**Time Series Analysis**

**Problem Statement:**

**Data of Covid positive cases and Death on account of same for India is provided in the sheet. Complete a time series analysis for the data set. Include smoothing, ACF PACF Study and transformation where required.**

**Answer:**

* The data includes Covid cases and deaths from March 2020 to May 2020 for India. In this period India had increased number of cases and death simultaneously. So we first understand the data and do the required changes. For the death, the 9th observation(i.e. on 1st April) shows only 1 death which is actually wrong so we replace it with the average of 8th and 10th observation. As this data has and increasing trend so we use cumulative to calculate the confirmed positive cases and confirmed deaths. We observe that the 28th observation may be a typo error so we replace 543 with 573.
* Now the data is been cleaned. We now, analyse both time dependent series by including trend analysis, smoothing, ACF , PACF and ,any more.

• **Overview of tools used**

1. **Trend**  **Analysis**

An analysis of the trend of the observations is needed to acquire an understanding of the progress of events leading to prevailing conditions. Time series data can have these four components:

* Secular trend :This is the long term growth or decline of the series.
* Seasonal variation: The seasonal variation of a time series is a pattern of change that recurs regularly over time.
* Cyclical variation: Cyclical variations also have recurring patterns but with a longer and more erratic time scale compared to Seasonal variations.
* Irregular variation: It usually occurs randomly and may be linked to events that also occur randomly.

1. **Double Exponential Smoothing**

The basic idea behind double exponential smoothing is to introduce a term to take into account the possibility of a series exhibiting some form of trend.

= -

1. **ACF**

The autocorrelation function (ACF) assesses the correlation between observations in a time series for a set of lags. The ACF for time series y is given by: Corr (yt,yt−k), k=1,2,….

1. **PACF**

The Partial Autocorrelation Function (PACF) is similar to the ACF, however it measures correlation between observations that are k time periods apart, after controlling for correlations at intermediate lags.

where,

&

1. **ARIMA (p,d,q) model**

The Autoregressive Integrated Moving Average (ARIMA) model uses time-series data and statistical analysis to interpret the data and make future predictions. The ARIMA model aims to explain data by using time series data on its past values and uses linear regression to make predictions. It is a generalization of an autoregressive moving average model.

1. **Gaussian White Noise**

A white noise process is one with no discernible structure. A definition of a white noise process is

= 0 , o.w.

If is normal, then it is called as Gaussian White Noise.

1. **Transformation**

Transformation is required to forecast future trends to aid in decision making. The power transformation is defined as a continuously varying function, with respect to the power parameter λ, in a piece-wise function form that makes it continuous at the point of singularity (λ = 0). For data vectors (y1,…, ) in which each yi >0, the power transform is,

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

* **Software put into use and the outputs**

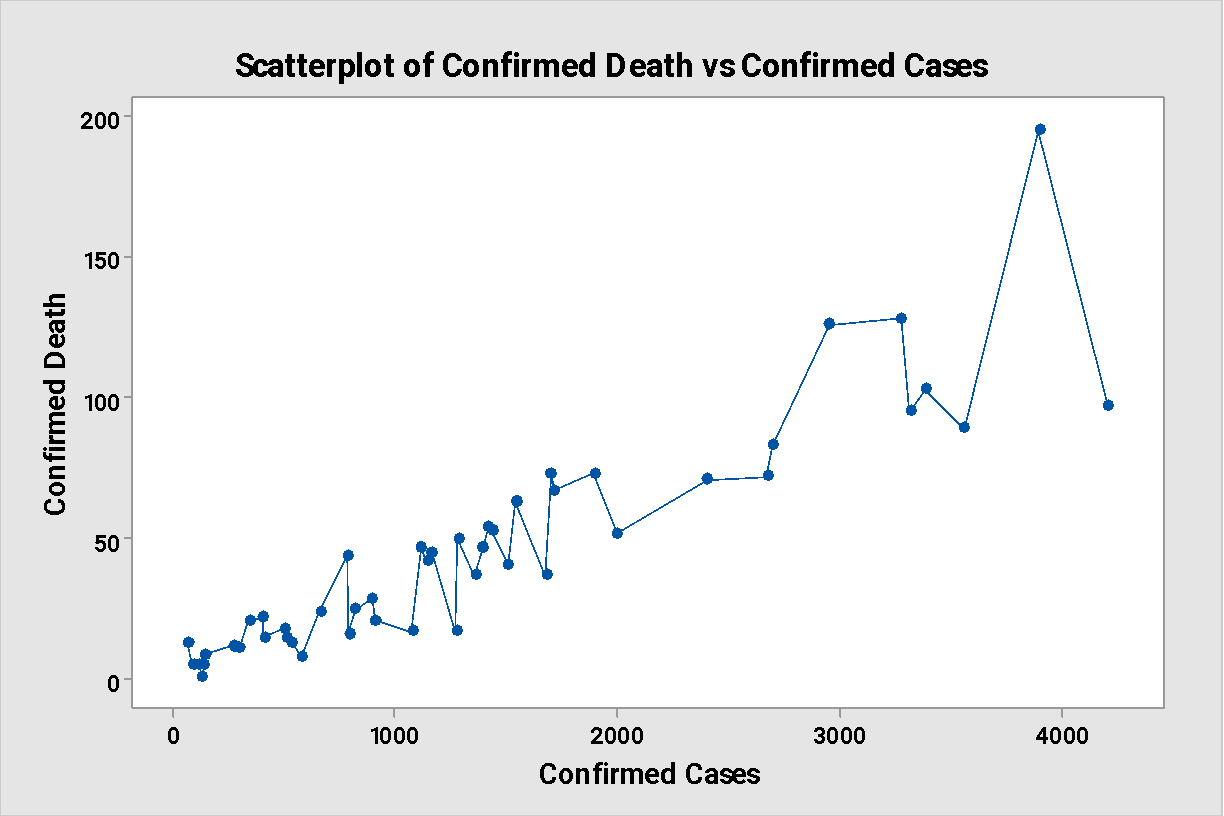
Minitab and Excel software were used to analyse the time series and interpret the results.

* **Interpretation and Results – tables/charts/ summary**

Descriptive Statistics: Covid Cases, Deaths

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | N | N\* | Mean | SE Mean | StDev | Minimum | Q1 | Median | Q3 | Maximum |
| Covid Cases | 49 | 0 | 20109 | 2728 | 19094 | 563 | 3901 | 14182 | 32191 | 67152 |
| Deaths | 49 | 0 | 657.6 | 90.9 | 636.2 | 0.0 | 117.0 | 462.0 | 1040.5 | 2206.0 |

