Time Series Analysis and Forecasting of COVID-19 Trends in Coffee County, Tennessee, United States

[1]Cyril Neba C.; [2]Gerard Shu F.; [3]Gillian Nsuh; [4]Philip Amouda A. , [5]Adrian Neba F., [6]Aderonke Adebisi, [7]P. Kibet., [8]F.Webnda

[1,4,5,6,7,8] Department of Mathematics and Computer Science, Austin Peay State University, Clarksville, Tennessee, USA
[2] Montana State University, Gianforte School of Computing, Bozeman, Monatana, USA
[3] School of Business Analytics, University of Quinnipiac, Hamden, Connecticut

Abstract:- The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has had a profound impact globally, including in the United States and Coffee County, Tennessee. This research project delves into the multifaceted effects of the pandemic on public health, the economy, and society. We employ time series analysis and forecasting methods to gain insights into the trajectory of COVID-19 cases specifically within Coffee County, Tennessee. The United States has witnessed significant repercussions from the COVID-19 pandemic, including public health crises, economic disruptions, and healthcare system strains. Vulnerable populations have been disproportionately affected, leading to disparities in health outcomes. Mental health challenges have also emerged. Accurate forecasting of COVID-19 cases is crucial for informed decision-making. Disease forecasting relies on time series models to analyze historical data and predict future trends. We discuss various modeling approaches, including epidemiological models, data-driven methods, hybrid models, and statistical time series models. These models play a vital role in public health planning and resource allocation. We employ ARIMA, AR, MA, Holt's Exponential Smoothing, and GARCH models to analyze the time series data of COVID-19 cases in Coffee County. The selection of the best model is based on goodness-of-fit indicators, specifically the AIC and BIC. Lower AIC and BIC values are favored as they indicate better model fit. The dataset for this research project was sourced from the Tennessee Department of Health and spans from 12/03/2020 to 12/11/2022. It comprises records of all Tennessee counties, including variables such as date, total cases, new cases, total confirmed, new confirmed, total probable, and more. Our analysis focuses on Coffee County, emphasizing County, Date, and Total cases. Among the models considered, the GARCH model proves to be the most suitable for forecasting COVID-19 cases in Coffee County, Tennessee. This conclusion is drawn from the model's lowest AIC values compared to ARIMA and Holt's Exponential Smoothing. Additionally, the GARCH model's residuals exhibit a distribution closer to normalcy. Hence, for this specific time series data, the GARCH model outperforms ARIMA, AR, MA, and Holt's Exponential Smoothing in terms of predictive accuracy and goodness of fit.

Keywords:- *COVID-19, Time Series Analysis, Disease Forecasting, ARIMA, AR, MA, Holt's Exponential Smoothing, GARCH, AIC, BIC.*

I. INTRODUCTION

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, emerged in late 2019 and quickly evolved into a global health crisis, profoundly impacting the United States and countries worldwide. This infectious disease, characterized by respiratory symptoms and a range of severity, prompted unprecedented public health responses, and introduced new challenges across various sectors. The pandemic has prompted extensive research, data analysis, and modeling efforts aimed at understanding its dynamics, forecasting its course, and developing effective strategies for mitigation and response.

A. *Impact of COVID-19 in the United States:*

The United States has experienced a significant and multifaceted impact from the COVID-19 pandemic. In addition to its severe toll on public health, the pandemic has had far-reaching economic, social, and healthcare consequences. The virus has strained healthcare systems, leading to shortages of medical supplies and intensive care unit (ICU) capacity in some regions. The economic repercussions have included job losses, business closures, and supply chain disruptions, affecting livelihoods and economic stability [7].

Moreover, the pandemic has exposed and exacerbated existing health disparities, disproportionately affecting vulnerable populations, including racial and ethnic minorities and individuals with preexisting health conditions [10]. The pandemic's impact on mental health has also been profound, with increased rates of anxiety, depression, and stress-related disorders reported among individuals of all ages [6].

Forecasting the trajectory of the COVID-19 pandemic is essential for informing public health interventions and resource allocation. Various methods and models have been employed for this purpose:

➢ *Epidemiological Models:*
Epidemiological models, including the Susceptible-Infected-Recovered (SIR) and its variants, have been used to simulate the spread of COVID-19. These models consider factors such as transmission rates, population susceptibility, and intervention measures to project future cases and deaths [4].

➢ *Data-Driven Models:*

Data-driven approaches use historical and real-time data to predict COVID-19 trends. Machine learning techniques, such as regression analysis, decision trees, and neural networks, ca capture complex relationships between variables and make short-term forecasts [3].

➢ *Hybrid Models:*

Some forecasting efforts combine epidemiological models with data-driven techniques to improve accuracy. These hybrid models aim to incorporate both the mechanistic understanding of disease spread and real-world data [1].

➢ *Statistical Time Series Models:*

Time series analysis, including methods like autoregressive integrated moving average (ARIMA) and seasonal decomposition, has been employed to forecast COVID-19 cases and deaths based on historical patterns and trends [8].

➢ *Machine Learning Ensembles:*

Ensemble methods, such as Random Forest and Gradient Boosting, have been used to enhance prediction accuracy by combining multiple models or model variants [12].

As the COVID-19 situation continues to evolve, ongoing research and refinements in forecasting techniques are critical for providing timely and accurate information to guide public health responses and decision-making.

➢ *Covid 19 In Coffee County, Tennessee, United States*

Based on media publications, Coffee County was one of the Counties with highest COVID-19 infection rates in Tennessee. We will therefore do a detailed analysis and make predictions for the total number of infections in this county, which would allow for better forecasting and a clearer picture of what the situation will be like in the coming months or years. For this project, we will use different time series models to forecast and compare amongst these models to see which model is the most appropriate or best.

➢ *Background On Disease Forecasting*

Time series models play a crucial role in disease forecasting, providing valuable insights into the past, present, and potential future trends of infectious diseases. These models use historical data on the disease's occurrence and spread to make predictions about its future trajectory. Here is a brief discussion with in-text citations on time series models for disease forecasting:

➢ *Historical Perspective:*

Time series analysis for disease forecasting has a rich history. Early applications include the work of [4] who developed compartmental models for infectious diseases. Time series data on disease cases have been used for decades to understand patterns, seasonality, and long-term trends [9].

➢ *Data Characteristics:*

• Time series data for disease forecasting typically involve daily, weekly, or monthly counts of cases or incidences.

• These data may exhibit various patterns, including seasonality, trends, and irregular fluctuations [5].

B. *Methods for Disease Forecasting:*

➢ *ARIMA Models:*

Autoregressive Integrated Moving Average (ARIMA) models are widely used in disease forecasting [5] can capture trends, seasonality, and autocorrelation in time series data.

➢ *Exponential Smoothing:*

Exponential smoothing methods, such as Holt-Winters, are effective for capturing seasonality and trends in disease data [5].

➢ *Machine Learning:*

Machine learning techniques, including regression, decision trees, and neural networks, have been applied to disease forecasting [1].

➢ *Hybrid Models:*

Some studies combine traditional time series models with machine learning approaches for improved accuracy [3].

➢ *Applications:*

Disease forecasting models have been applied to various infectious diseases, including influenza, COVID-19, and vector-borne diseases like dengue and malaria. These models are used by public health authorities to plan resource allocation, implement control measures, and communicate risk to the public [8].

➢ *Challenges:*

Disease forecasting is inherently uncertain due to factors like changing human behavior, vaccine distribution, and new variants. Models must adapt to evolving conditions and incorporate real-time data [5].

➢ *Real-world Impact:*

Accurate disease forecasting can help authorities prepare for outbreaks, allocate medical resources, and implement timely interventions. For example, during the COVID-19 pandemic, forecasting models guided decisions on lockdowns, mask mandates, and vaccine distribution [8].

II. METHODOLOGY

The following models will be employed for the time series analysis: ARIMA, AR, MA, HOLT EXPONENTIAL SMOOTHING AND GARCH. To analyze and decide on which of our time series model is best, we will use the goodness of fit, specifically the AIC(Akaike Information Criterion) and BIC(Bayesian Information Criterion) indicators. The idea behind using the AIC and BIC indicators is that they penalize the models with more estimated parameters to prevent overfitting hence smaller AIC and BIC values are preferred. All other things being equal, a model with lower AIC and BIC values is a better fit when compared to another model with higher AIC and BIC values.

➢ *Dataset*

From the Tennessee Department of Health website, the dataset for this project was obtained. The dataset ranges from 12/03/2020 to 12/11/2022. The dataset is made up the following variables: Counties in Tennessee which are character types and Total Cases, New Cases, Total Confirmed, New Confirmed, Total Confirmed, New Confirmed, Total Probable, New Probable, Post Test, New Post Tests, Negative Test, New Negative Tests, Total Test, New Tests, New Deaths, Total Deaths, New Recovered, Total Recovered, New Active, Total Active, New Inactive Recovered, Total Inactive Recovered, New Hospitalized and Total Hospitalized which are all integer types.

