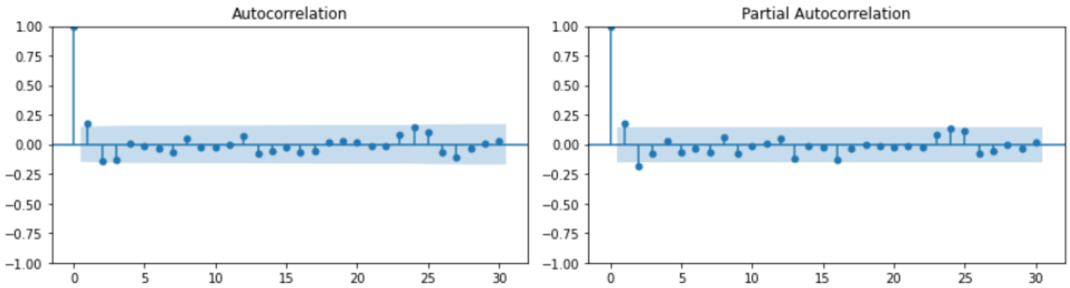



Figure 3.18: ACF and PACF plots of differenced adjusted wheat prices

function in PYTHON to find an optimal $\lambda$ for Box-Cox transformation.

For the weighted average wheat price, the dataset lambda value is equal to $-2.38$. The new data set consists of transformed prices for $\lambda$.

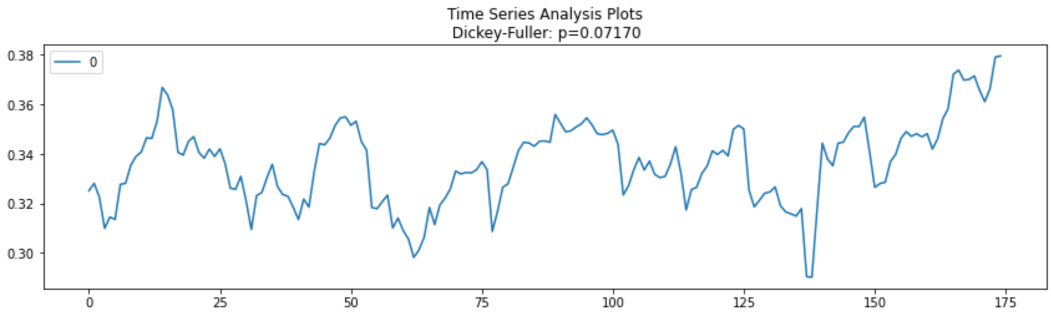



Figure 3.19: Time Series transformed adjusted wheat price

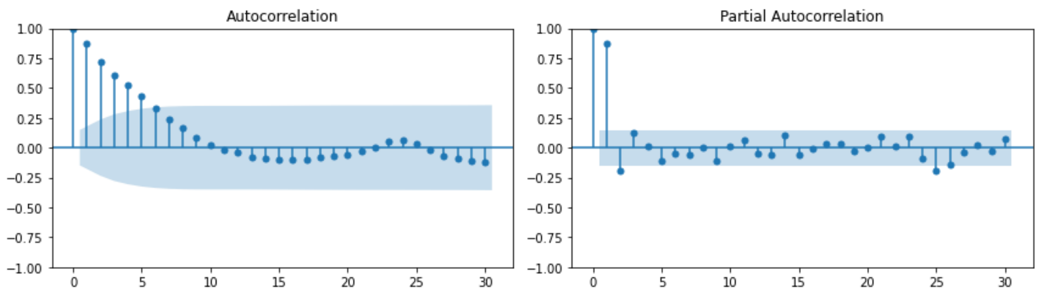



Figure 3.20: ACF and PACF of transformed adjusted wheat price

KPSS test is applied transformed dataset shows that the new data set is still non-stationary. Also, there is no indication of seasonality since the ACF plot do not follow seasonal pattern. As it is seen from the Table 3.20, ACF function has the decay as the number of lag increases. This is the indication of non-stationarity in the dataset. Thus, the first difference is retaken. After the differencing, the new data set became stationary. By looking the ACF and PACF plots in the Figure 3.21, ARIMA(1,1,1) model can be suggested.

Also, ARIMA(2,1,2)(2,0,0) model is suggested by AUTO.ARIMA() function in RSTUDIO. Finally, ARIMA(2,1,2), ARIMA(1,1,1) and SARIMA(2,1,2)(2,0,0) are suggested as candidate models.

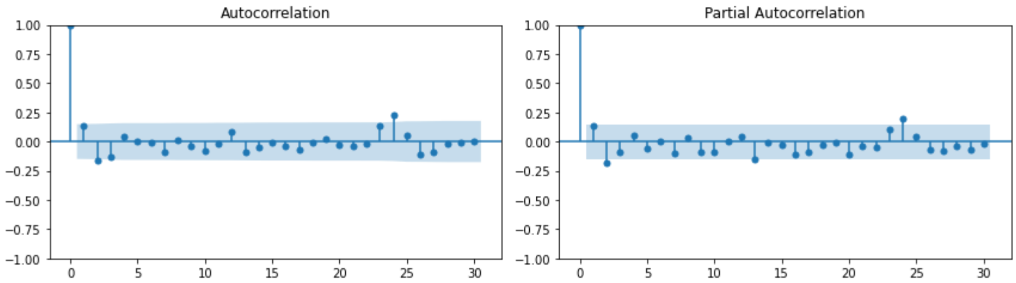



Figure 3.21: ACF and PACF of differenced transformed adjusted wheat price

Table 3.13 shows that the parameter of ARIMA(1,1,1) is insignificant. Thus, this model cannot be one of the selected models.

Table 3.13: Model output for ARIMA(1,1,1) for adjusted wheat prices

| | Coefficients | Standart Error | z | p-value |
|---|---|---|---|---|
| $\phi_1$ | $-0.108975$ | 0.328109 | $-0.3321$ | 0.7398 |
| $\theta_1$ | 0.252479 | 0.311714 | 0.8100 | 0.4180 |

| | | | |
|---|---|---|---|
| Mean dependent var | $-0.001682$ | S.D. dependent var | 0.070182 |
| Mean of innovations | $-0.001640$ | S.D. of innovations | 0.069254 |
| $R^2$ | 0.768319 | Adjusted $R^2$ | 0.766603 |
| Log-likelihood | 171.3816 | Akaike criterion | $-336.7633$ |
| Schwarz criterion | $-328.0034$ | Hannan–Quinn | $-333.2035$ |

According to the results of ARIMA(2,1,2) model, the parameters are significant, so this model whose results are shown in Table 3.14 is selected as one of the candidate models.