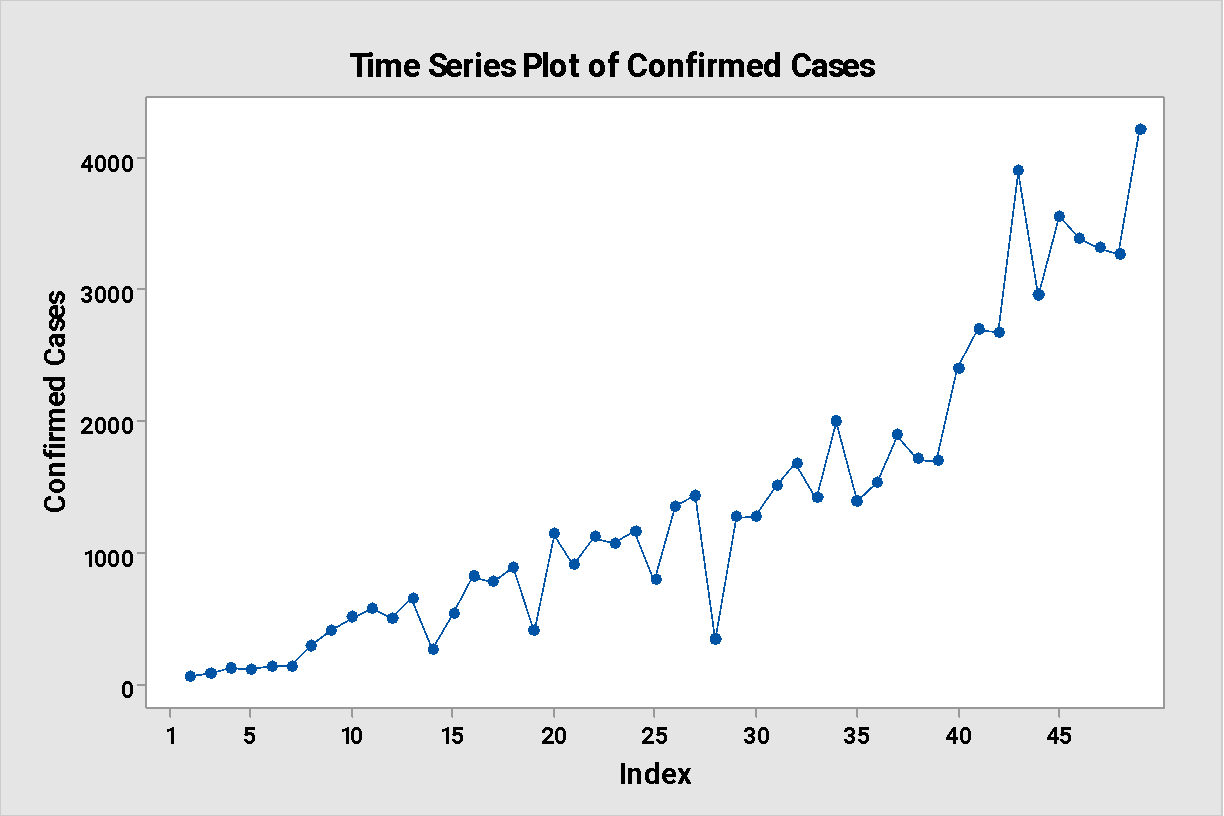
Scatterplot of Confirmed Death vs Confirmed Cases



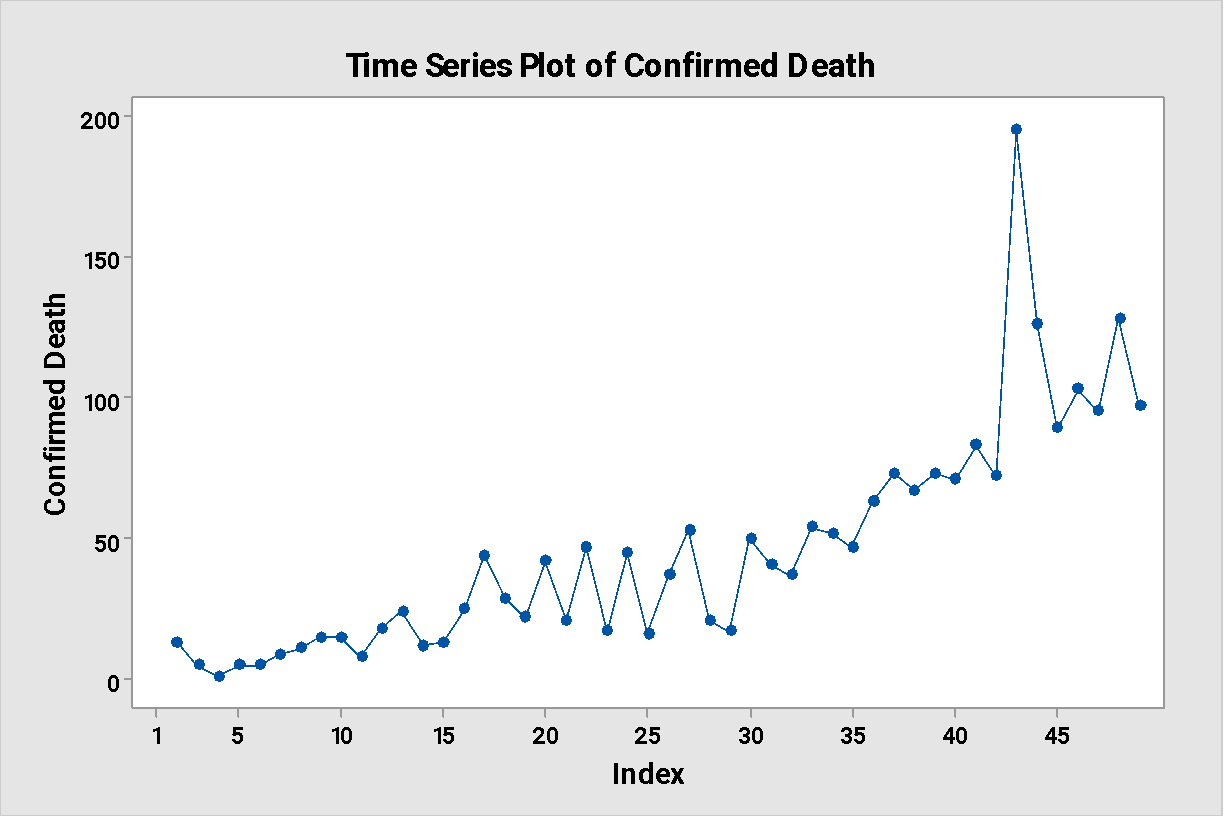
Interpretation:

There is an increasing number of death as cases increase.

Time Series Plot of Confirmed Cases



Time Series Plot of Confirmed Death



Interpretation:

* We see an increasing trend in both plot.
* Seasonality is absent.

Trend Analysis for Confirmed Cases

Method

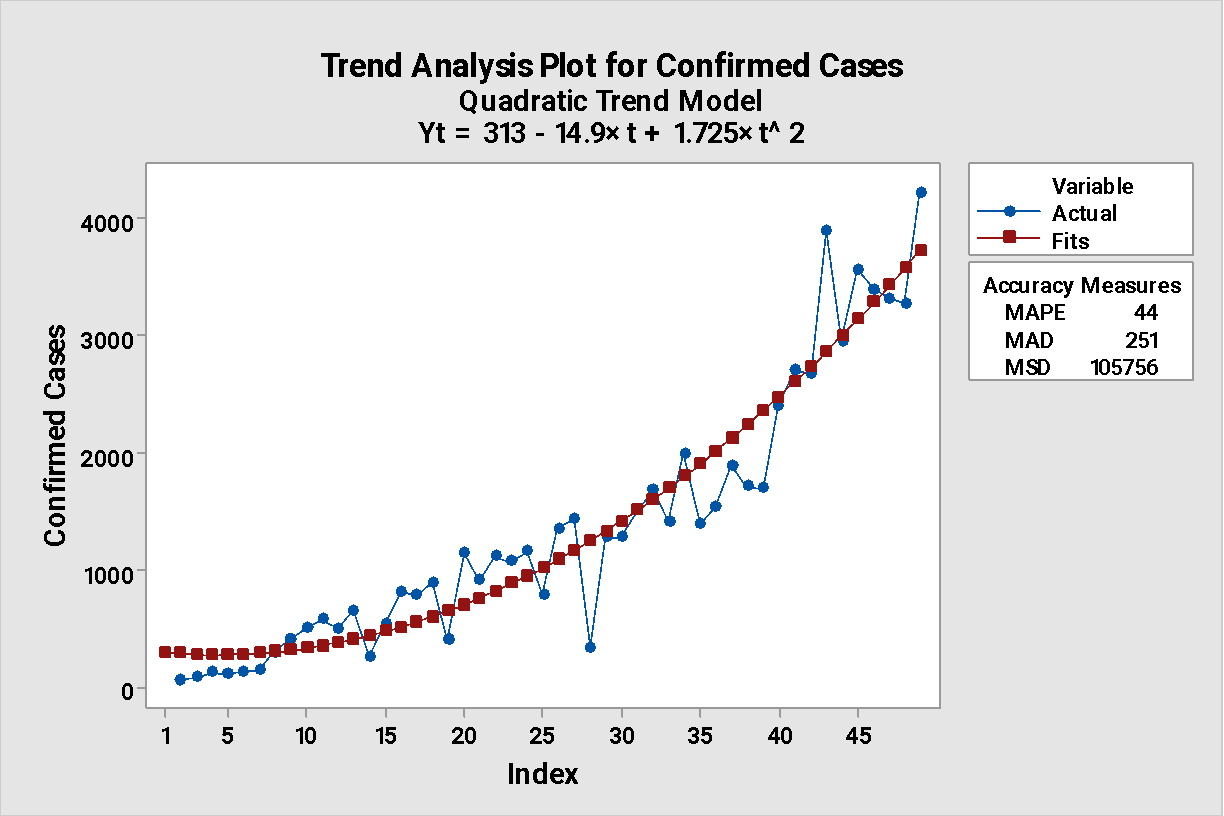
|  |  |
| --- | --- |
| Model type | Quadratic Trend Model |
| Data | Confirmed Cases |
| Length | 49 |
| NMissing | 1 |