➢ *Importing Covid-19 Data for Coffee County*

For this study, we imported the data into R studio, uploaded the required packages and then proceeded to convert the data into a time series object for the purpose of the time series analysis. The date variable was converted from character type to date type.
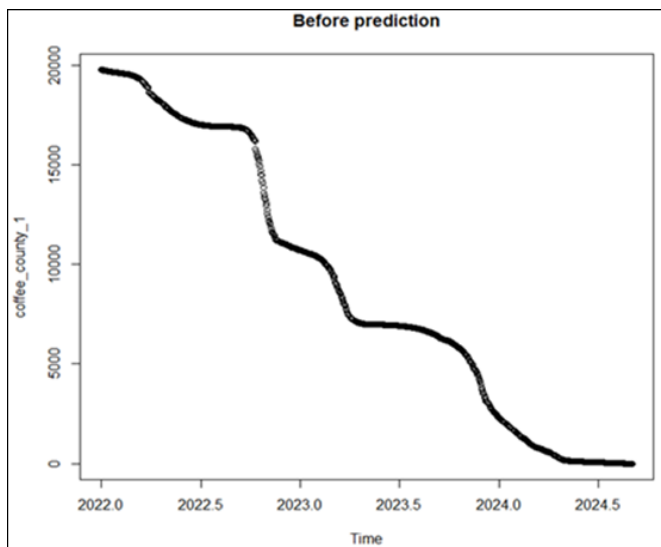
Fig 1 Time Series Plot Depicting COVID-19 Trends in Coffee County before Forecasting

➢ *Conducting Augmented Dickey-Fuller Test to Assess Stationarity of the Time Series*

Augmented Dickey-Fuller Test

data: coffee_county_1

Dickey-Fuller = -3.9558, Lag order = 9, p-value = 0.01121

alternative hypothesis: stationary

Observing the p-value from the Dickey-Fuller Test, we observe that it exceeds 0.01, leading to our decision not to reject the null hypothesis. Consequently, we have sufficient statistical evidence to assert that the time series lacks stationarity. To address this, we proceed with a transformation to induce stationarity.

➢ *First Differencing of Time Series Data*
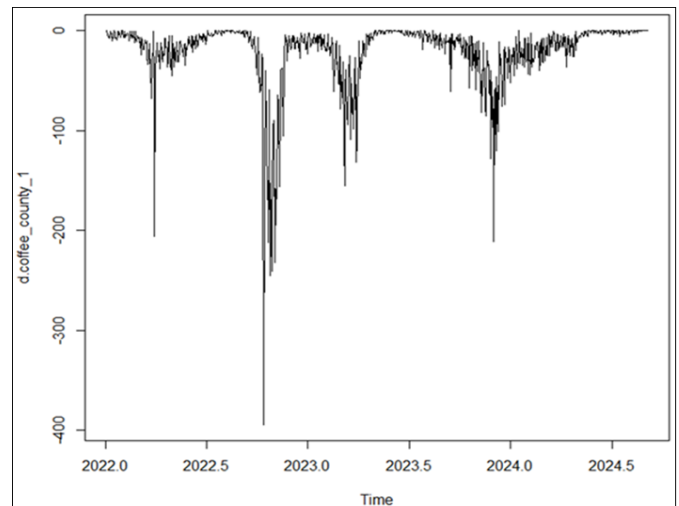
Fig 2 Plot of First Differencing of the Time Series Data

Augmented Dickey-Fuller Test

data: d.coffee_county_1

Dickey-Fuller = -3.1963, Lag order = 9, p-value = 0.08856

alternative hypothesis: stationary

Upon examining the First Differencing plot, it becomes evident that the time series remains non-stationary. This observation is further substantiated by the Dickey-Fuller test, which produces a p-value of 0.08856. As this p-value exceeds 0.01, we refrain from rejecting the null hypothesis, thereby accumulating adequate statistical evidence to affirm that the time series data remains non-stationary. Consequently, our next step involves implementing a second differencing to attain stationarity.

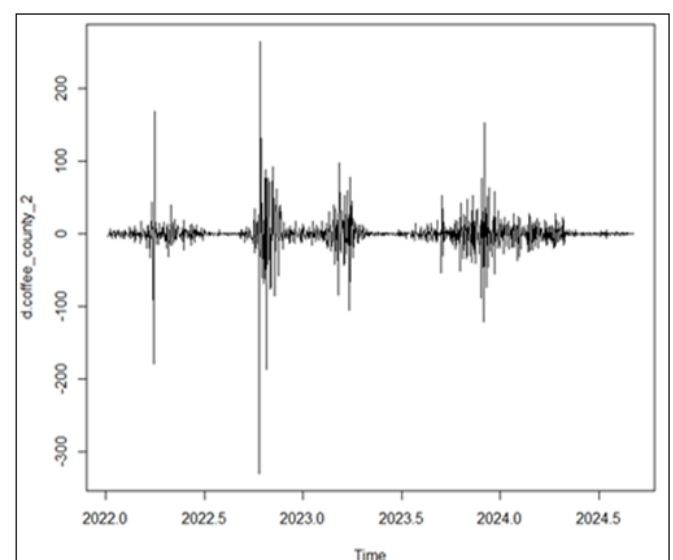➢ *Second Differencing of Time Series Data*

Fig 3 Plot of Second Differencing of the Time Series Data

Augmented Dickey-Fuller Test

data: d.coffee_county_2

Dickey-Fuller = -10.92, Lag order = 9, p-value = 0.01

alternative hypothesis: stationary

Observing the p-value resulting from the Dickey-Fuller Test applied to the second differencing, we note its significance as it falls below the threshold of 0.01. Consequently, we confidently reject the null hypothesis, substantiating the presence of ample statistical evidence to establish the stationarity of the time series. This affirmation enables us to move forward with the construction of our time series model.

Box-Ljung test

data: d.coffee_county_2

X-squared = 204.1, df = 5, p-value < 2.2e-16

Upon examining the p-values obtained from the Ljung-Box test applied to the second differencing of the time series data, it becomes evident that they consistently fall below the significance level of 0.01. Consequently, we confidently reject the null hypothesis, establishing robust statistical evidence supporting the presence of correlation among the residuals. This substantiates our ability to employ ARCH/GARCH models for further analysis.

III. ARIMA MODEL

ARIMA(0,0,5) with zero mean

Coefficients:

	ma1	ma2	ma3	ma4	ma5
	-0.6981	-0.0998	-0.0733	-0.0681	0.2678
s.e.	0.0314	0.0387	0.0397	0.0415	0.0301

sigma^2 = 406.3: log likelihood = -4305.69

AIC=8623.38 AICc=8623.46 BIC=8652.67

➢ *Making Forecasts*
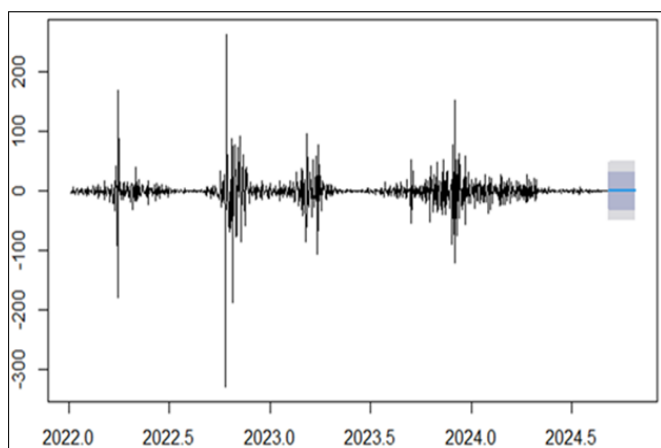
Fig 4 Forecasting COVID-19 Trends in Coffee County with ARIMA
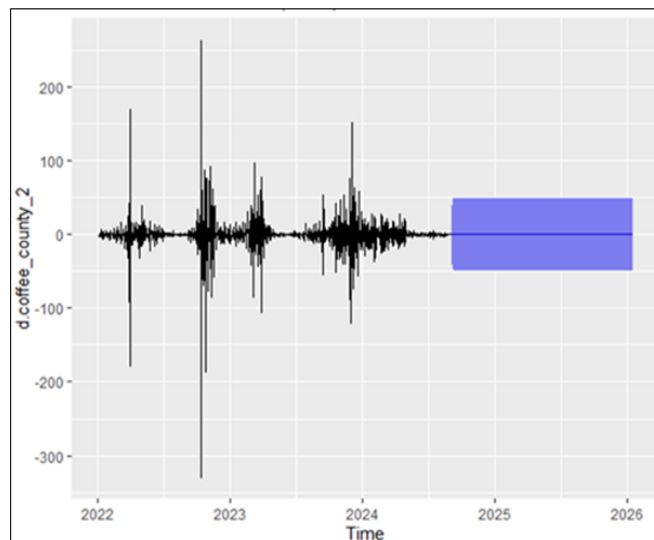
Fig 5 ARIMA(0, 0, 5) Forecast with Zero Mean

- *Model Validation Through Lag Value Selection*

lag=5, type="Ljung-Box"

Box-Ljung test

data: model_coffee_county_1$resid

X-squared = 2.4354, df = 5, p-value = 0.7862

lag=10, type="Ljung-Box"

Box-Ljung test

data: model_coffee_county_1$resid

X-squared = 38.218, df = 10, p-value = 3.477e-05

lag=15, type="Ljung-Box"

Box-Ljung test

data: model_coffee_county_1$resid

X-squared = 76.564, df = 15, p-value = 2.95e-10

Upon observing the p-values, it becomes evident that they are less than 0.01, leading us to reject the null hypothesis. Consequently, there exists sufficient statistical evidence to affirm that the residuals exhibit correlation, thereby allowing for the application of ARCH/GARCH.
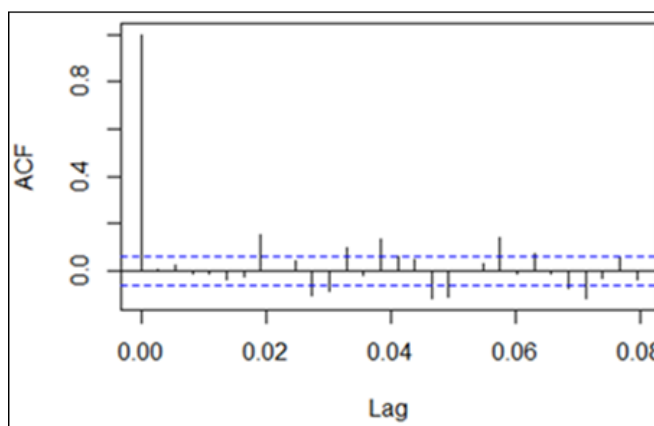