Table 3.14: Model output for ARIMA(2,1,2) for adjusted wheat prices

| | Coefficients | Standart Error | z | p-value |
|---|---|---|---|---|
| $\phi_1$ | 0.522612 | 0.362034 | 1.444 | 0.1489 |
| $\phi_2$ | $-0.606802$ | 0.241240 | $-2.515$ | 0.0119 |
| $\theta_1$ | $-0.388903$ | 0.401146 | $-0.9695$ | 0.3323 |
| $\theta_2$ | 0.451238 | 0.273916 | 1.647 | 0.0995 |

| | | | |
|---|---|---|---|
| Mean dependent var | $-0.001682$ | S.D. dependent var | 0.070182 |
| Mean of innovations | $-0.001706$ | S.D. of innovations | 0.068290 |
| $R^2$ | 0.770740 | Adjusted $R^2$ | 0.765569 |
| Log-likelihood | 173.2586 | Akaike criterion | $-336.5171$ |
| Schwarz criterion | $-321.9172$ | Hannan–Quinn | $-330.5841$ |

On the other hand, SARIMA(2,1,2)(2,0,0) model has the significant parameters with 79% of $R^2$ value.

Table 3.15: Model output of SARIMA(2,1,2)(2,0,0) )

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| $\phi_1$ | 0.399343 | 0.0321787 | 12.41 | 0.0000 |
| $\phi_2$ | −0.977285 | 0.0244760 | −39.93 | 0.0000 |
| $\Phi_1$ | 0.143734 | 0.0904944 | 1.588 | 0.1122 |
| $\Phi_2$ | 0.204892 | 0.100506 | 2.039 | 0.0415 |
| $\theta_1$ | −0.342459 | 0.0385662 | −8.880 | 0.0000 |
| $\theta_2$ | 1.00000 | 0.0486651 | 20.55 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | −0.001682 | S.D. dependent var | 0.070182 |
| Mean of innovations | −0.002028 | S.D. of innovations | 0.064640 |
| $R^2$ | 0.793836 | Adjusted $R^2$ | 0.785967 |
| Log-likelihood | 177.9357 | Akaike criterion | −341.8714 |
| Schwarz criterion | −321.4315 | Hannan–Quinn | −333.5652 |

When the model results are compared, SARIMA(2,1,2)(2,0,0) has the lowest AIC value with the highest $R^2$ value. Also, Figure 3.22 shows there is no relationship between residuals' current value and its past values. Also we see that standard residuals are white noise. Figure 3.23 shows the forecasts and actuals of test set between 2019 and 2021. The forecast results show that in the summer season, adjusted prices decrease.

The forecast evaluation statistics for SARIMA(2,1,2)(2,0,0) are shown in Table 3.16. The model has the 3.25% MAPE value. The residuals derived from this model are used in the next Chapter for the copula analyses. Test values, predictions, and standard errors are shown in Appendix A.10.

Table 3.16: Forecast evaluation statistics

| | |
|---|---|
| Mean Error | 0.0274 |
| Root Mean Squared Error | 0.1051 |
| Mean Absolute Error | 0.0713 |
| Mean Percentage Error | 1.1989 |
| Mean Absolute Percentage Error (MAPE) | 3.2511 |

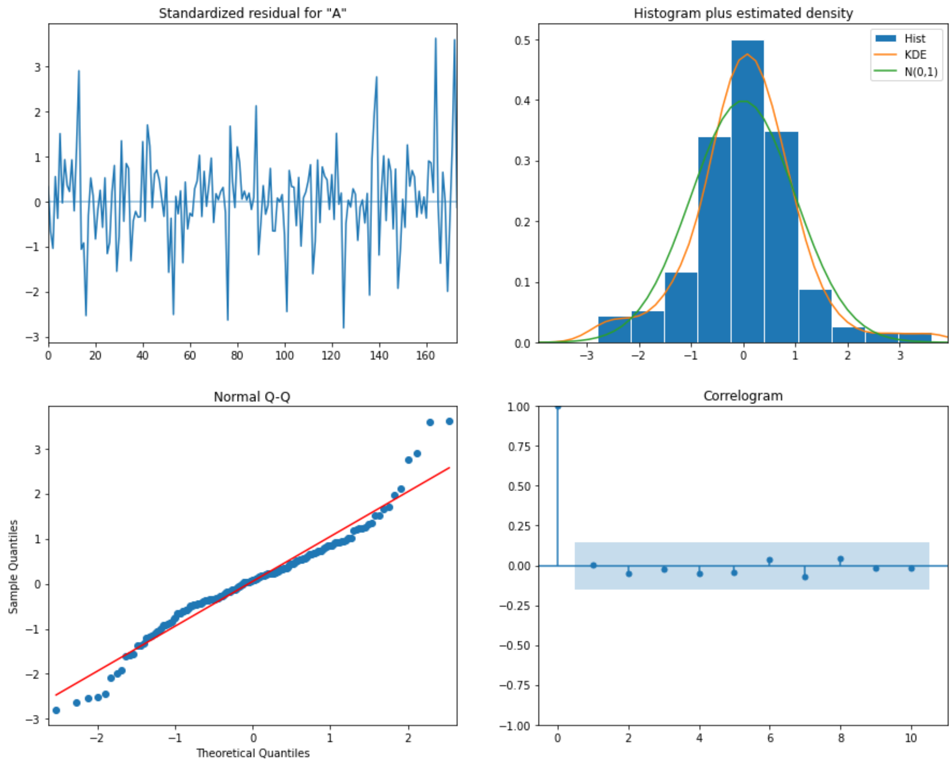



Figure 3.22: Diagnostic check of SARIMA(2,1,2)(2,0,0)

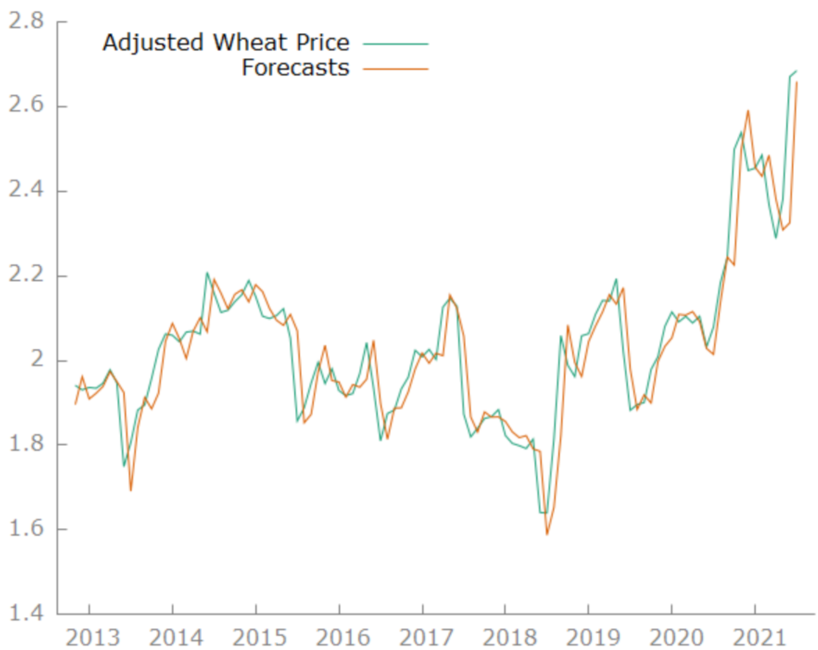



Figure 3.23: Actual and forecasting adjusted wheat prices (Static)