Fitted Trend Equation

|  |
| --- |
| Yt = 313 - 14.9×t + 1.725×t^2 |

Accuracy Measures

|  |  |
| --- | --- |
| MAPE | 44 |
| MAD | 251 |
| MSD | 105756 |
|  |  |



Trend Analysis for Confirmed Death

Method

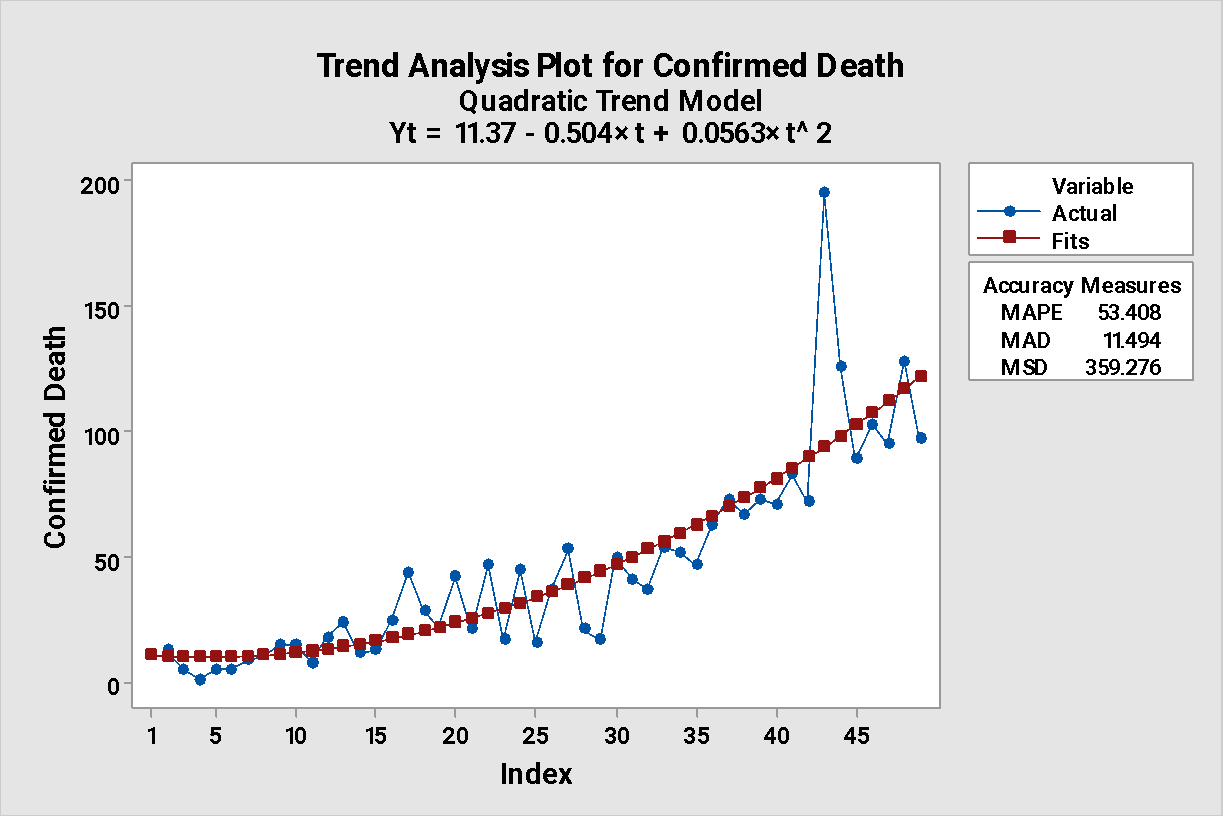
|  |  |
| --- | --- |
| Model type | Quadratic Trend Model |
| Data | Confirmed Death |
| Length | 49 |
| NMissing | 1 |

Fitted Trend Equation

|  |
| --- |
| Yt = 11.37 - 0.504×t + 0.0563×t^2 |

Accuracy Measures

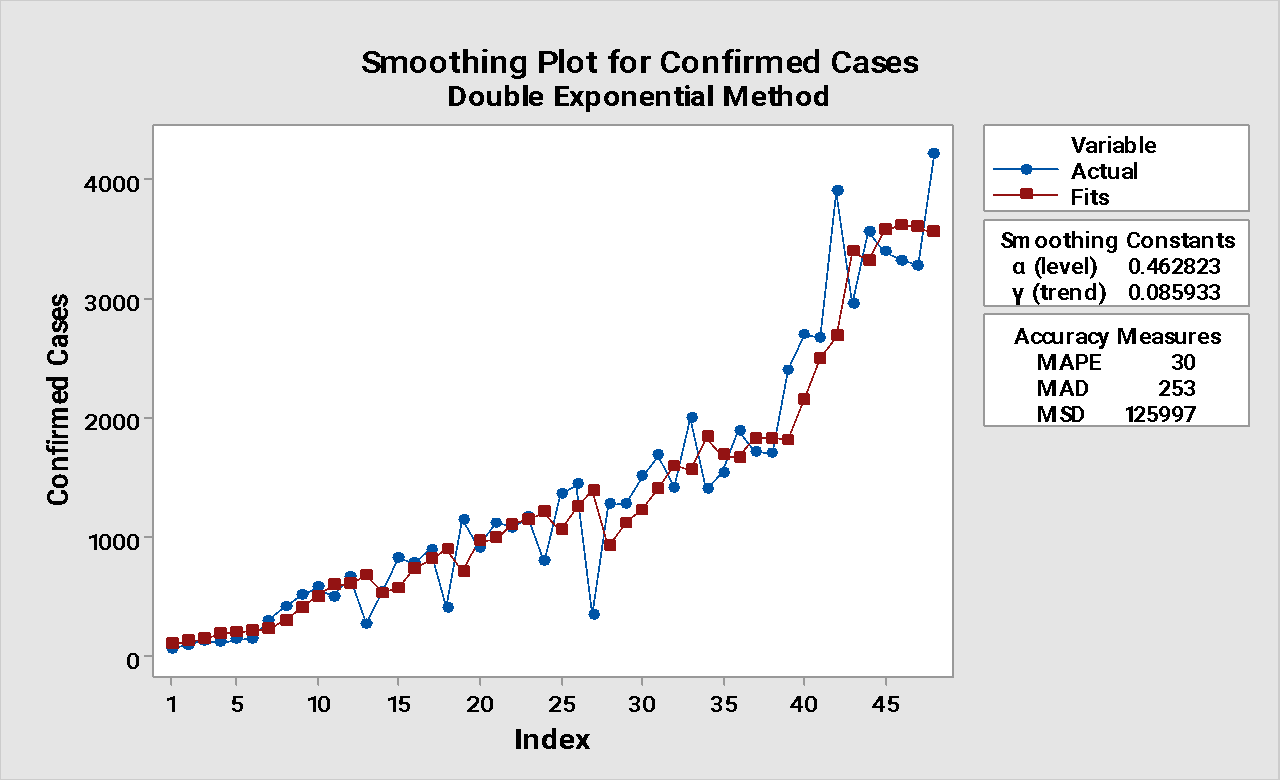
|  |  |
| --- | --- |
| MAPE | 53.408 |
| MAD | 11.494 |
| MSD | 359.276 |