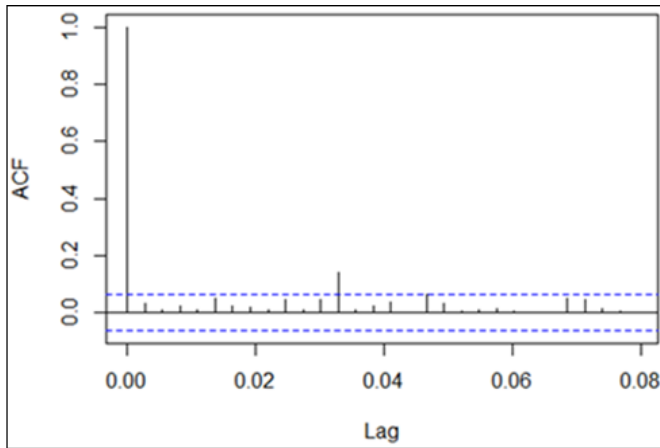Fig 6 Autocorrelation Function (ACF) Plot for the ARIMA Model's Residuals

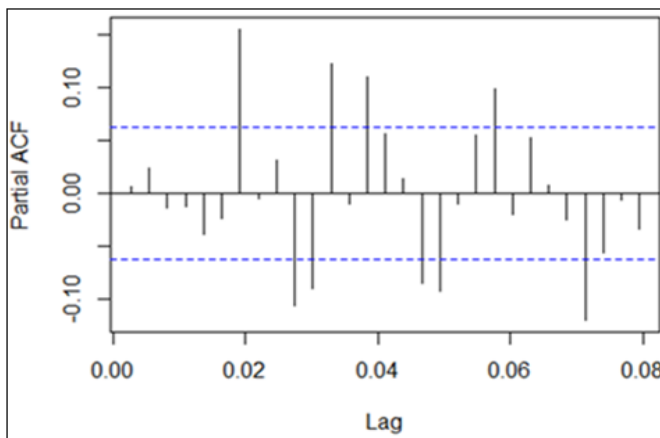Fig 7 Autocorrelation Function (ACF) Plot for the Squared Residuals of the ARIMA Model

Fig 8 Partial Autocorrelation Function (PACF) Plot for the ARIMA Model's Residuals

Analyzing the autocorrelation function (ACF) plot of squared residuals above, it becomes evident that there is a degree of correlation present between the residuals. This signifies the suitability of employing ARCH/GARCH models. Additionally, the ACF pattern of the squared series resembles an Autoregressive Moving Average (ARMA) pattern, as both the ACF and PACF gradually taper off. This observation suggests that a GARCH(1,1) model may be appropriate.
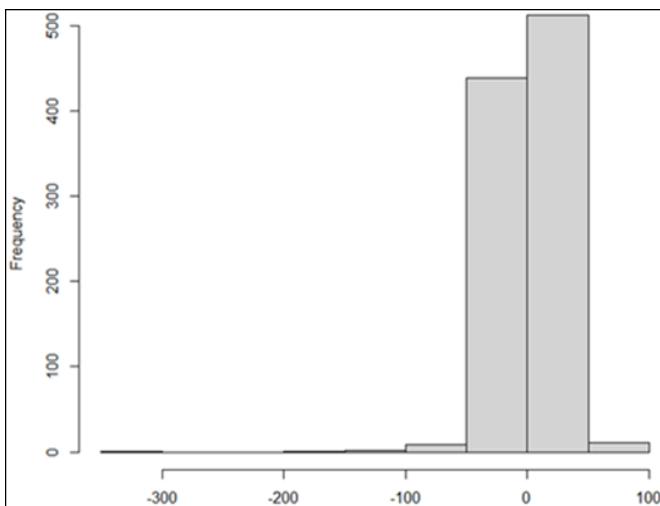
Fig 9 Histogram of ARIMA models Residuals
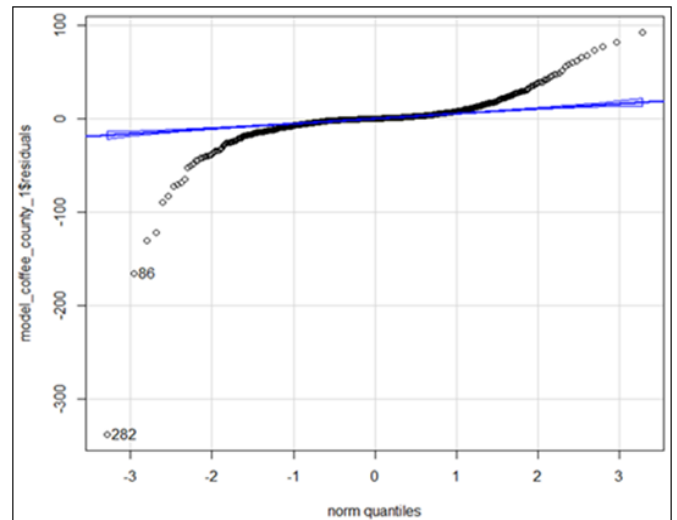
Fig 10 QQ Plot of ARIMA Model's Residuals

IV. AR MODEL

Call:
arima(x = d.coffee_county_2, order = c(1, 0, 0))
Coefficients:
     ar1 intercept
   -0.4085   0.0014
s.e. 0.0292   0.5206
sigma^2 estimated as 523.4: log likelihood = –4430.97,
aic = 8867.94
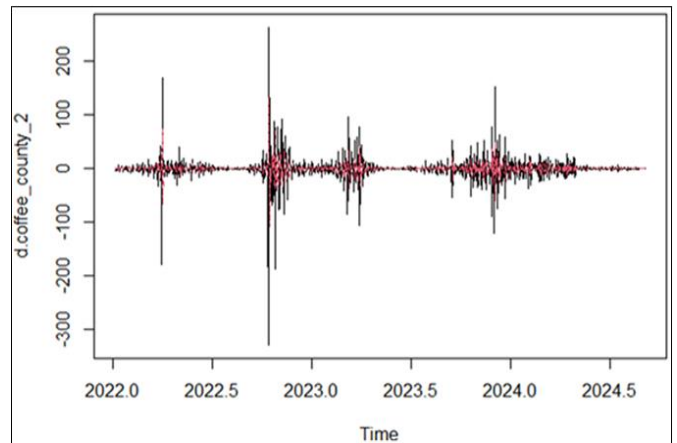
Fig 11 Time Series Visualization of COVID-19 Cases in Coffee County Using AR

➤ *1-Step Forecast for AR*

Spred
Time Series:
Start = c(2024, 249)
End = c(2024, 249)
Frequency = 365
[1] 0.002013471
Sse
Time Series:
Start = c(2024, 249)
End = c(2024, 249)

Frequency = 365
[1] 22.87902

➤ *AR Prediction*

$pred

Time Series:

Start = c(2024, 249)

End = c(2024, 258)

Frequency = 365

[1]   0.002013471   0.001191031   0.001526972   0.001389750
      0.001445801 0.001422906 0.001432258 0.001428438

[9] 0.001429998 0.001429361

$se

Time Series:

Start = c(2024, 249)

End = c(2024, 258)

Frequency = 365

[1] 22.87902 24.71408 25.00715 25.05571 25.06380 25.06515
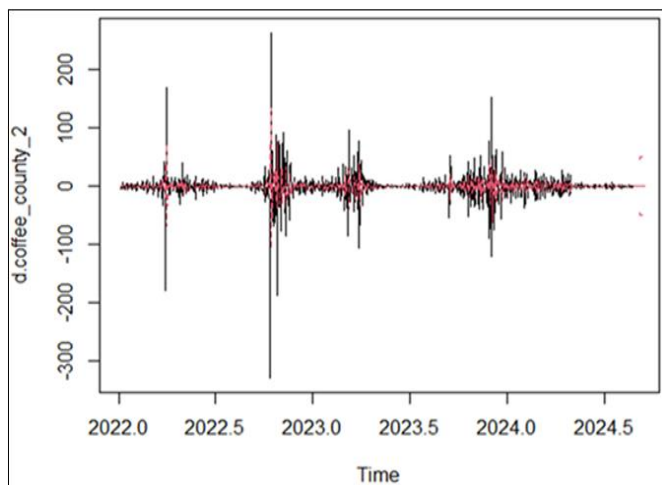   25.06538 25.06542 25.06542 25.06542

Fig 12 Plot of Covid 19 Forecast for Coffee County Using AR
Model

➤ *Comparison between AR with ARIMA*
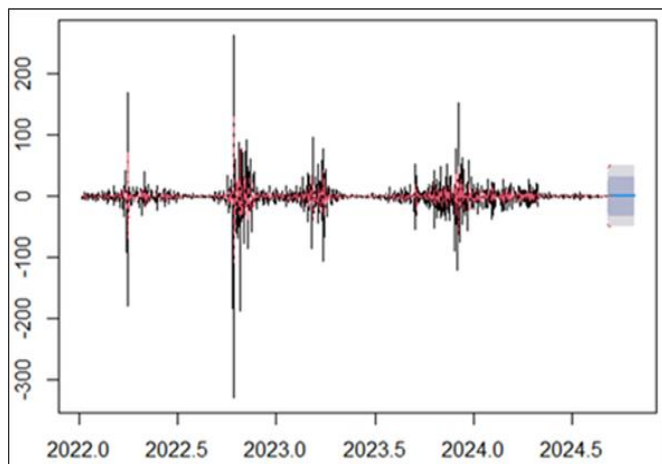
Fig 13 Comparative Forecasting Plot of Covid-19 for Coffee
County Using AR and ARIMA

Upon examination of the plot above, it becomes evident that the forecasts generated by AR and ARIMA are nearly indistinguishable, if not identical. The red line corresponds to AR, while the black line corresponds to MA.