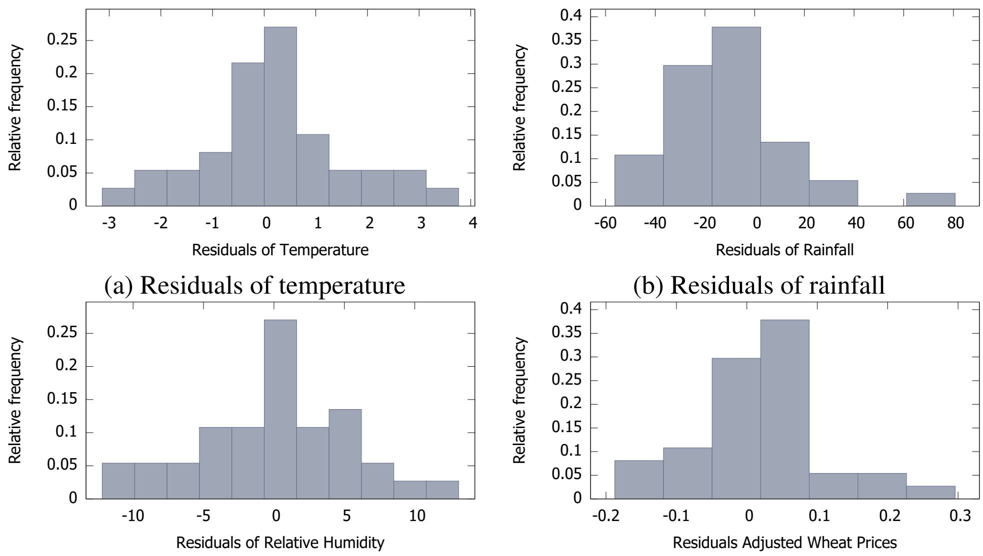
# CHAPTER 4

# COPULA

In this Chapter, copula is used to model interdependence between the probability distribution of temperature, rainfall, humidity, and adjusted wheat prices, whose models are determined as ARIMA (1,0,0)(2,0,0)[12], ARIMA (0,1,1)(0,1,1)[12], Triple Exponential Smoothing and ARIMA(2,1,2)(2,0,0) [12] for the prescribed variables respectively. The prescribed variables are predicted for the test set, then the residuals in the test sets of each model are obtained for further analyses.

The analyses are done using KDECOPULA and VINECOPULA libraries in RSTUDIO. With the help of this library, Frank, Gumbel, Clayton, and student t-copulas can be applied to data sets, and densities and random sets can be generated.

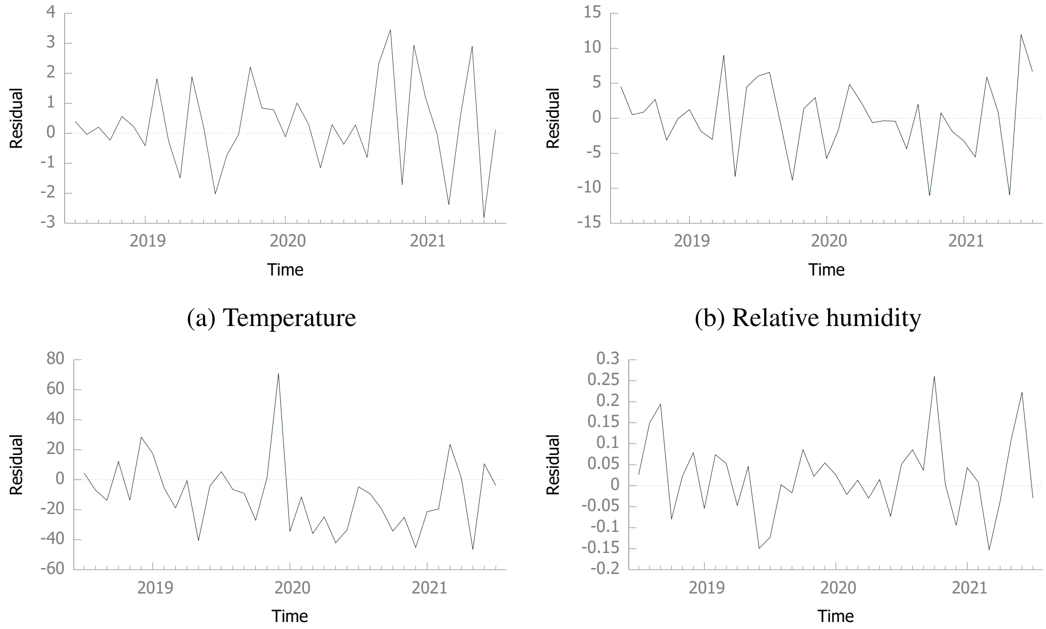
Residuals from suggested ARIMA and TES models are obtained, and then the JB test is applied to understand whether the datasets are normally distributed or not. According to the results, which are shown in Figure 4.1, all residuals are normally distributed except residuals of the precipitation dataset.




(a) Residuals of temperature
(b) Residuals of rainfall
(c) Residuals of relative humidity
(d) Residuals of adjusted wheat price

Figure 4.1: Histogram of residuals

The results of the KPSS test indicate that the residuals are stationary because the p-values are greater than 0.05. The stationary time series are shown in Figure 4.2 at which the movements of residuals, which are the differences between actual and fitted values for climate components and adjusted wheat prices, can be followed visually. The plots show that the values fluctuate around zero.




(a) Temperature (b) Relative humidity

(c) Precipitation (d) Adjusted wheat price

Figure 4.2: Time series plot of residuals

The first step of the copula part is to measure the correlation between temperature, precipitation, relative humidity, and adjusted wheat prices. First, Pearson, Spearman, and Kendall coefficient values are found, and correlation matrices are obtained. Pearson correlation matrix in the Figure 4.3 shows that correlation is the largest between the residuals of temperature and relative humidity, even though it is less than 0.80. There is a moderate negative correlation between temperature and relative humidity residuals. Also, there is a moderate positive correlation between relative humidity and precipitation residuals. Furthermore, the correlation between the residuals of precipitation and relative prices is weak.

A correlation coefficient of normal random variables is dependent on the marginal distributions. Even though perfect dependence is observed for two normal variates, the correlation coefficient can have values lower than 1. As mentioned in the methodology, the linear correlation coefficient gives us the correct measure if and only if these variables are jointly elliptically distributed. Moreover, the definition of the linear correlation coefficient is valid for pairs of variables with finite variance. Thus, the means of copula is used to distinguish the marginal distributions from the dependence structure between the random variables.

In order to make the analyses with copula, the following steps are applied:

i. Pseudo-observations which create normalize ranked data is obtained through elementwise transformation that is applied to residual variables.

ii. Copula model, which is the joint density function of the transformed variables are predicted by using the matrix of pseudo-uniform variables.

iii. Generation of data sets for the pseudo-uniformly distributed variables for the relative humidity and wheat prices

iv. Adjusted wheat prices are simulated.

v. The performance of ARIMA model and copula model is compared.