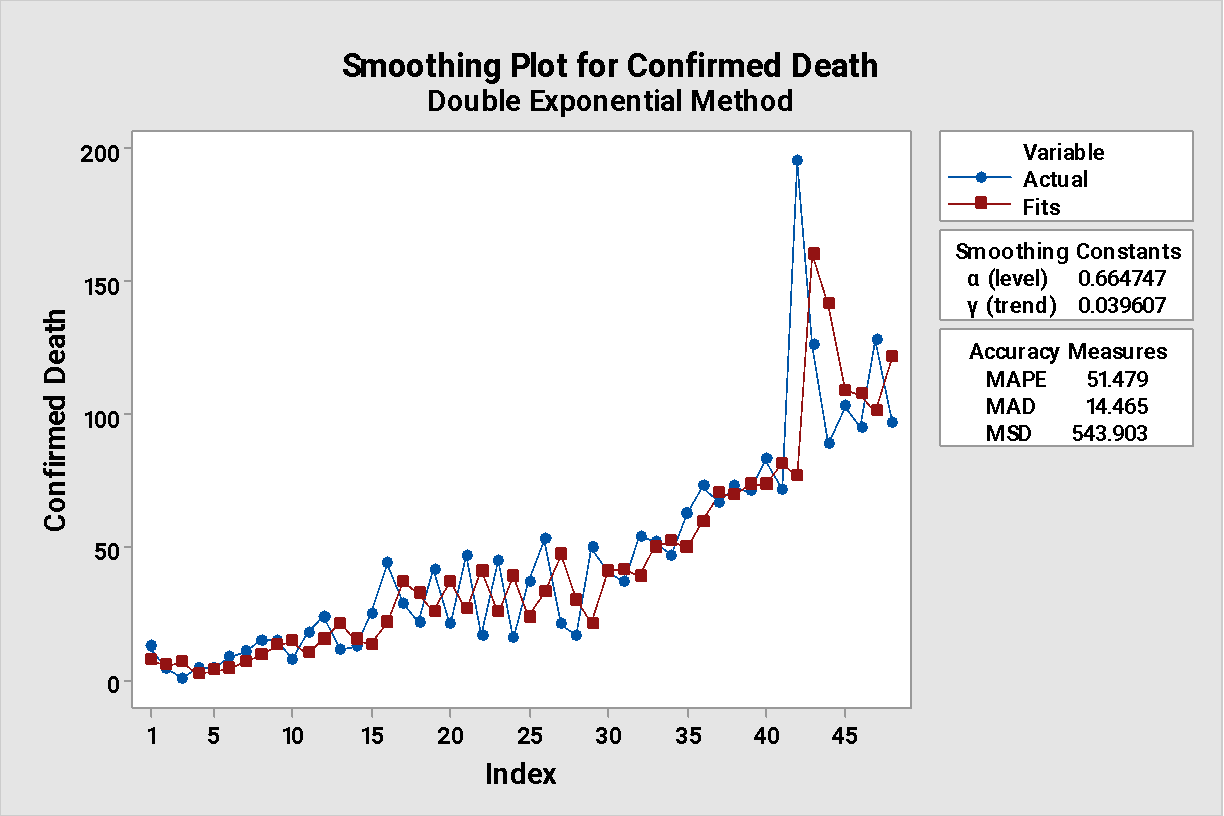
Interpretation:

* We see the quadratic factor has less effect still fits better than linear smoothing. Most of the effect is due to constant. And their MSD values are also smaller than the respective linear smoothing.
* Double Exponential Smoothing for Confirmed Cases & Confirmed Deaths