V. MA MODEL

Call:
arima(x = d.coffee county 2, order = c(0, 0, 1))

Coefficients:
      ma1   intercept
    -0.7343   0.0041
s.e.   0.0207   0.1807

sigma^2 estimated as 447.8: log likelihood = -4355.29, aic = 8716.58
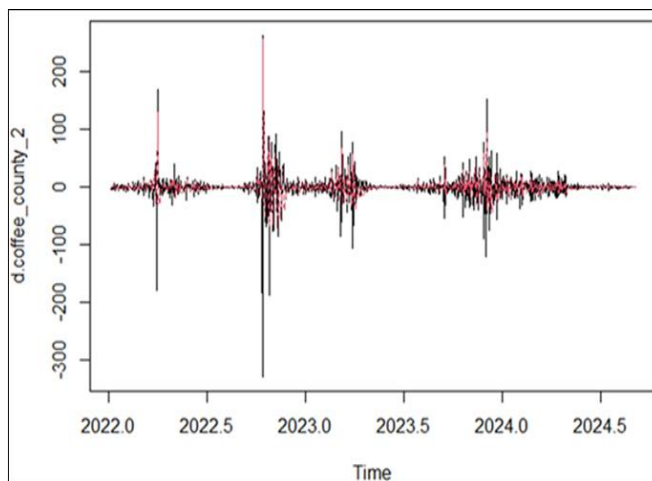
Fig 14 Time Series Visualization of Covid-19 for Coffee
County Using MA

➤ *MA Prediction*

$pred

Time Series:

Start = c(2024, 249)

End = c(2024, 258)

Frequency = 365

[1] 0.003260081 0.004078130 0.004078130 0.004078130
     0.004078130 0.004078130 0.004078130 0.004078130

[9] 0.004078130 0.004078130

$se

Time Series:

Start = c(2024, 249)

End = c(2024, 258)

Frequency = 365

[1]   21.16215   26.25438   26.25438   26.25438   26.25438
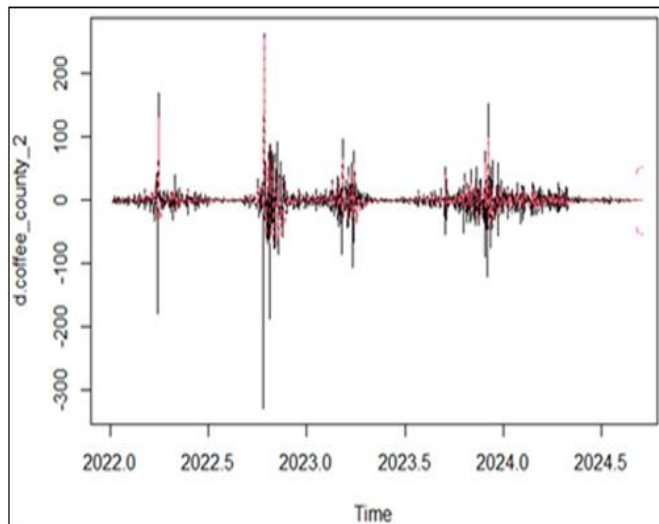   26.25438 26.25438 26.25438 26.25438 26.25438

Fig 15 Plot Comparing MA with ARIMA for Forecasting COVID-19 Cases in Coffee County
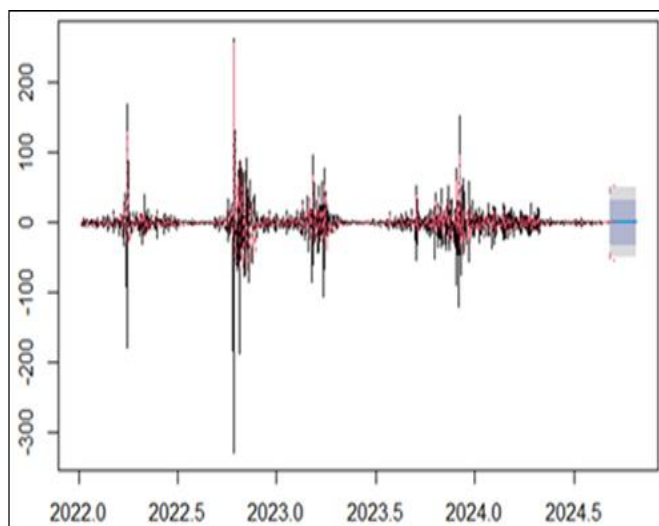

Fig 16 Forecasting Plot Comparing MA with ARIMA Covid 19 for Coffee County

➢ *Finding the correlation between fit for AR and fit MA*

Correlation: 0.7868456

AIC of AR = 8867.944

AIC of MA = 8716.58

BIC of AR = 8882.588

BIC of (MA) = 8731.225

Although there are various models available for time series data, AR and MA are among the most commonly used. Upon examining the provided output, the AR model exhibits an AIC value of 8867.944 and a BIC value of 8882.588, whereas the MA model demonstrates an AIC value of 8716.58 and a BIC value of 8731.225. The lower AIC and BIC values in the MA model suggest that, in comparison to the AR model, it provides a better fit for the time series analysis of Coffee County. However, it's worth noting that both models yield similar results, making them suitable for forecasting purposes.

VI. HOLTS EXPONENTIAL SMOOTHING

Call:

ets(y = d.coffee_county_2)

Smoothing parameters:

alpha = 1e-04

Initial states:

l = 4e-04

sigma: 25.0976

AIC	AICc	BIC
12984.45	12984.48	12999.09

accuracy(fit)

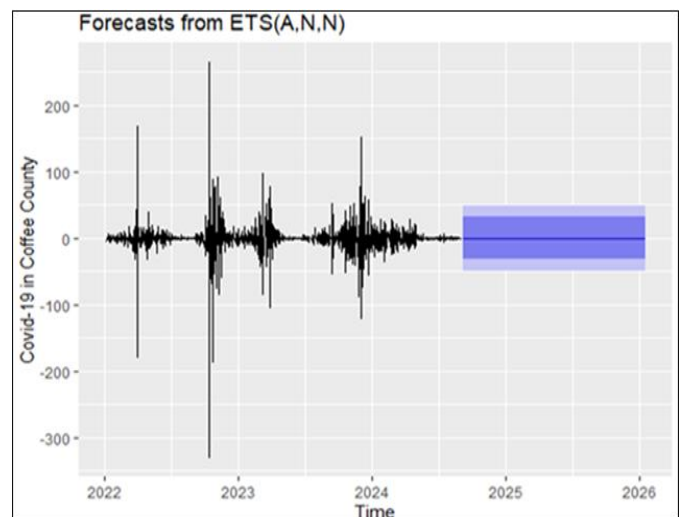|	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.0015	25.072	11.032	NaN	Inf	0.634	-0.4089

Fig 17 Visualization Showing Forecast of COVID-19 Trends in Coffee County Through Holt's Exponential Smoothing

Observing the plot above, it becomes apparent that the generated forecast appears reasonable, showing a consistent decline in COVID-19 cases over time.
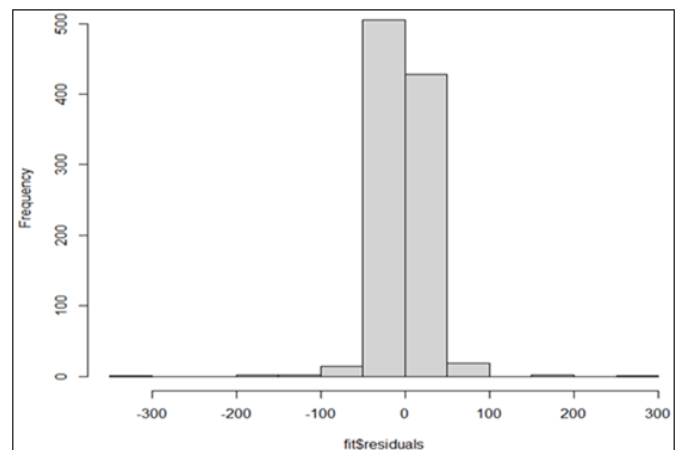

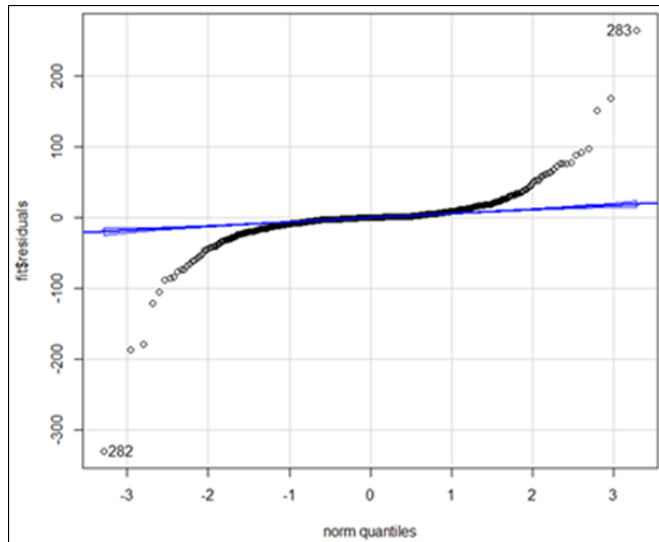Fig 18 Histogram Visualization of Residuals from Holts Exponential Smoothing Model

Fig 19 QQ Visualization of Residuals Derived from the Holts Exponential Smoothing Model

VII. GARCH MODEL

> *GARCH(1,1) Model*

Call:

garch(x = model_coffee_county_1$residual, grah = "numerical", trace = F)

Coefficient(s):

   a0   a1   b1

3.642e+02 2.711e-01 4.514e-13

t(confint(arch_model_coffee_county_1))

       a0   a1   b1

2.5 % 332.9201 0.2340534 -0.0444925

97.5 % 395.4461 0.3081728 0.0444925

summary(arch_model_coffee_county_1)

Call:

garch(x = model_coffee_county_1$residual, grah = "numerical", trace = F)