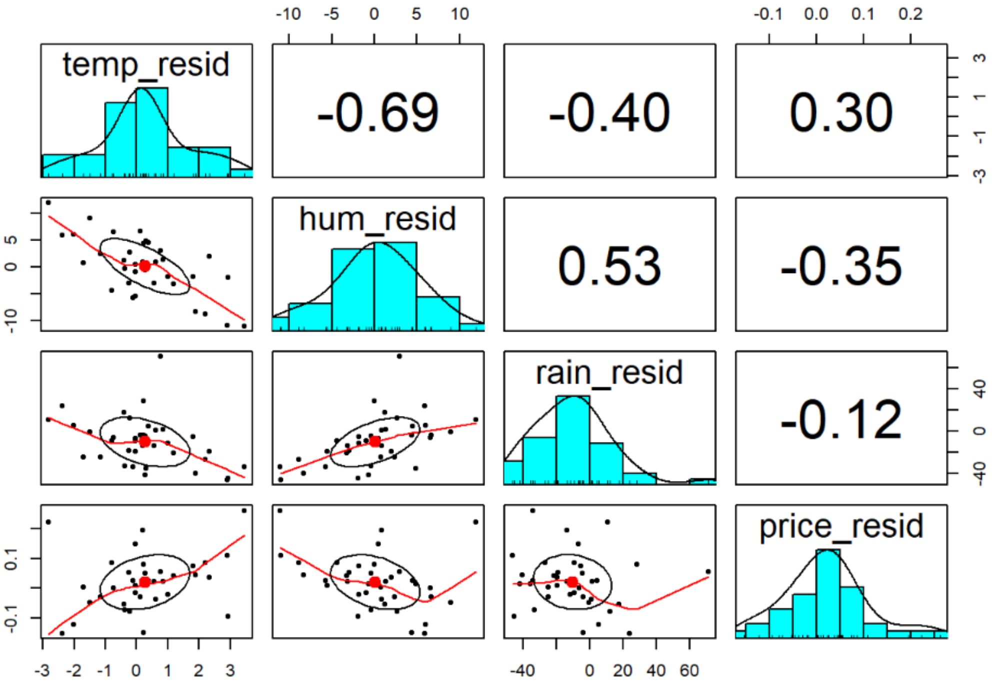



Figure 4.3: Pairs panel and copula structures of variables

For the copula analyses, parameters of the time series of climate components are estimated, and the interdependence relation between the climate components is found. The parameters and best models are shown in the previous chapter, and the residuals are obtained from these models. As seen in the Figure 4.2, residuals are stationary.

Then, the pseudo-uniform variables of the residuals for temperature, relative humidity, and precipitation are obtained by using the pobs function of Vine copula package of RSTUDIO. In the next block of the statements, parameters for various families are estimated, and the best family for each pair is selected with RVINESTRUCTURESELECT function.

As a result, the C-Vine copula is selected as the best copula family. The results of RVINESTRUCTURESELECT function are shown in Figure 4.4. Results in Figure 4.4 demonstrate that Survival BB7 and Survival Gumbel are the most suitable copulas, which model dependence structure for the pairs for relative humidity-temperature and relative humidity-precipitation, respectively. The upper tail dependence coefficient of a copula means the the lower tail dependence of its survival [21]. Therefore, we can say that the survival BB7 copula (rotated 180 degrees), i.e, survival Clayton-Joe copula, enables us to model the dependence in both the upper and the lower tail. Also, Survival Gumbel copula is useful while modeling lower tail dependence. On the other hand, Joe (Survival Joe) copula is useful to model the upper (lower) tails.

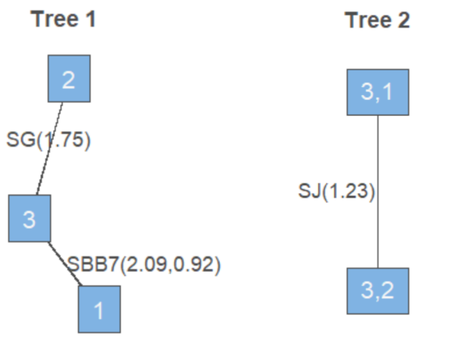
```
C-vine copula with the following pair-copulas:
Tree 1:
3,1  Survival BB7 (par = 2.09, par2 = 0.92, tau = 0.5)
3,2  Survival Gumbel (par = 1.75, tau = 0.43)

Tree 2:
2,1;3  Survival Joe (par = 1.23, tau = 0.12)

---
1 <-> rtemp,   2 <-> rrain,   3 <-> rhum
```

Figure 4.4: Copula family for climate components.

There are three trees for three variables constructed. These are the residuals of temperature, rainfall, and relative humidity. The graphical representation of the classification, called vine, is shown in the Figure 4.5. The vines are represented with $k(k-1)/2$ pair-copulas with $k-1$ trees, where k represents the number of variables. Since we have three variables for the copula analyses, three pair copulas and two trees are shown in the Figure 4.5. In Figure 4.4, the residuals of temperature, precipitation and relative humidity are represented with numbers 1,2, and 3, respectively. Figure 4.5 demonstrates that Survival BB07 copula, is used to model the dependence structure for temperature and relative humidity, and survival Gumbel copula is used to model relative humidity and precipitation. Then, conditioned on residuals of relative humidity, residuals of temperature and precipitation are modeled with survival Joe in the Figure 4.6.




(a) First tree (b) Second tree

Figure 4.5: Copula trees

Residuals of relative humidity are at the core of the dependence modeling between the residuals of climate components. For this reason, we take the residuals of relative humidity as a core component while modeling the dependency. Instead of adding these three variables into the copula analyses, we can use only the residuals of relative humidity to improve wheat prices'

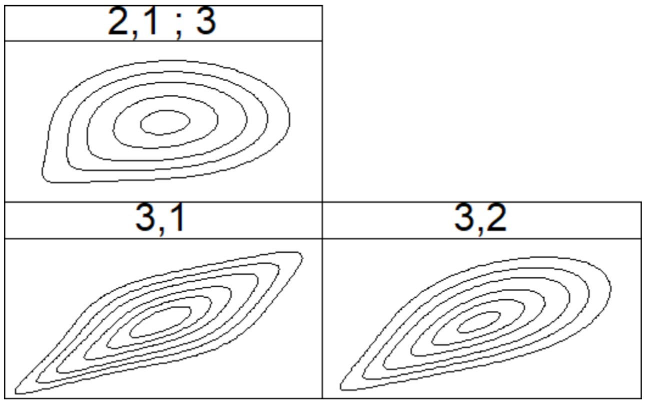



Figure 4.6: Copula climate contour tree

predictions.