Double Exponential Smoothing for Confirmed Cases

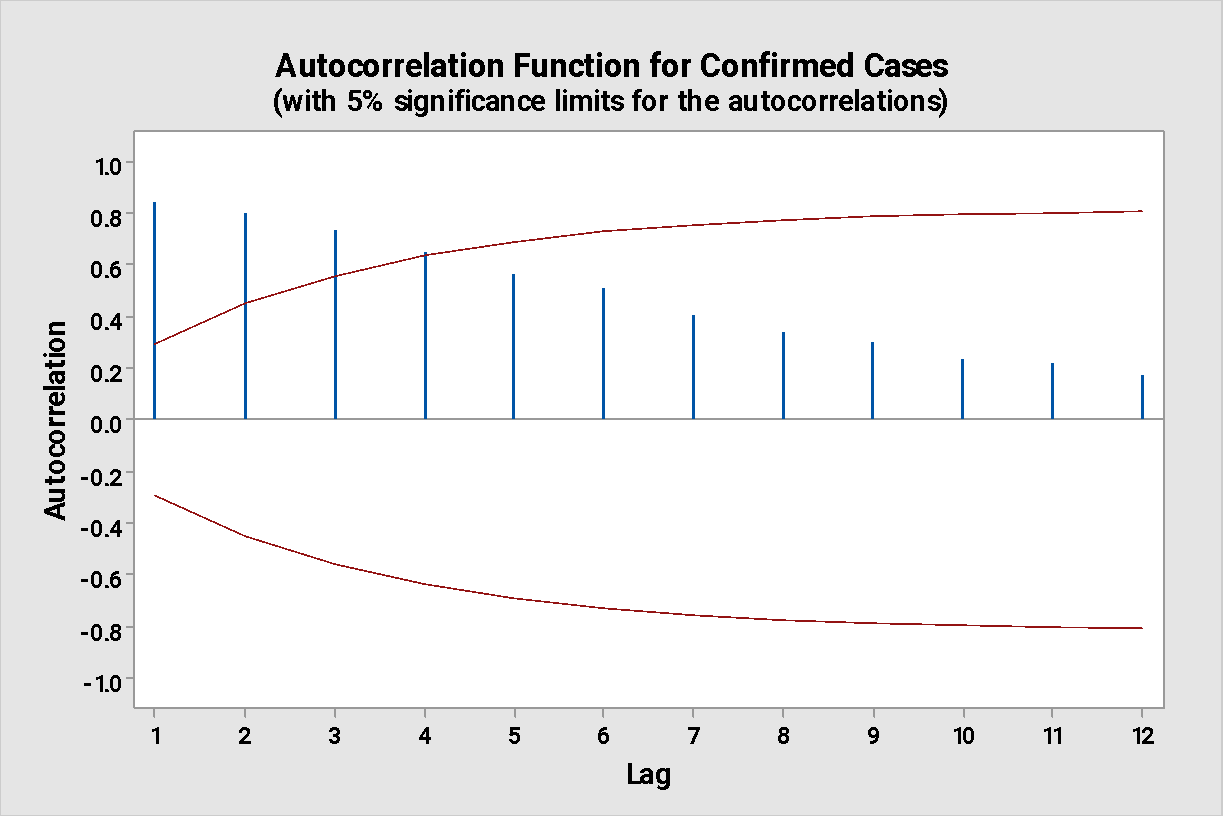


Double Exponential Smoothing for Confirmed Death

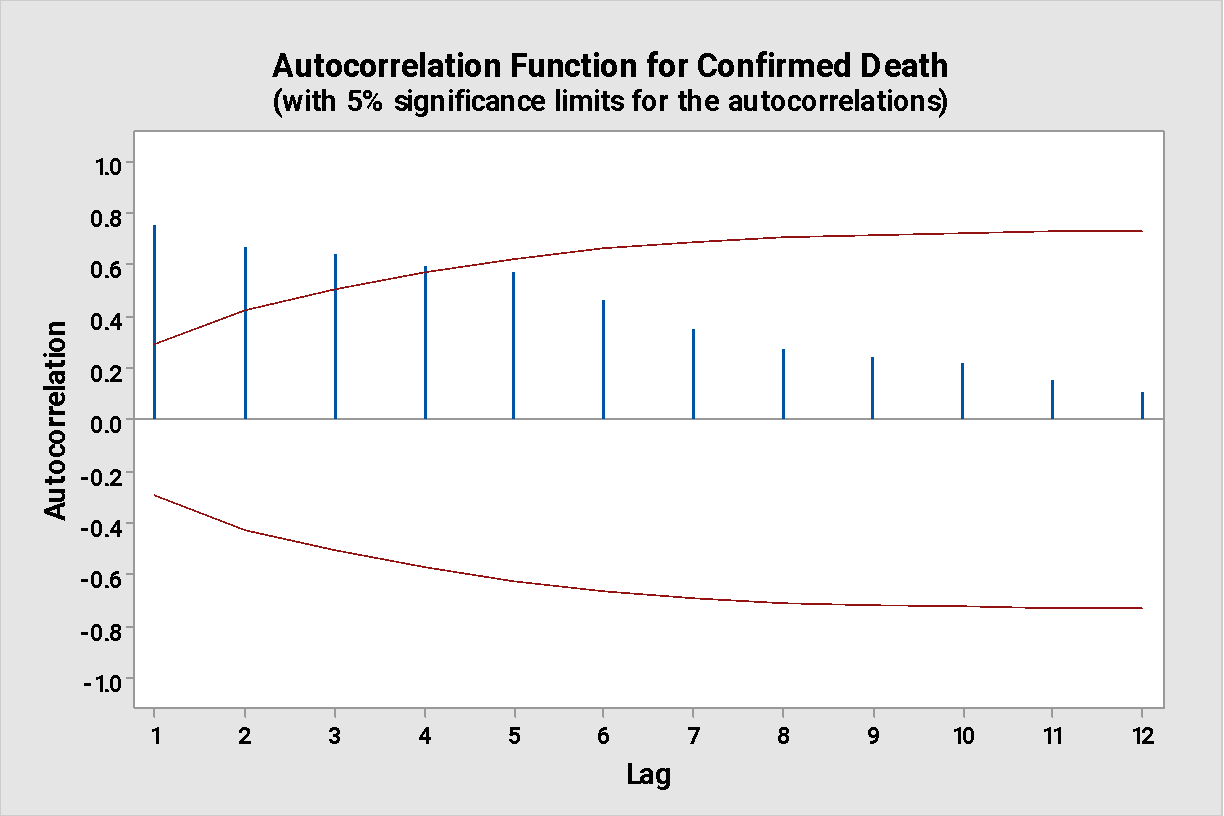


* Autocorrelation Function Plot (ACF plot)

Autocorrelation Function: Confirmed Cases



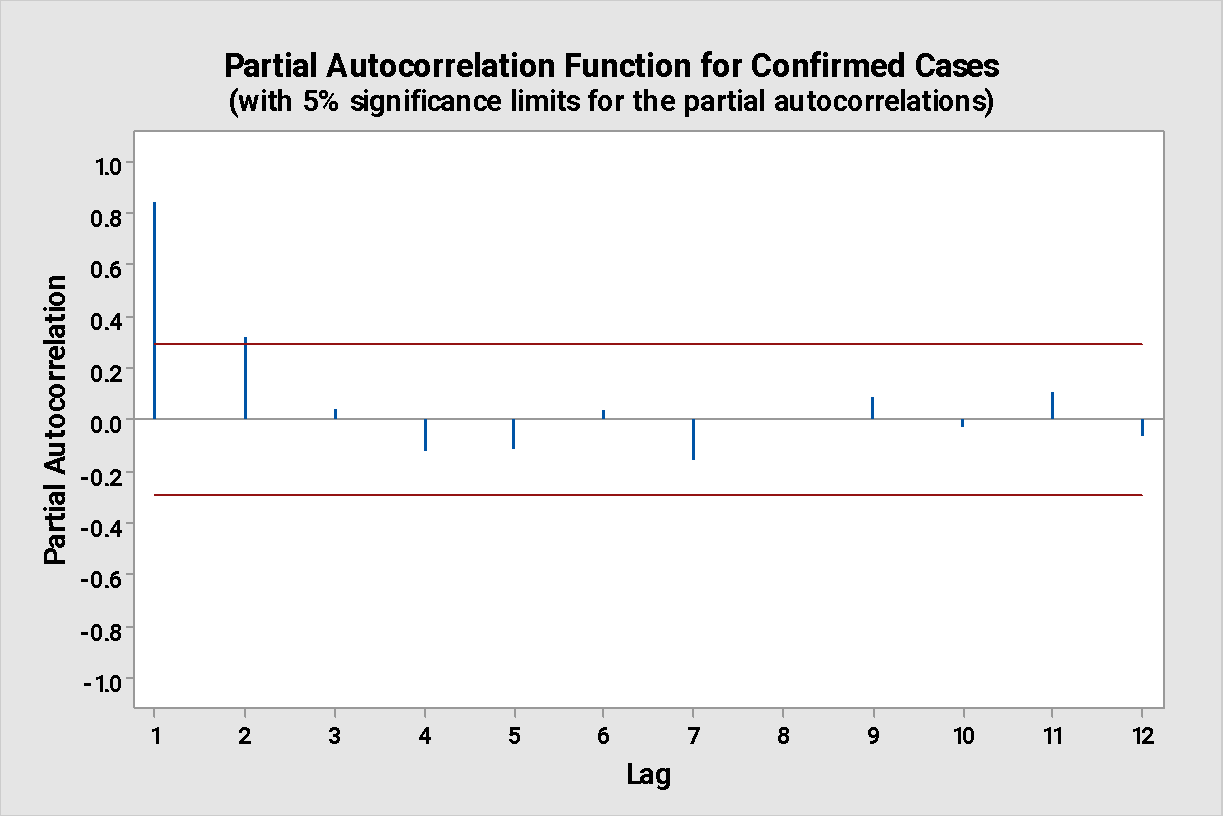
Autocorrelation Function: Confirmed Death