Model:

GARCH(1,1)

Residuals:

  Min   1Q   Median   3Q   Max

-14.63177 -0.16255 0.01545 0.19979 4.82561

Coefficient(s):

  Estimate Std. Error t value Pr(>|t|)

a0 3.642e+02 1.595e+01 22.83 <2e-16 ***

a1 2.711e-01 1.891e-02 14.34 <2e-16 ***

b1 4.514e-13 2.270e-02 0.00 1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:

  Jarque Bera Test

data: Residuals

X-squared = 338177, df = 2, p-value < 2.2e-16

  Box-Ljung test

data: Squared.Residuals

X-squared = 0.063855, df = 1, p-value = 0.8005

Fig 20 Forecasting COVID-19 Trends for Coffee County from the GARCH(1,1) Model
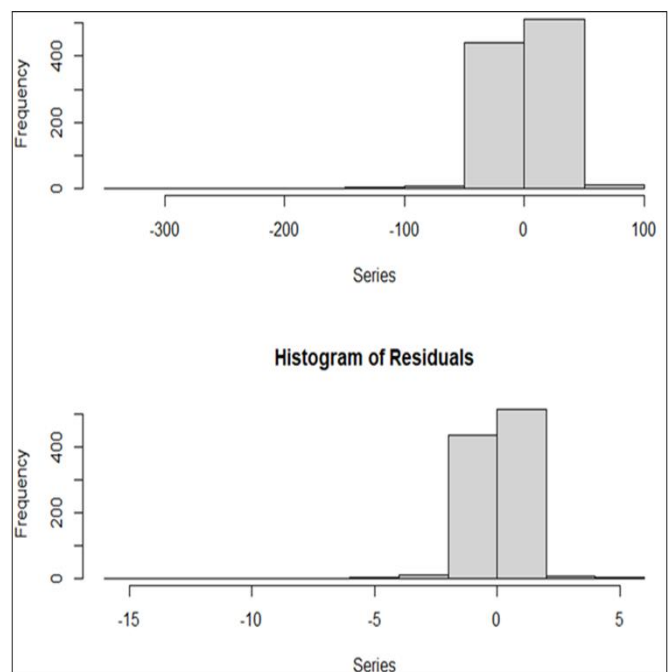

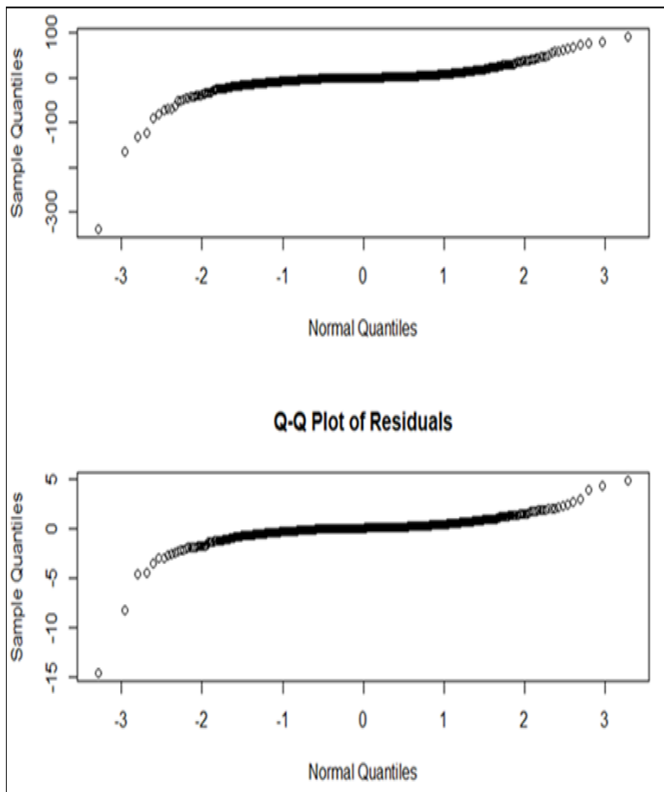Fig 21 Histogram Visualization of Residuals from the GARCH(1,1) Model

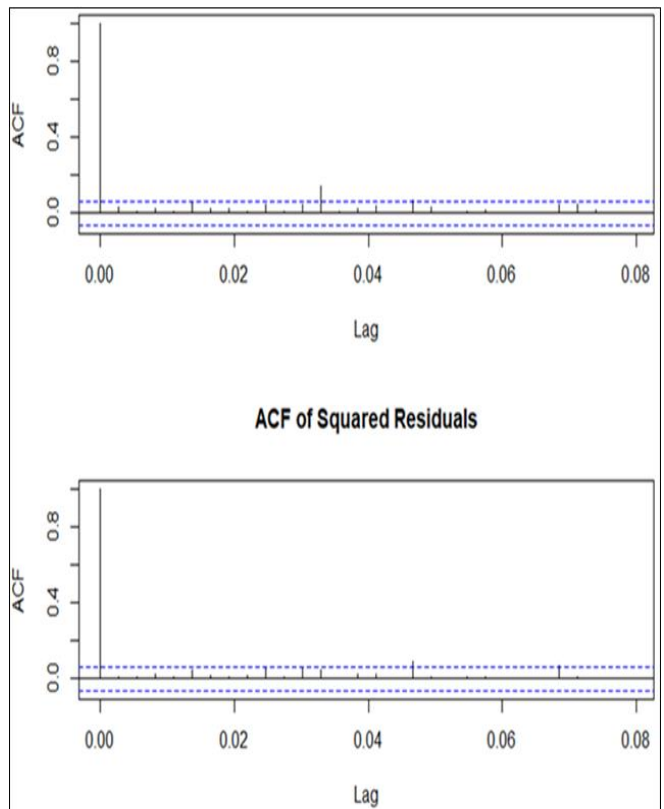Fig 22 QQ Visualization of Residuals Derived from
GARCH(1,1) Model

Fig 23 AFC Visualization of Residuals from GARCH(1,1)

➢ *GARCH(0,1) Model*

Call:
garch(x = model_coffee_county_1$residual, order = c(0, 1),
trace = F)
Coefficient(s):
  a0   a1
82.629 2.299
t(confint(arch01_model_coffee_county_1))
      a0   a1
2.5 % 80.33296 2.204260

97.5 % 84.92404 2.393222
summary(arch01_model_coffee_county_1)
Call:
garch(x = model_coffee_county_1$residual, order = c(0, 1),
trace = F)
Model:
GARCH(0,1)
Residuals:
   Min   1Q   Median   3Q   Max
-8.88651 -0.21730 0.02671 0.27932 9.97103
Coefficient(s):
   Estimate Std. Error t value Pr(>|t|)
a0 82.62850   1.17122   70.55 <2e-16 ***
a1 2.29874   0.04821   47.69 <2e-16 ***
---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Diagnostic Tests:
    Jarque Bera Test
data: Residuals
X-squared = 49800, df = 2, p-value < 2.2e-16
    Box-Ljung test
data: Squared.Residuals
X-squared = 0.14346, df = 1, p-value = 0.7049

#AIC FOR GARCH(0,1)
(loglik01_model_coffee_county_1=logLik(arch01_model_coff
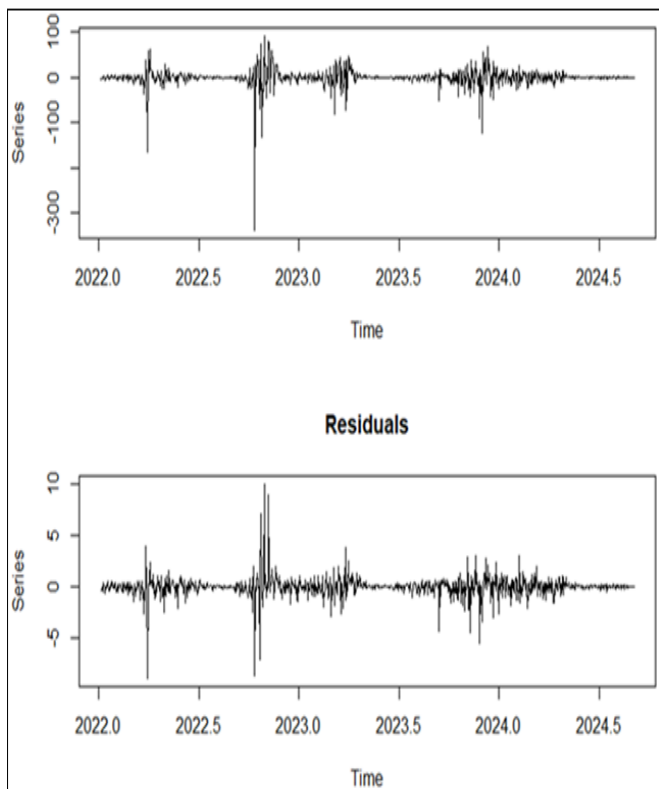ee_county_1))
'log Lik.' -3964.428 (df=2)

Fig 24 Forecasting COVID-19 Trends for Coffee County from the GARCH(0,1)
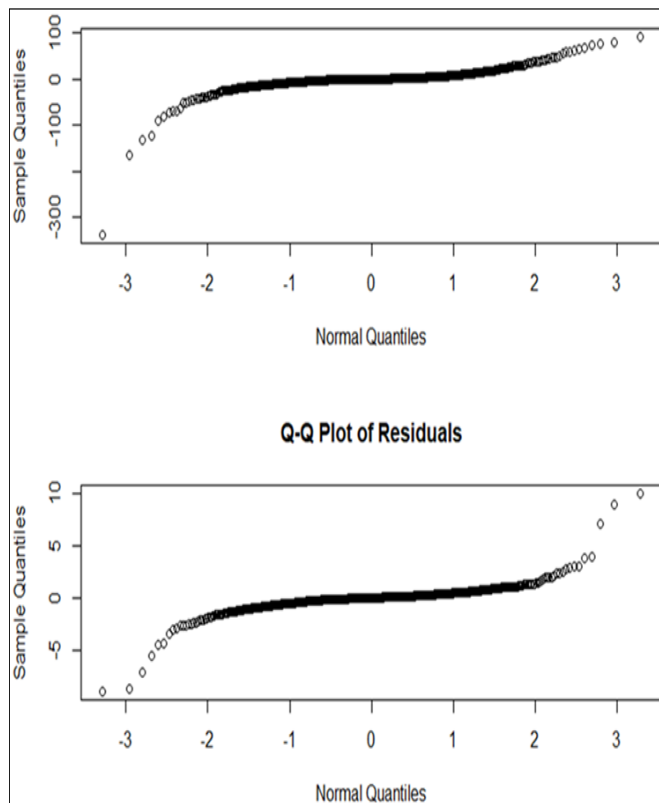