To find the best bivariate copula family function, BICOPSELECT() is used in RSTUDIO. This function gives us the best copula family with its corresponding parameters obtained by maximum likelihood estimation. The results show that the dependence relation between the residuals of the relative humidity and adjusted wheat prices are explained better with t-copula. Figure 4.7 shows that the best $rho$ parameter is 0.47 with with $df$ 2.82.

```
```{r}
library(VineCopula)
u <- pobs(as.matrix(cbind(hum_resid,price_resid)))[,1]
v <- pobs(as.matrix(cbind(hum_resid,price_resid)))[,2]
selectedCopula <- BiCopSelect(u,v,familyset=2)
selectedCopula
```

Bivariate copula: t (par = -0.47, par2 = 2.82, tau = -0.31)
```

Figure 4.7: Bivariate copula selection for relative humidity and adjusted Wheat Price

Therefore, to analyze the dependence relation between residuals of relative humidity and wheat price, t-copula is used. Firstly, a random sample of 40 observations is generated based on the size of the test set. The correlation of the random sample is shown in Figure 4.1, there is a weak negative correlation between the actual values from student t-copula.

Table 4.1: Spearman correlation matrix for residuals

| | Relative Humidity | Adjusted Price |
|---|---|---|
| Relative Humidity | 1.00 | -0.52 |
| Adjusted Price | -0.52 | 1.00 |

We observe a low correlation between the observations coming from t-copula. We can validate it by looking at the random samples from the t-copula, which indicates that the residuals are distributed independently.

After we sample from the t-copula, we transform the margins individually by using MVDC() function RSTUDIO which enables us to generate random values for a multivariate distribution via copula.

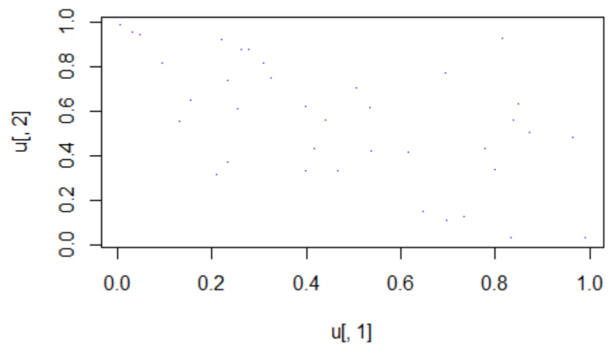



Figure 4.8: The random samples from the copula

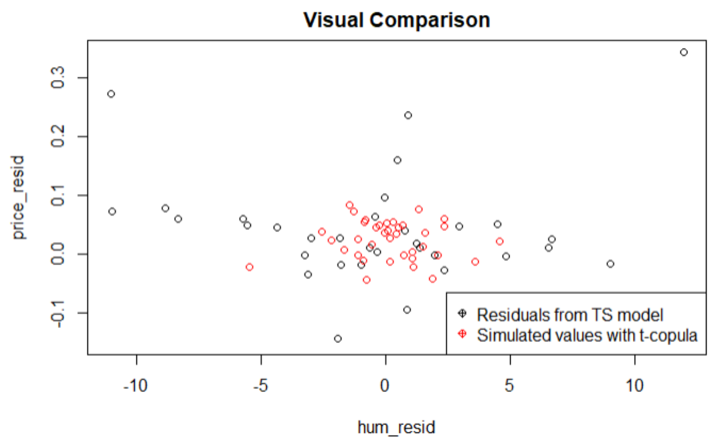



Figure 4.9: Visual comparison of residuals residuals and simulated values

We generate 100 samples for residuals, each has the size of the test set with t-copula, but as seen in Figure 4.10, especially extreme points could not be fully captured, even though the selected copula function is the best fitting one. The simulated observations are shown in Figure A.10.

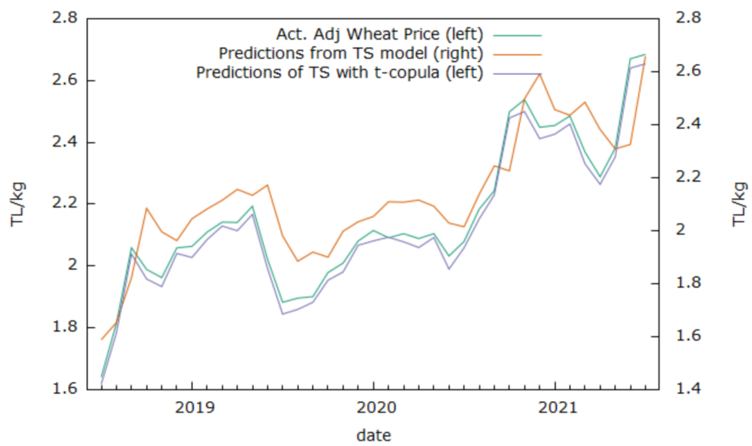



Figure 4.10: Time series of adjusted wheat prices with different models

Finally, we estimate the adjusted price from the residuals simulated by the t-copula. The time series of the adjusted wheat prices are shown in Figure 4.10.

The mean absolute error (MAE) of the residuals simulated with t-copula is found as 0.04, while the MAE value from the ARIMA(2,1,2)(2,0,0) is equal to 0.07. Since the sum of absolute residuals with the same sample size is lower while modeling with t-copula, it is concluded that the copula approach gives more accurate results.

# CHAPTER 5

# CONCLUSION

The future of agricultural products has become an increasingly considered topic in the world as well as in Türkiye. Wars, economic crises, and epidemics have fueled these discussions. In addition to being a food substance that feeds living things, agricultural commodities have also been contracted in financial markets and subjected to transactions. For this reason, the factors mentioned above affect all areas of life. In this study, the dependence relationship between climatic parameters and agricultural products in the financial market is modeled together.

Within the scope of this study, Konya is selected as the representative city for Türkiye, and the interdependence relationships between temperature, relative humidity, rainfall, and spot adjusted wheat prices are modeled with the copula approach.

This thesis is one of the first ones in the literature that estimates the local adjusted spot wheat prices under the impact of climate components in Türkiye. Each region has its own unique climatic conditions. Therefore, when a price estimation study is made for the whole Türkiye, the effect of local climatic conditions would be ignored. Through this study, we see how local climatic factors shape the local spot price.

First of all, climatic factors are modeled using time series or exponential smoothing methods, and the best models are selected by estimating these models. Then, time series analysis is also conducted for spot-adjusted wheat prices, and it is found that the best model is SARIMA(2,1,2)(2,0,0). The model results demonstrate that the forecasts are close to reality with error without the impact of climate components. To minimize the errors, we have used the copula approach. In the next step, instead of modeling all climate components and spot prices together, we aim to find the residuals of which climate component explains the residuals of others under the dependence structure. Because the results imply that relative humidity is the star variable among the other climate components, we modeled the residuals of the spot wheat prices and the relative humidity with t-copula. This model has been used to predict the spot-adjusted wheat prices with minimized errors. The obtained results show that, with the copula analyses, the predictions are much closer to the actual values, with much lower error rates.