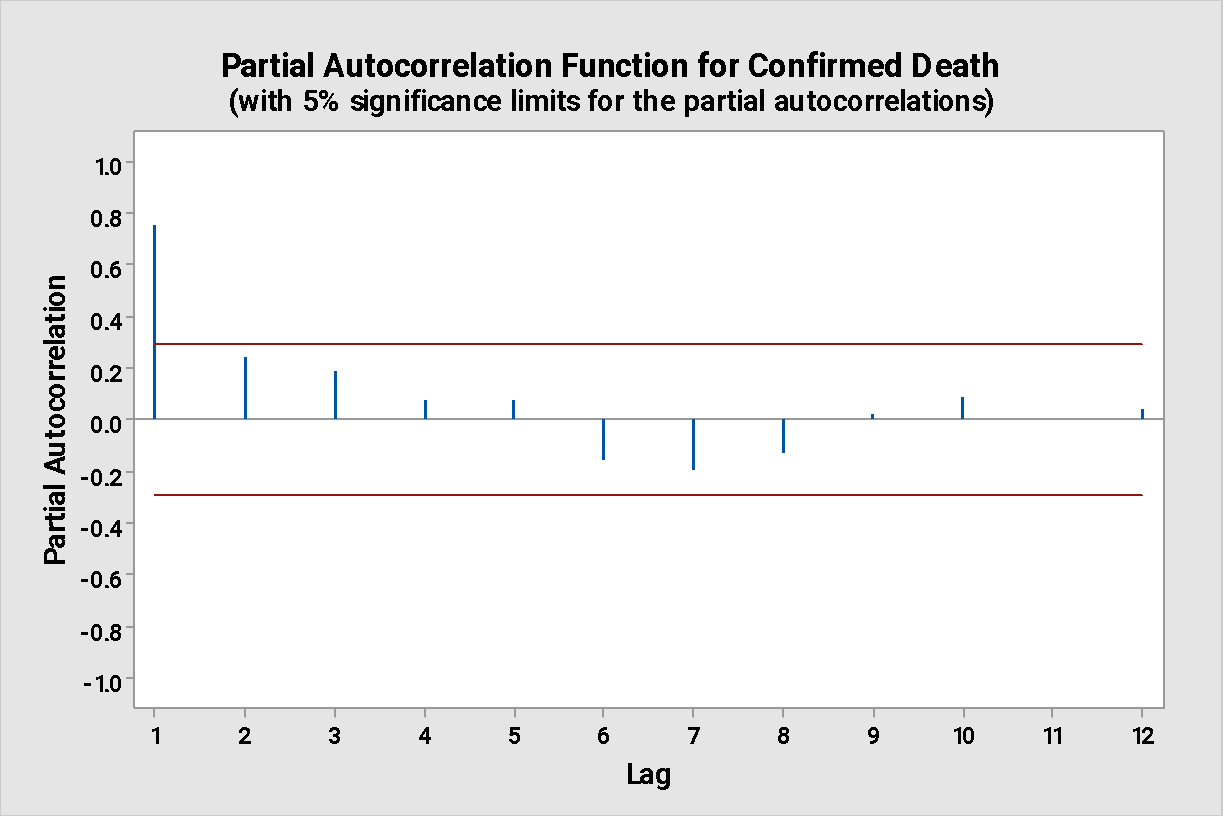
Interpretation:

* We see that both series are not stationary. Three lags are out (3 parameter) so the plots suggests MA(3).
* Partial Autocorrelation Function plot (PACF plot)

Partial Autocorrelation Function: Confirmed Cases

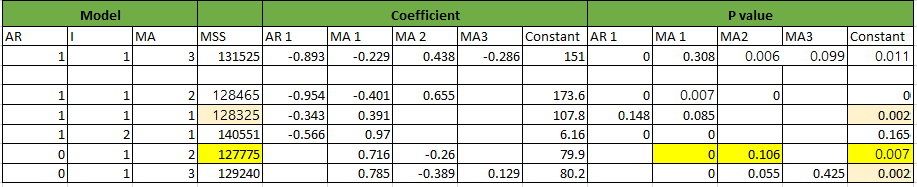


Partial Autocorrelation Function: Confirmed Death



Interpretation:

* The PACF plot for both series shows 1 lag out(1 parameter), suggesting AR(1) model.
* ARIMA for Confirmed Cases



**From above table, ARIMA(0,1,2):**

ARIMA Model: Confirmed Cases

Final Estimates of Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Coef | SE Coef | T-Value | P-Value |
| MA   1 | 0.716 | 0.155 | 4.63 | 0.000 |
| MA   2 | -0.260 | 0.157 | -1.65 | 0.106 |
| Constant | 79.9 | 28.3 | 2.82 | 0.007 |

Residual Sums of Squares

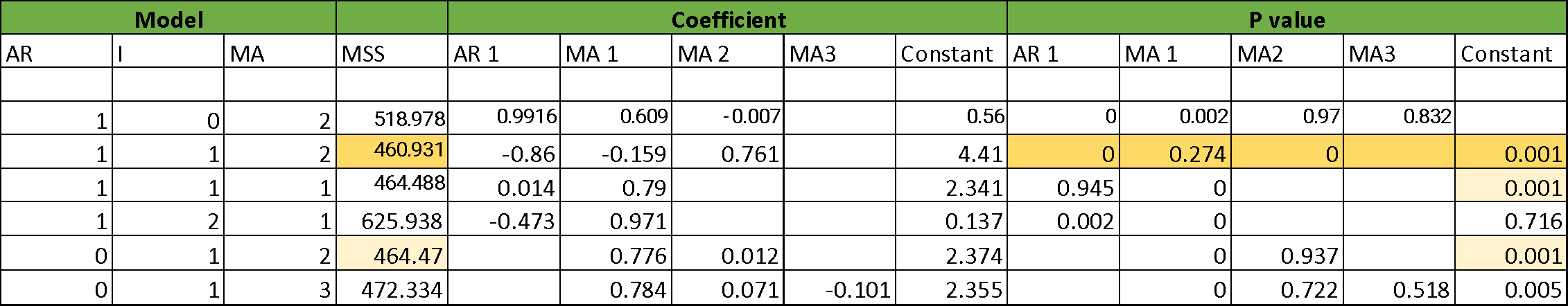
|  |  |  |
| --- | --- | --- |
| DF | SS | MS |
| 44 | 5622096 | 127775 |

*Back forecasts excluded*

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lag | 12 | 24 | 36 | 48 |
| Chi-Square | 8.05 | 17.95 | 22.72 | \* |
| DF | 9 | 21 | 33 | \* |
| P-Value | 0.529 | 0.652 | 0.910 | \* |

ARIMA for Confirmed Death



**From above table, ARIMA(1,1,2):**

ARIMA Model: Confirmed Death

Final Estimates of Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Coef | SE Coef | T-Value | P-Value |
| AR   1 | -0.860 | 0.147 | -5.85 | 0.000 |
| MA   1 | -0.159 | 0.144 | -1.11 | 0.274 |
| MA   2 | 0.761 | 0.110 | 6.89 | 0.000 |
| Constant | 4.41 | 1.25 | 3.54 | 0.001 |