Fig 26 Visualization of Residuals Derived from GARCH (0,1)
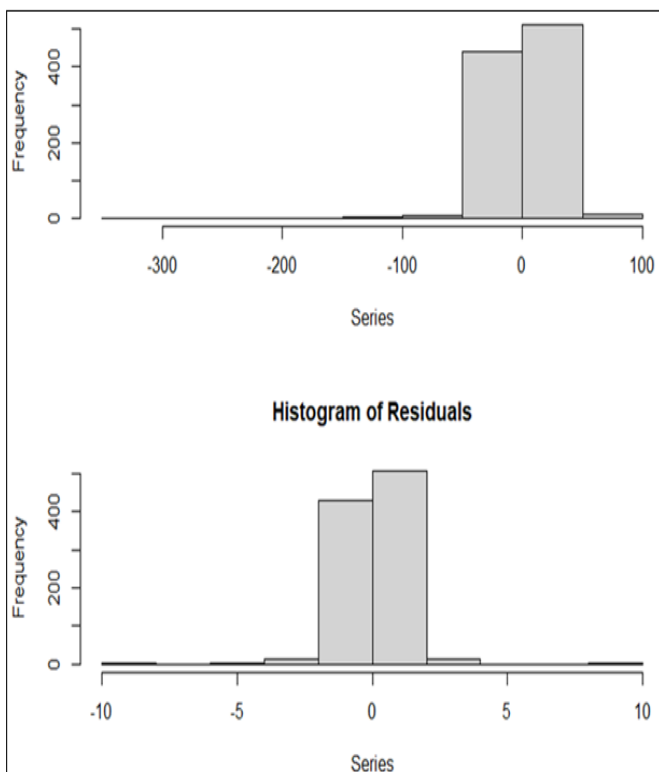

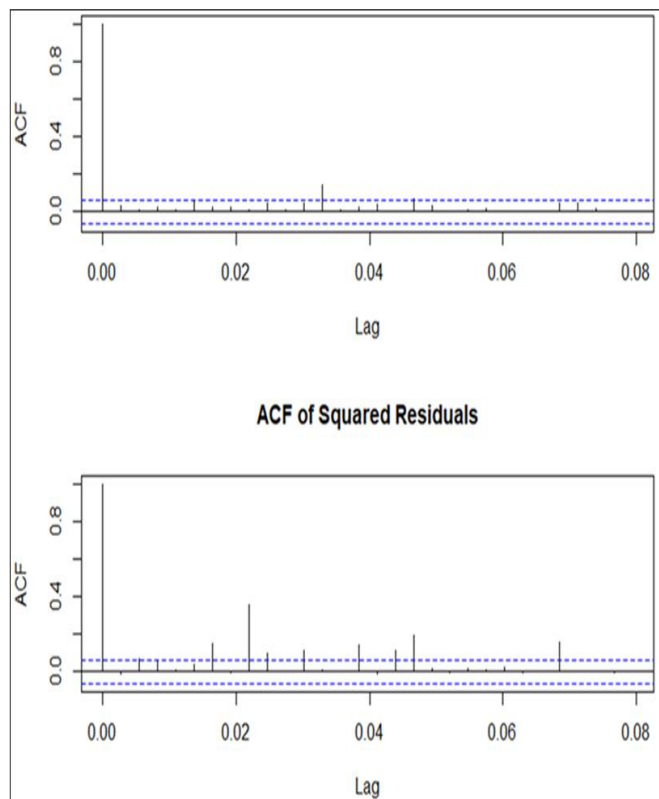Fig 25 Histogram Visualization of Residuals from GARCH(0,1)

Fig 27 AFC Visualizations of Residuals from GARCH(0,1)

Upon examining the Box-Ljung test results for both GARCH(1,1) and GARCH(0,1), it becomes evident that all p-values are comfortably above the 0.05 threshold, indicating the suitability of these GARCH models.

Furthermore, when assessing the histograms and QQ plots of residuals for both GARCH(1,1) and GARCH(0,1), it is apparent that the residuals tend to follow a roughly normal distribution.

While both GARCH(1,1) and GARCH(0,1) exhibit favorable characteristics, my preference leans towards GARCH(1,1) due to its lower p-value compared to GARCH(0,1), signifying a better fit for the time series model.

Call:

garch(x = arch01_model_coffee_county_2)

Model:

GARCH(1,1)

Residuals:

Min	1Q	Median	3Q	Max
-14.63177	-0.16255	0.01545	0.19979	4.82561

Coefficient(s):

	Estimate	Std. Error	t value	Pr(>|t|)
a0	3.642e+02	1.595e+01	22.83	<2e-16 ***
a1	2.711e-01	1.891e-02	14.34	<2e-16 ***
b1	4.514e-13	2.270e-02	0.00	1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:

Jarque Bera Test

data: Residuals

X-squared = 338177, df = 2, p-value < 2.2e-16

Box-Ljung test

data: Squared.Residuals

X-squared = 0.063855, df = 1, p-value = 0.8005

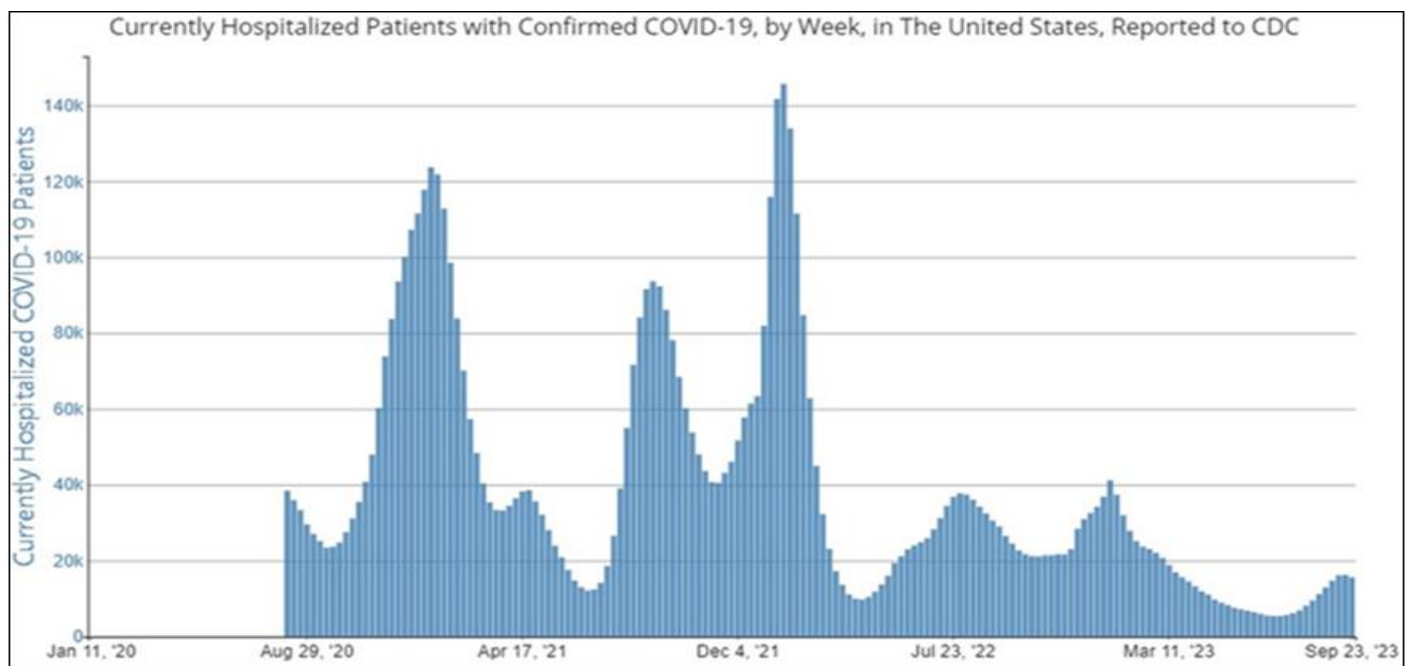➢ *Comparing the Different Models with the CDC Report for the Same Time Period*

Fig 28 Currently Hospitalized Patients with Confirmed COVID-19, by Week, in The United States, Reported to CDC [2]