The product whose price is estimated in this thesis is a financial asset. Therefore, the results show that the prices of financial assets whose underlying assets are agricultural products give better results when the prices of products are modeled with the copula approach. These results could be beneficial for the experts who are interested in investing in these instruments and the policy-makers.

However, in this work, because the dataset size does not cover the entire 2021, we could not consider the impact of drought observed, especially in the 2021 harvest season, in price estimation. Thus, we could not see if the drought had a significant impact on pricing.

Last but not least, in further studies, it can be fruitful to focus on premium estimation for agricultural insurance contracts by using the commodity prices, which are influenced by the climate component. Furthermore, the impact of climatic parameters can be modeled daily for several cities for Türkiye and wider date intervals. Thanks to it, the result of the recent drought on the spot wheat pricing can be considered.

# REFERENCES


[1] 2022 Dönemi Hububat Satış Fiyatları, https://www.tmo.gov.tr/hububat/1/bugday-arpabr-cavdar-yulaf/106/2022-donemi-hububat-satis-fiyatlari, accessed: 2022-04-26.

[2] KMEX Wheat Prices, http://www.ktb.org.tr/?cat=50, accessed: 2022-04-26.

[3] Tarım Ürünleri Piyasaları, https://arastirma.tarimorman.gov.tr/tepge/Belgeler/PDF%20Tar%C4%B1m%20%C3%9Cr%C3%BCnleri%20Piyasalar%C4%B1/2021-Haziran%20Tar%C4%B1m%20%C3%9Cr%C3%BCnleri%20Raporu/Bu%C4%9Fday,%20Haziran-2021,%20Tar%C4%B1m%20%C3%9Cr%C3%BCnleri%20Piyasa%20Raporu,%20TEPGE.pdf, accessed: 2022-04-26.

[4] The Importance Of Ukraine And The Russian Federation For Global Agricultural Markets And The Risks Associated With The War In Ukraine, https://www.fao.org/3/cb9013en/cb9013en.pdf, accessed: 2022-04-26.

[5] Wheat pulls back from 13-1/2 year peak as Ukraine conflict assessed, https://www.reuters.com/business/energy/wheat-pulls-back-13-12-year-peak-ukraine-conflict, accessed: 2022-04-26.

[6] K. Aas, C. Czado, A. Frigessi, and H. Bakken, Pair-copula constructions of multiple dependence, Insurance: Mathematics and economics, 44(2), pp. 182–198, 2009.

[7] D. E. Allen, M. McAleer, and A. K. Singh, Risk measurement and risk modelling using applications of vine copulas, Sustainability, 9(10), p. 1762, 2017.

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

[9] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.

[10] R. G. Brown, Exponential smoothing for predicting demand, in *Operations Research*, volume 5, pp. 145–145, Inst Operations Research Management Sciences 901 Elkridge Landing Rd, Ste, 1957.

[11] U. Cherubini and E. Luciano, Bivariate option pricing with copulas, Applied Mathematical Finance, 9(2), pp. 69–85, 2002.

[12] R.-G. Cong and M. Brady, The interdependence between rainfall and temperature: copula analyses, The Scientific World Journal, 2012.

[13] C. Czado and T. Nagler, Vine copula based modeling, Annual Review of Statistics and Its Application, 9(1), pp. 453–477, 2022.

[14] C. Czado, U. Schepsmeier, and A. Min, Maximum likelihood estimation of mixed c-vines with application to exchange rates, Statistical Modelling, 12(3), pp. 229–255, 2012.

[15] T. M. Erhardt, C. Czado, and U. Schepsmeier, R-vine models for spatial time series with an application to daily mean temperature, Biometrics, 71(2), pp. 323–332, 2015.

[16] O. Evkaya, C. Yozgatligil, and A. Sevtap Selcuk-Kestel, Drought analysis using copula approach: a case study for turkey, Communications in Statistics: Case Studies, Data Analysis and Applications, 5(3), pp. 243–260, 2019.

[17] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, Forecasting of demand using arima model, International Journal of Engineering Business Management, 10, p. 1847979018808673, 2018.

[18] E. W. Frees and E. A. Valdez, Understanding relationships using copulas, North American actuarial journal, 2(1), pp. 1–25, 1998.

[19] L. Garcia-Jorcano and S. Benito, Studying the properties of the bitcoin as a diversifying and hedging asset through a copula analysis: Constant and time-varying, Research in International Business and Finance, 54, p. 101300, 2020.

[20] M. Gronwald, J. Ketterer, and S. Trück, The relationship between carbon, commodity and financial markets: A copula analysis, Economic record, 87, pp. 105–124, 2011.

[21] T. Großmaß, *Copulae and tail dependence*, Master's thesis, Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, 2007.

[22] Z. Haofei, X. Guoping, Y. Fangting, and Y. Han, A neural network model based on the multi-stage optimization approach for short-term food price forecasting in china, Expert Systems with Applications, 33(2), pp. 347–356, 2007.

[23] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting, 20(1), pp. 11–13, 2004.

[24] S. Hylleberg, *Modelling seasonality*, Oxford University Press, 1992.

[25] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with exponential smoothing: the state space approach*, Springer Science & Business Media, 2008.

[26] N. Iqbal, K. Bakhsh, A. Maqbool, and A. S. Ahmad, Use of the arima model for forecasting wheat area and production in pakistan, Journal of Agriculture and Social Sciences, 1(2), pp. 120–122, 2005.

[27] G. Jain and B. Mallick, A study of time series models arima and ets, Available at SSRN 2898968, 2017.

[28] H. Joe, *Multivariate models and multivariate dependence concepts*, CRC press, 1997.

[29] H. Joe, Asymptotic efficiency of the two-stage estimation method for copula-based models, Journal of multivariate Analysis, 94(2), pp. 401–419, 2005.

[30] A. J. Patton, A review of copula models for economic time series, Journal of Multivariate Analysis, 110, pp. 4–18, 2012.

[31] B. Pfaff, *Financial risk modelling and portfolio optimization with R*, John Wiley & Sons, 2016.

[32] A. Sarraf, S. F. Vahdat, and A. Behbahaninia, Relative humidity and mean monthly temperature forecasts in ahwaz station with arima model in time series analysis, in *International Conference on Environment and Industrial Innovation IPCBEE, Singapore*, volume 12, IACSIT Press Singapore, 2011.

[33] J. H. Shih and T. A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics, pp. 1384–1399, 1995.

[34] R. d. S. Silva and H. F. Lopes, Copula, marginal distributions and model selection: a bayesian note, Statistics and Computing, 18(3), pp. 313–320, 2008.

[35] M. Sklar, Fonctions de repartition an dimensions et leurs marges, Publ. inst. statist. univ. Paris, 8, pp. 229–231, 1959.

[36] M. Smith, A. Min, C. Almeida, and C. Czado, Modeling longitudinal data using a pair-copula decomposition of serial dependence, Journal of the American Statistical Association, 105(492), pp. 1467–1479, 2010.

[37] M. S. Smith, Copula modelling of dependence in multivariate time series, International Journal of Forecasting, 31(3), pp. 815–833, 2015.