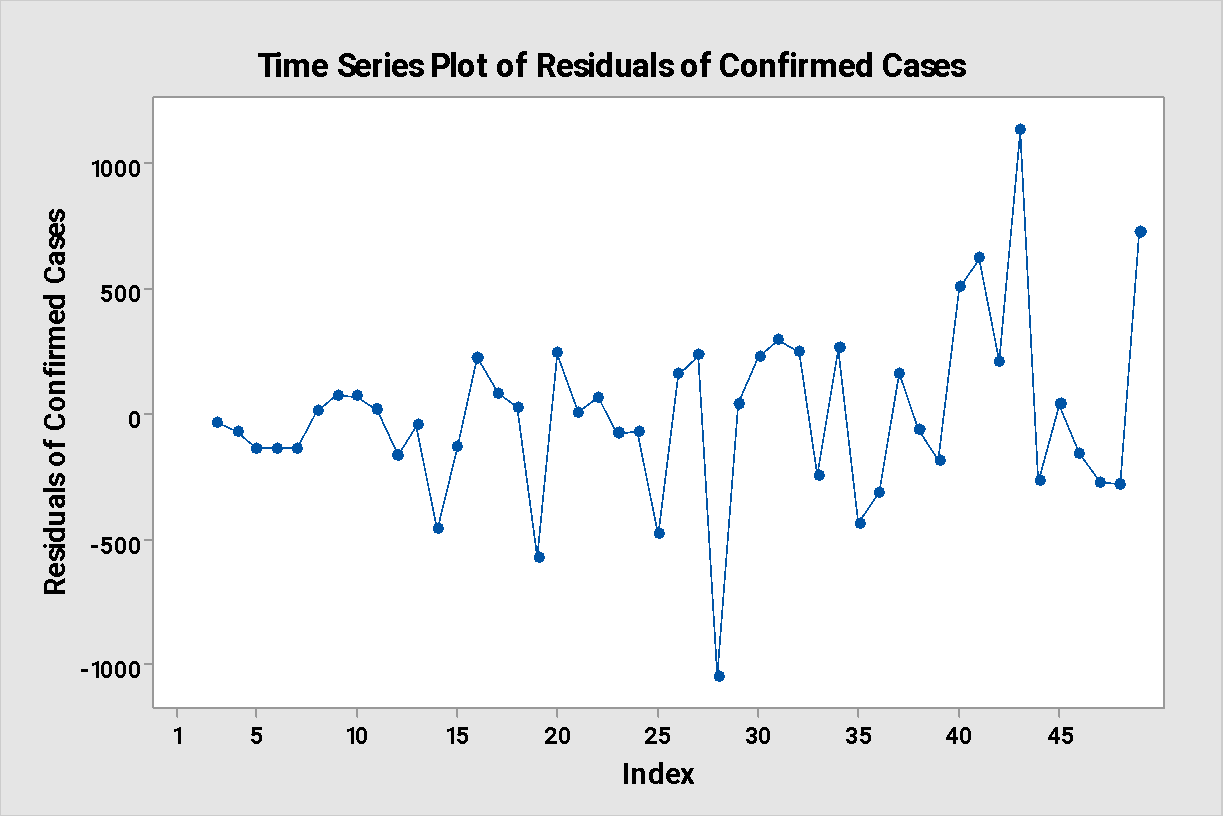
Residual Sums of Squares

|  |  |  |
| --- | --- | --- |
| DF | SS | MS |
| 43 | 19820.0 | 460.931 |

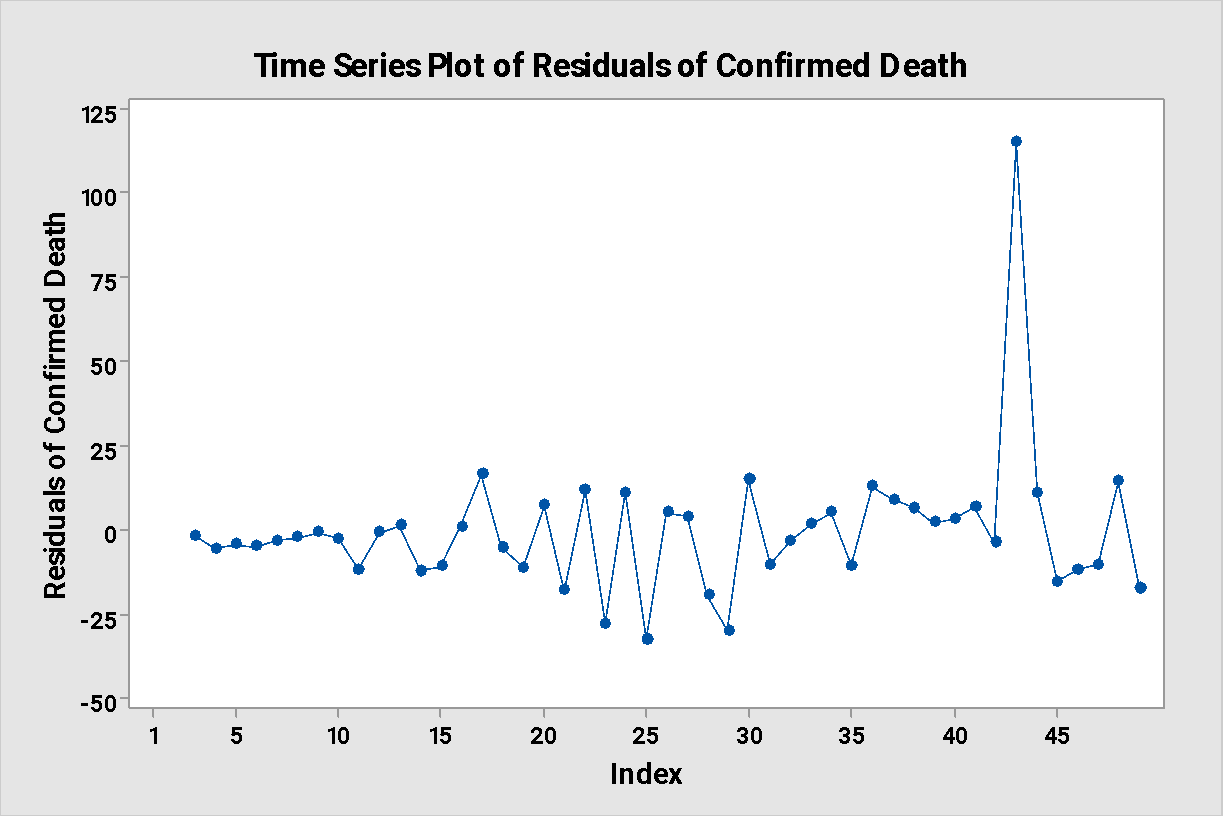
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lag | 12 | 24 | 36 | 48 |
| Chi-Square | 2.60 | 9.92 | 14.86 | \* |
| DF | 8 | 20 | 32 | \* |
| P-Value | 0.957 | 0.970 | 0.996 | \* |

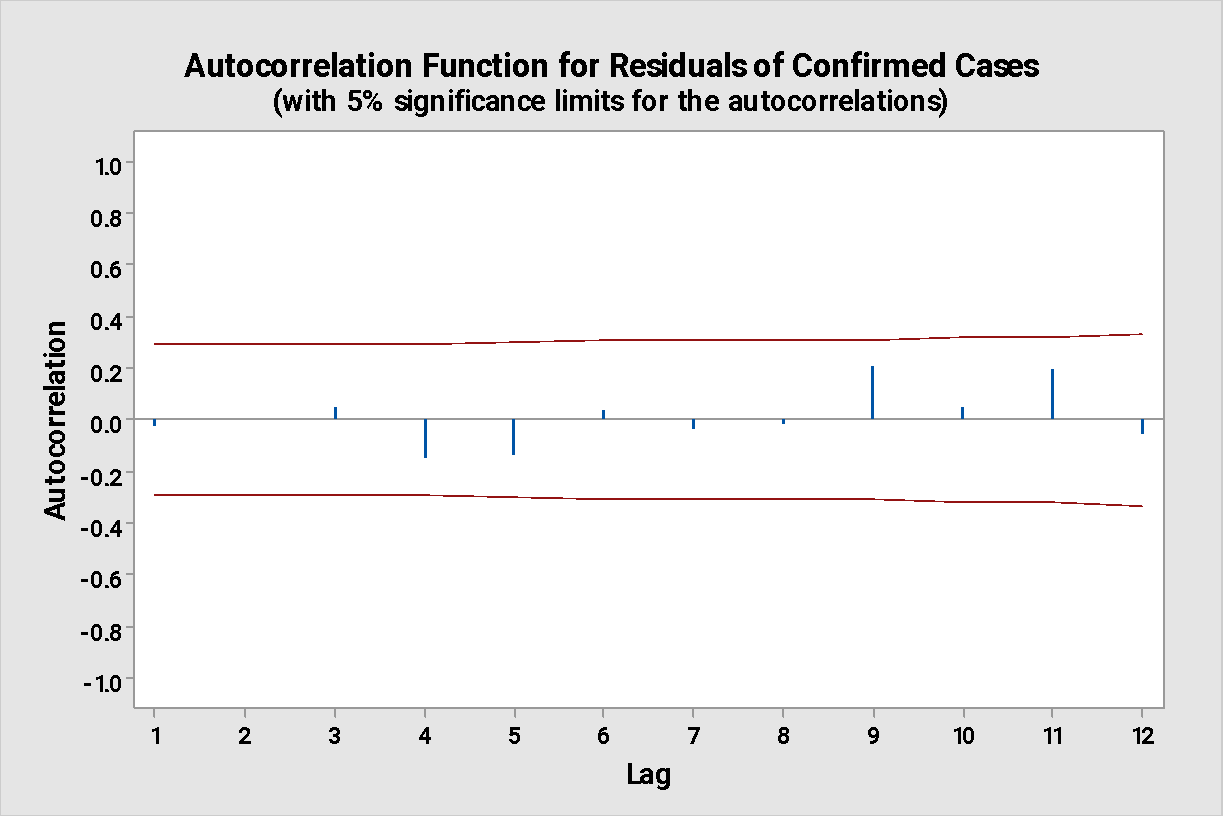
Time series plot of Residuals for Confirmed Cases