Looking at Figure 27, we can conclude that all the models were predicting correctly and because all the models predicted spikes in 2023 which can be confirmed by Covid 19 spikes in January 2023 and August 2023. We can conveniently say that the models to a certain level are accurately predicting. Figures 28, 29, 30, 31, 32, and 33 show a comparison between the Model Predictions and the CDC Report for the same period.
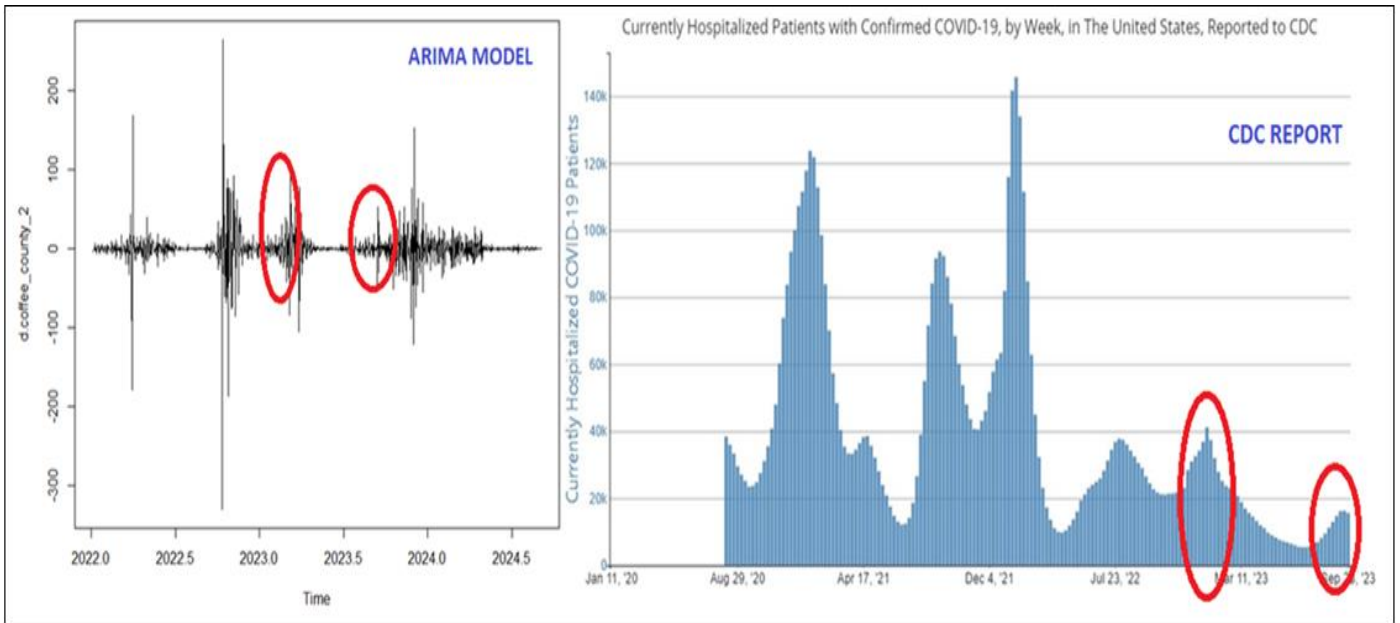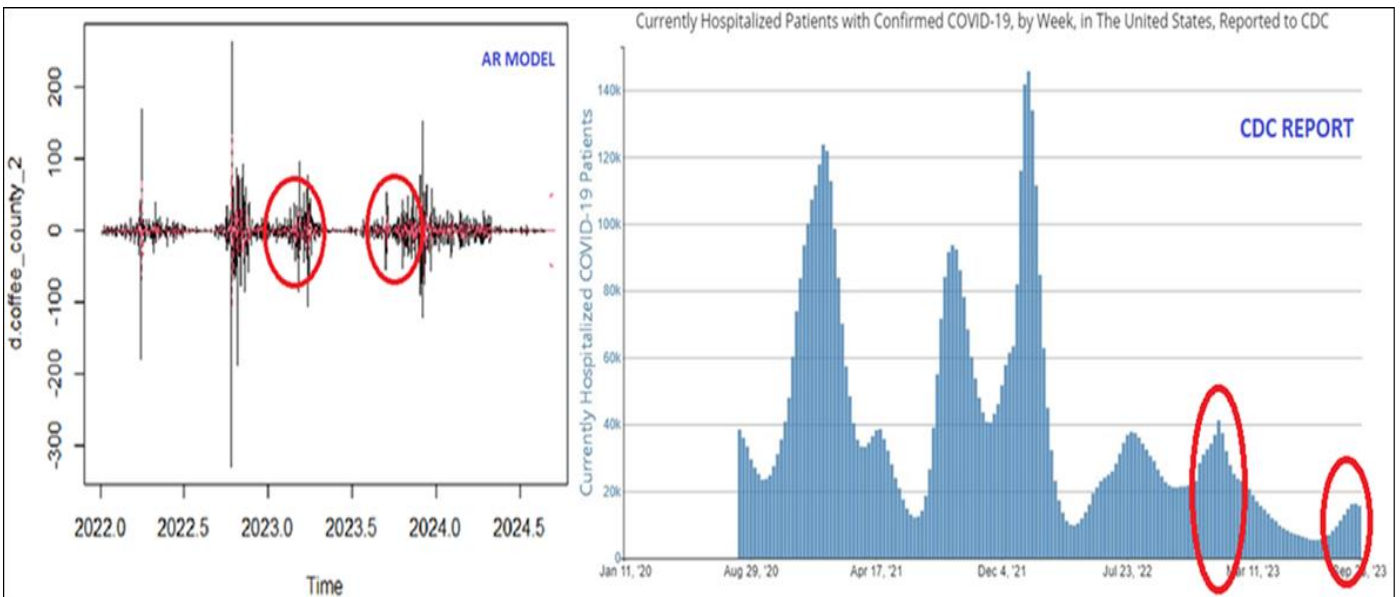
Fig 29 ARIMA Model Vs CDC Report
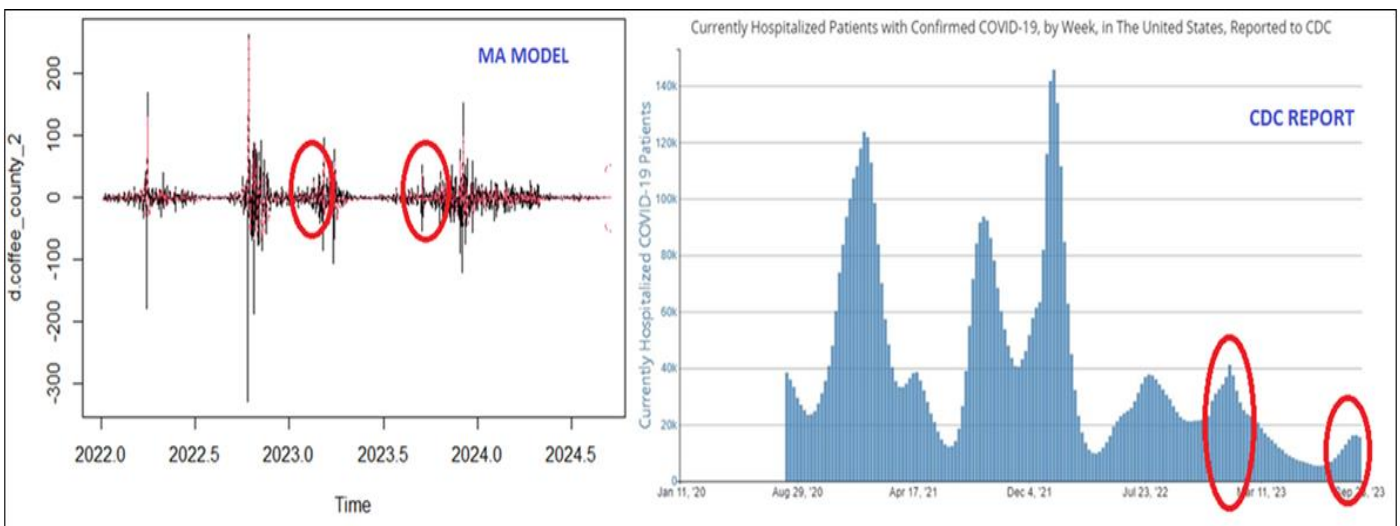