[38] M. Tektaş, Weather forecasting using anfis and arima models, Environmental Research, Engineering and Management, 51(1), pp. 5–10, 2010.

[39] P. R. Winters, Forecasting sales by exponentially weighted moving averages, Management science, 6(3), pp. 324–342, 1960.

[40] J. J. Xu, *Statistical modelling and inference for multivariate and longitudinal discrete response data*, Ph.D. thesis, University of British Columbia, 1996.

# APPENDIX A

# DETAILED OUTCOMES OF TIME SERIES MODELS AND COPULA

## A.1 Summaries of Alternative Models Climate Components

Table A.1: ARIMA(1,0,0)(2,0,0)[12] model for temperature

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | 1.91775 | 0.340836 | 5.627 | 0.0000 |
| $\phi_1$ | 0.352448 | 0.0774340 | 4.552 | 0.0000 |
| $\Phi_1$ | −0.365462 | 0.0843782 | −4.331 | 0.0000 |
| $\Phi_2$ | −0.188427 | 0.0884497 | −2.130 | 0.0331 |
| dm1 | −1.69177 | 0.384673 | −4.398 | 0.0000 |
| dm2 | 0.528199 | 0.454136 | 1.163 | 0.2448 |
| dm3 | 5.01402 | 0.473850 | 10.58 | 0.0000 |
| dm4 | 9.76344 | 0.479742 | 20.35 | 0.0000 |
| dm5 | 14.2818 | 0.481996 | 29.63 | 0.0000 |
| dm6 | 19.0979 | 0.482370 | 39.59 | 0.0000 |
| dm7 | 22.9979 | 0.481737 | 47.74 | 0.0000 |
| dm8 | 22.7846 | 0.479443 | 47.52 | 0.0000 |
| dm9 | 18.2434 | 0.472652 | 38.60 | 0.0000 |
| dm10 | 11.3824 | 0.452304 | 25.17 | 0.0000 |
| dm11 | 4.97393 | 0.389214 | 12.78 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 12.48000 | S.D. dependent var | 8.821711 |
| Mean of innovations | 0.001800 | S.D. of innovations | 1.655336 |
| $R^2$ | 0.964556 | Adjusted $R^2$ | 0.961039 |
| Log-likelihood | −279.9228 | Akaike criterion | 591.8457 |
| Schwarz criterion | 639.4734 | Hannan–Quinn | 611.1984 |

Table A.2: ARIMA(1,0,1)(1,1,1)[12] model for temperature

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | 0.00326722 | 0.0624555 | 0.05231 | 0.9583 |
| $\phi_1$ | 0.418747 | 0.228014 | 1.836 | 0.0663 |
| $\Phi_1$ | -0.232467 | 0.102696 | -2.264 | 0.0236 |
| $\theta_1$ | −0.0740056 | 0.250220 | −0.2958 | 0.7674 |
| $\Theta_1$ | −0.990525 | 1.92537 | −0.5145 | 0.6069 |

| | | | |
|---|---|---|---|
| Mean dependent var | 0.100000 | S.D. dependent var | 3.205595 |
| Mean of innovations | −0.016888 | S.D. of innovations | 1.819083 |
| $R^2$ | 0.956745 | Adjusted $R^2$ | 0.955681 |
| Log-likelihood | −270.9447 | Akaike criterion | 553.8894 |
| Schwarz criterion | 570.9071 | Hannan–Quinn | 560.8031 |

Table A.3: ARIMA(1,0,1)(1,1,1)[12] model for relative humidity

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | −1.05320 | 0.464444 | −2.268 | 0.0302 |
| $\phi_1$ | 0.820629 | 0.101540 | 8.082 | 0.0000 |
| $\Phi_1$ | −0.106636 | 0.117731 | −0.9058 | 0.3651 |
| $\theta_1$ | −0.596685 | 0.133436 | −4.472 | 0.0000 |
| $\Theta_1$ | −0.708780 | 0.0998111 | −7.101 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | −0.871429 | S.D. dependent var | 9.652862 |
| Mean of innovations | −0.088862 | S.D. of innovations | 6.720879 |
| $R^2$ | 0.8587 | Adjusted $R^2$ | 0.852691 |
| Log-likelihood | −424.0019 | Akaike criterion | 860.0037 |
| Schwarz criterion | 877.0214 | Hannan–Quinn | 866.9175 |

Table A.4: ARIMA(1,0,1)(1,0,1)[12] model for relative humidity

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | 60.6095 | 10.2293 | 5.925 | 0.0000 |
| $\phi_1$ | 0.804112 | 0.0963681 | 8.344 | 0.0000 |
| $\Phi_1$ | 0.990238 | 0.00727383 | 136.1 | 0.0000 |
| $\theta_1$ | −0.479535 | 0.146654 | −3.270 | 0.0011 |
| $\Theta_1$ | −0.726436 | 0.0774553 | −9.379 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 60.17029 | S.D. dependent var | 17.76185 |
| Mean of innovations | −0.956588 | S.D. of innovations | 6.883648 |
| $R^2$ | 0.851965 | Adjusted $R^2$ | 0.848651 |
| Log-likelihood | −475.1163 | Akaike criterion | 962.2327 |
| Schwarz criterion | 979.7962 | Hannan–Quinn | 969.3701 |

Table A.5: Summary statistics of ARIMA(1,0,1) for precipitation

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | 30.1646 | 2.38665 | 12.64 | 0.0000 |
| $\phi_1$ | $-0.0217495$ | 0.483656 | $-0.04497$ | 0.9641 |
| $\theta_1$ | 0.136840 | 0.473874 | 0.2888 | 0.7728 |

| | | | |
|---|---|---|---|
| Mean dependent var | 30.14203 | S.D. dependent var | 25.47352 |
| Mean of innovations | 0.004787 | S.D. of innovations | 25.21437 |
| $R^2$ | 0.013092 | Adjusted $R^2$ | 0.005836 |
| Log-likelihood | $-641.2034$ | Akaike criterion | 1290.407 |
| Schwarz criterion | 1302.116 | Hannan–Quinn | 1295.165 |

Table A.6: Summary statistics of SARIMA(0,1,1)(0,1,1)[12] for precipitation

| | Coefficient | Standart Error | z | p-value |
|---|---|---|---|---|
| const | $-0.0218715$ | 0.0330799 | $-0.6612$ | 0.5085 |
| $\theta_1$ | $-1.00000$ | 0.0558671 | $-17.90$ | 0.0000 |
| $\Theta_1$ | $-0.849916$ | 0.110514 | $-7.691$ | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 0.443200 | S.D. dependent var | 44.39604 |
| Mean of innovations | $-1.360460$ | S.D. of innovations | 23.76052 |
| $R^2$ | 0.165685 | Adjusted $R^2$ | 0.158902 |
| Log-likelihood | $-584.5045$ | Akaike criterion | 1177.009 |
| Schwarz criterion | 1188.322 | Hannan–Quinn | 1181.605 |