Fig 30 AR Model Vs CDC Report

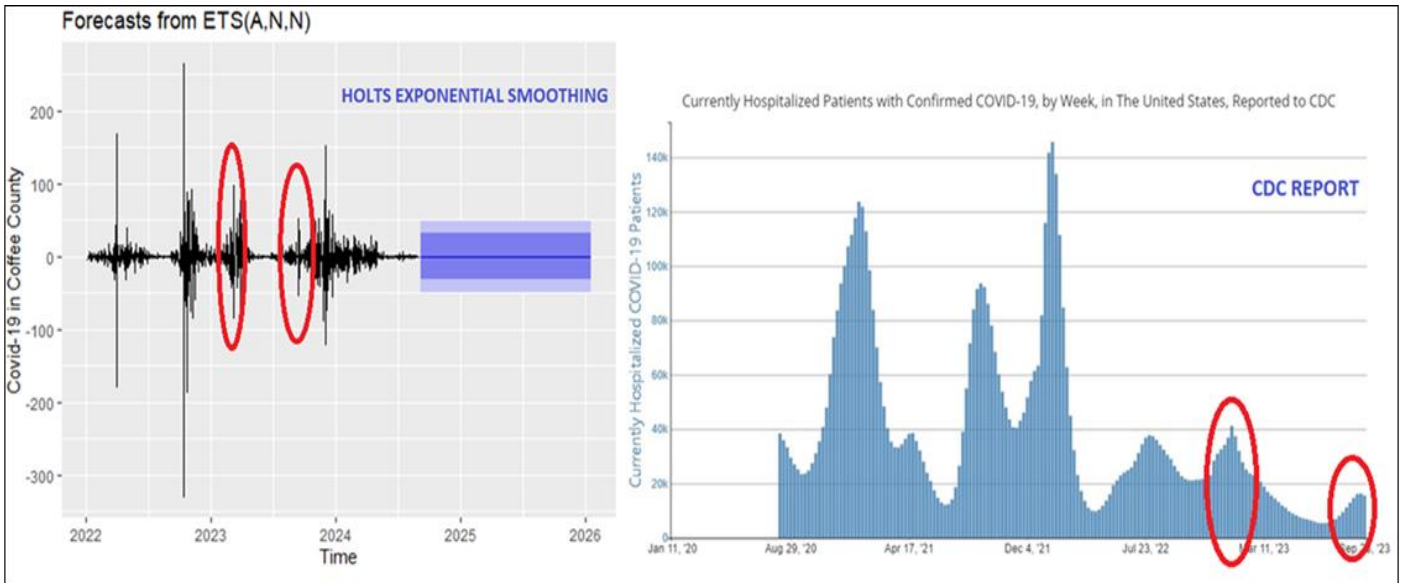Fig 31 MA Model Vs CDC Report

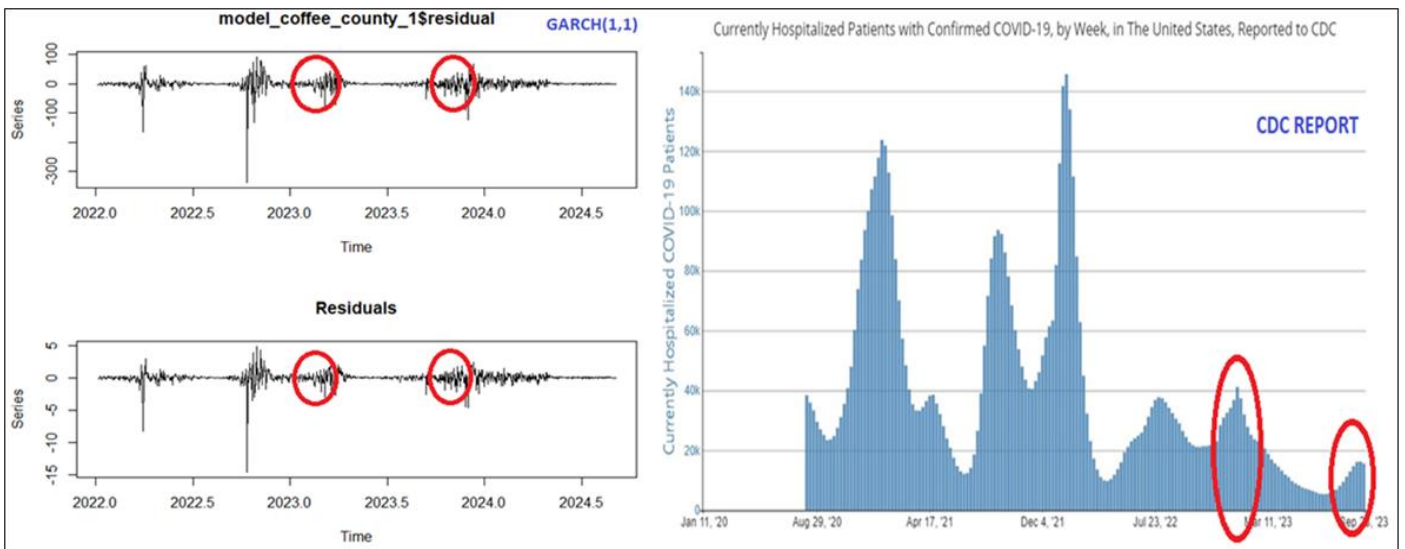Fig 32 Holts Exponential Smoothing Vs CDC Report

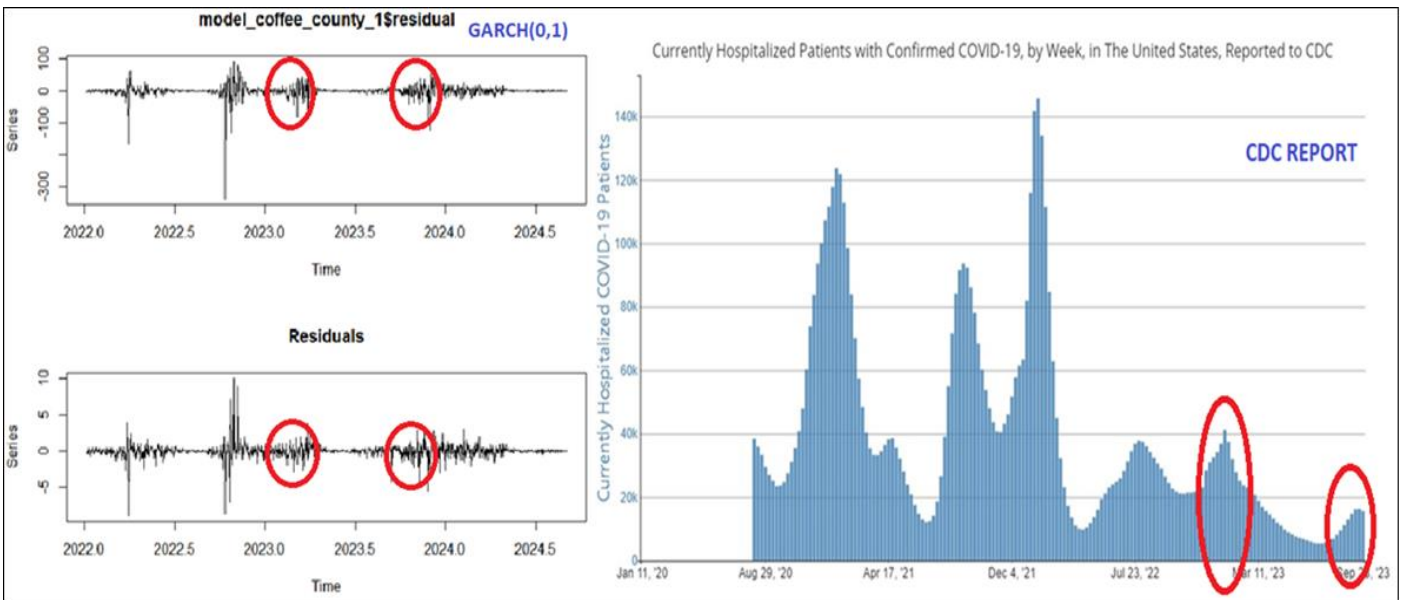Fig 33 GARCH(1,1) Model Vs CDC Report

Fig 34 GARCH(0,1) Model Vs CDC Report

VIII. CONCLUSION

In conclusion, we notice the five models considered are making reasonably accurate predictions up to a certain point. though among these models, GARCH emerges as the most suitable option. This assertion is grounded in the GARCH model's favorable characteristics, as evidenced by its lowest AIC values of -3964.428 for GARCH(0,1) and -4195.056 for GARCH(1,1)), in stark contrast to the AIC values of 8623.38 for ARIMA and 12984.45 for Holt's Exponential Smoothing. Additionally, a notable advantage of the GARCH model lies in its residuals, which exhibit a distribution closely resembling a normal distribution, as observed in the histogram and QQ plots. This stands in contrast to the residuals of ARIMA and Holt's Exponential Smoothing, which display a somewhat skewed distribution.

In light of these findings, it is evident that the GARCH model outperforms the other models, including ARIMA, AR, MA, and Holt's Exponential Smoothing, making it the preferred choice for our time series data analysis. In summary, an analysis of Figure 27 indicates that all the models made correct predictions.

This is particularly evident as they consistently forecasted spikes in 2023, a pattern that aligns with the observed spikes in Covid-19 cases in January 2023 and August 2023. These findings suggest a noteworthy level of accuracy in the predictive capabilities of these models. Furthermore, Figures 28 to 33 provide a comprehensive comparison between the Model Predictions and the CDC Report for the corresponding period, further validating the reliability of the models' predictions.

REFERENCES

[1]. Chakraborty, T., Ghosh, I., & Chattopadhyay, J. (2020). Estimation of future COVID-19 spread in metropolitan cities of India with hybrid model. Chaos, Solitons & Fractals, 138, 109945..

[2]. Centers for Disease Control and Prevention. COVID Data Tracker, Atlanta, GA: U.S. Department of Health and Human Services, CDC; 2023, September 30. Https://covid.cdc.gov/covid-data-tracker

[3]. Gupta, M., George, A. T., & Kumar, S. (2021). Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India using hybrid machine learning approaches. Chaos, Solitons & Fractals, 142, 110415.

[4]. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 115(772), 700-721.

[5]. Nau, Bob. "Stationary and Differencing." Duke University. Durham, NC. n.d. Jong, Robert. "Stationary Time Series." Ohio State University. Columbus, OH. n.d.

[6]. Pfefferbaum, B., & North, C. S. (2020). Mental health and the Covid-19 pandemic. New England Journal of Medicine, 383(6), 510-512.

[7]. Russo, V., Bottino, F., Mupo, D., Capunzo, M., Di Giuseppe, G., Piscopo, G., ... & Moauro, F. (2020). Effects of COVID-19 lockdown on air quality in the urban area of Naples (Italy). Science of the Total Environment, 750, 141576.

[8]. Sharma, A., Ahmad Farouk, M., & Dhiman, G. (2021). Forecasting of COVID-19 cases using hybrid model. Alexandria Engineering Journal, 60(3), 2855-2867.

[9]. Shumway, R. H., & Stoffer, D. S. (2017). Time Series Analysis and Its Applications With R Examples Fourth Edition.

[10]. Yancy, C. W. (2020). COVID-19 and African Americans. JAMA, 323(19), 1891-1892.

[11]. Dataset:

[12]. https://www.tn.gov/content/dam/tn/health/documents/cedep/novelcoronavirus/datasets/Public-Dataset-County-New.XLSX

[13]. Dal Molin Ribeiro M. H., Coelho L. D. S. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Appl. Soft Comput. 86, 105837. 10.1016/j.asoc.2019.105837

[14]. Webpages

[15]. https://www.news-journal.com/counties-with-highest-covid-19-infection-rates-in-tennessee/collection_2761d96c-7171-57c3-bcfe-1a243ae20dc0.html#1

[16]. https://rpubs.com/JSHAH/481706?fbclid=IwAR2hlb12jBxLqPNPh5vAdBUcGNIXszkOUrKelG_-GznwLuzC4M18eTncQgk