Table A.7: Summary statistics of SARIMA(1,1,1)(1,1,1)[12] for precipitation

| | Coefficients | Standart Error | z | p-value |
|---|---|---|---|---|
| const | $-0.0218139$ | 0.0309329 | $-0.7052$ | 0.4807 |
| $\phi_1$ | $-0.0290206$ | 0.0925122 | $-0.3137$ | 0.7538 |
| $\Phi_1$ | $-0.0776133$ | 0.119227 | $-0.6510$ | 0.5151 |
| $\theta_1$ | $-0.999999$ | 0.0612930 | $-16.32$ | 0.0000 |
| $\Theta_1$ | $-0.806618$ | 0.126990 | $-6.352$ | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 0.443200 | S.D. dependent var | 44.39604 |
| Mean of innovations | $-1.476286$ | S.D. of innovations | 23.80595 |
| $R^2$ | 0.164577 | Adjusted $R^2$ | 0.143864 |
| Log-likelihood | $-584.2549$ | Akaike criterion | 1180.510 |
| Schwarz criterion | 1197.480 | Hannan–Quinn | 1187.404 |

Table A.8: Summary statistics of SARIMA(0,0,1)(0,1,1)[12] for precipitation

| | Coefficients | Standart Error | z | p-value |
|---|---|---|---|---|
| const | 0.992575 | 0.600528 | 1.653 | 0.0984 |
| $\theta_1$ | −0.0390902 | 0.0974058 | −0.4013 | 0.6882 |
| $\Theta_1$ | −0.880666 | 0.131438 | −6.700 | 0.0000 |

| | | | |
|---|---|---|---|
| Mean dependent var | 1.484127 | S.D. dependent var | 32.09498 |
| Mean of innovations | 0.817707 | S.D. of innovations | 23.48579 |
| $R^2$ | 0.181346 | Adjusted $R^2$ | 0.174743 |
| Log-likelihood | −585.1235 | Akaike criterion | 1178.247 |
| Schwarz criterion | 1189.592 | Hannan–Quinn | 1182.856 |

## A.2 Residuals of Temperature, Relative Humidity, Precipitation and Adjusted Wheat Prices

Table A.9: Residuals of temperature, relative humidity, precipitation and adjusted wheat prices

| Date | Temperature | Relative Humidity | Precipitation | Adjusted Wheat Prices |
|---|---|---|---|---|
| 1.07.2018 | -0.30 | 4.51 | 4.43 | 0.05 |
| 1.08.2018 | 0.20 | 0.49 | -6.97 | 0.16 |
| 1.09.2018 | -1.00 | 0.88 | -13.65 | 0.24 |
| 1.10.2018 | -0.20 | 2.70 | 12.06 | -0.10 |
| 1.11.2018 | -0.40 | -3.11 | -13.59 | -0.03 |
| 1.12.2018 | -0.40 | -0.05 | 28.36 | 0.10 |
| 1.01.2019 | 0.50 | 1.23 | 17.45 | 0.02 |
| 1.02.2019 | -2.90 | -1.85 | -5.54 | 0.03 |
| 1.03.2019 | 0.10 | -3.01 | -18.83 | 0.03 |
| 1.04.2019 | 1.40 | 9.01 | -0.66 | -0.02 |
| 1.05.2019 | -2.40 | -8.32 | -40.55 | 0.06 |
| 1.06.2019 | 0.10 | 4.44 | -4.15 | -0.15 |
| 1.07.2019 | 2.20 | 6.01 | 5.38 | -0.10 |
| 1.08.2019 | 0.70 | 6.56 | -6.43 | 0.01 |
| 1.09.2019 | -0.30 | -0.97 | -9.11 | -0.02 |
| 1.10.2019 | -2.90 | -8.85 | -27.00 | 0.08 |
| 1.11.2019 | -0.70 | 1.36 | 1.55 | 0.01 |
| 1.12.2019 | -0.80 | 2.95 | 70.90 | 0.05 |
| 1.01.2020 | 0.50 | -5.73 | -34.41 | 0.06 |
| 1.02.2020 | -1.70 | -1.78 | -11.59 | -0.02 |
| 1.03.2020 | -0.10 | 4.84 | -35.88 | 0.00 |
| 1.04.2020 | 1.60 | 2.34 | -24.72 | -0.03 |
| 1.05.2020 | -1.20 | -0.62 | -42.01 | 0.01 |
| 1.06.2020 | 0.70 | -0.35 | -33.41 | 0.00 |
| 1.07.2020 | 0.20 | -0.43 | -4.68 | 0.06 |
| 1.08.2020 | 1.10 | -4.37 | -9.29 | 0.05 |
| 1.09.2020 | -2.80 | 1.99 | -19.36 | 0.00 |
| 1.10.2020 | -4.40 | -11.04 | -34.25 | 0.27 |
| 1.11.2020 | 1.80 | 0.75 | -25.10 | 0.04 |
| 1.12.2020 | -3.40 | -1.92 | -45.16 | -0.14 |
| 1.01.2021 | -0.90 | -3.24 | -21.27 | 0.00 |
| 1.02.2021 | -0.60 | -5.54 | -19.45 | 0.05 |
| 1.03.2021 | 2.30 | 5.91 | 23.66 | -0.12 |
| 1.04.2021 | -0.30 | 0.88 | 0.83 | -0.09 |
| 1.05.2021 | -3.60 | -10.97 | -46.47 | 0.07 |
| 1.06.2021 | 3.30 | 11.97 | 10.53 | 0.35 |
| 1.07.2021 | -0.10 | 6.69 | -3.74 | 0.03 |

## A.3 Simulation results from t-copula

Table A.10: Simulations of t-copula

| Simulatations of Residual of Relative Humidity | Simulations of Residual of Adjusted Wheat Price |
|---|---|
| -8.83 | 0.03 |
| 1.13 | 0.02 |
| 2.02 | 0.00 |
| -1.15 | 0.02 |
| -1.79 | -0.01 |
| -3.39 | 0.00 |
| -4.71 | 0.02 |
| -0.90 | -0.03 |
| 1.31 | -0.09 |
| -0.27 | -0.06 |
| 9.94 | -0.07 |
| 4.55 | 0.01 |
| 5.01 | -0.08 |
| -8.02 | 0.09 |
| -5.40 | 0.17 |
| 2.79 | 0.13 |
| 2.56 | 0.03 |
| -5.49 | -0.04 |
| -5.93 | 0.12 |
| 3.62 | -0.04 |
| 2.24 | 0.01 |
| -5.25 | 0.11 |
| 1.13 | -0.08 |
| 9.07 | -0.10 |
| 2.90 | -0.05 |
| -3.68 | -0.04 |
| 3.53 | -0.06 |
| -2.69 | -0.10 |
| 1.69 | 0.10 |
| 3.57 | 0.00 |
| 2.69 | 0.10 |
| 1.06 | 0.02 |
| 15.11 | -0.22 |
| -10.25 | 0.08 |
| -4.62 | 0.00 |
| -3.80 | 0.11 |
| 1.34 | 0.07 |