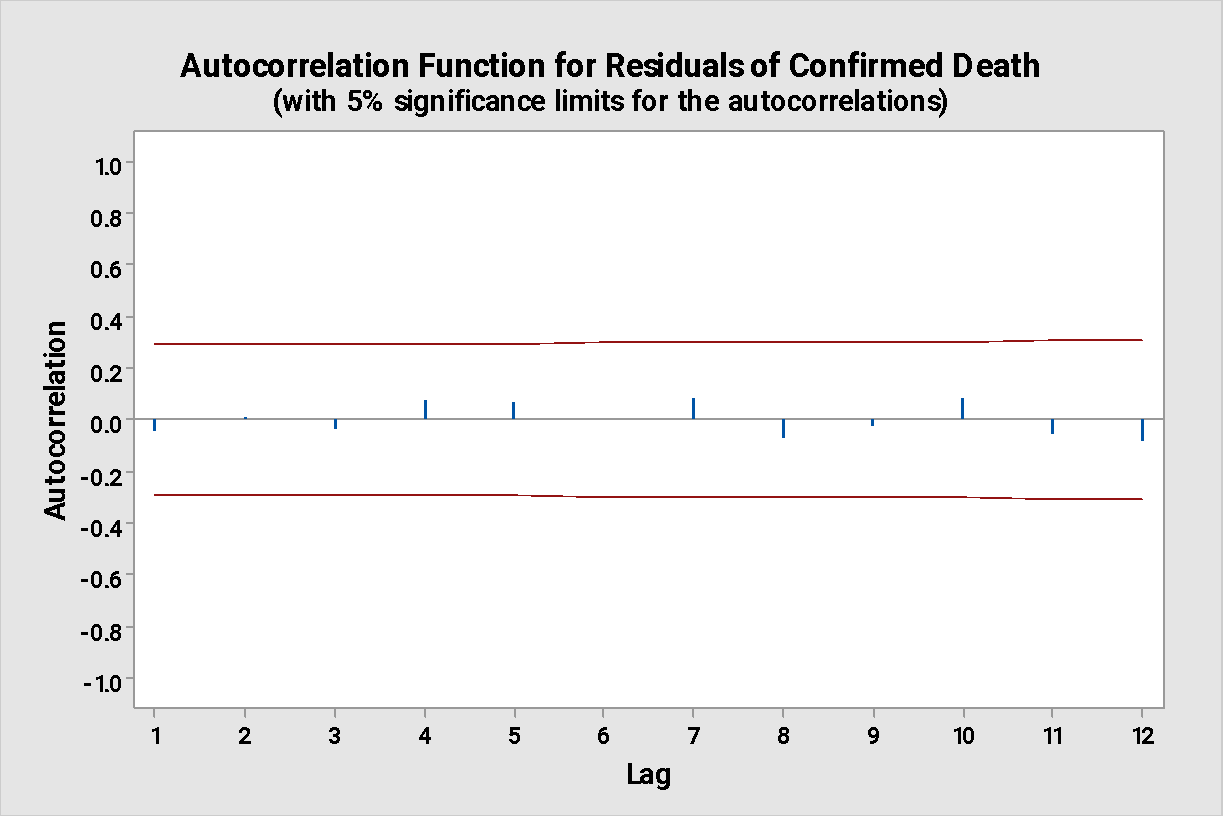


Time series plot of Residuals for Confirmed Death



ACF Plot Residuals of confirmed cases & confirmed death





Interpretation:

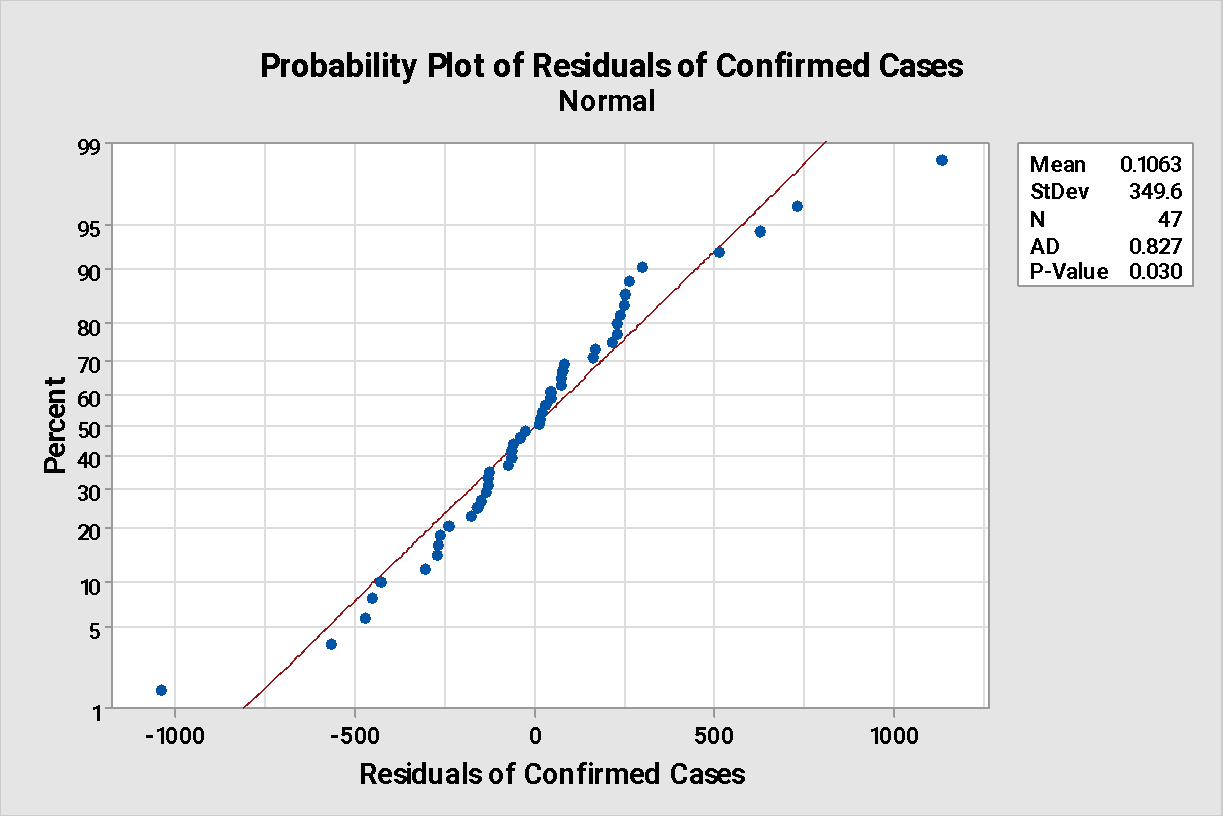
* ACF plots for both series shows that the series are weakly stationary. Therefore, the noise is a white noise.

Normality Test :

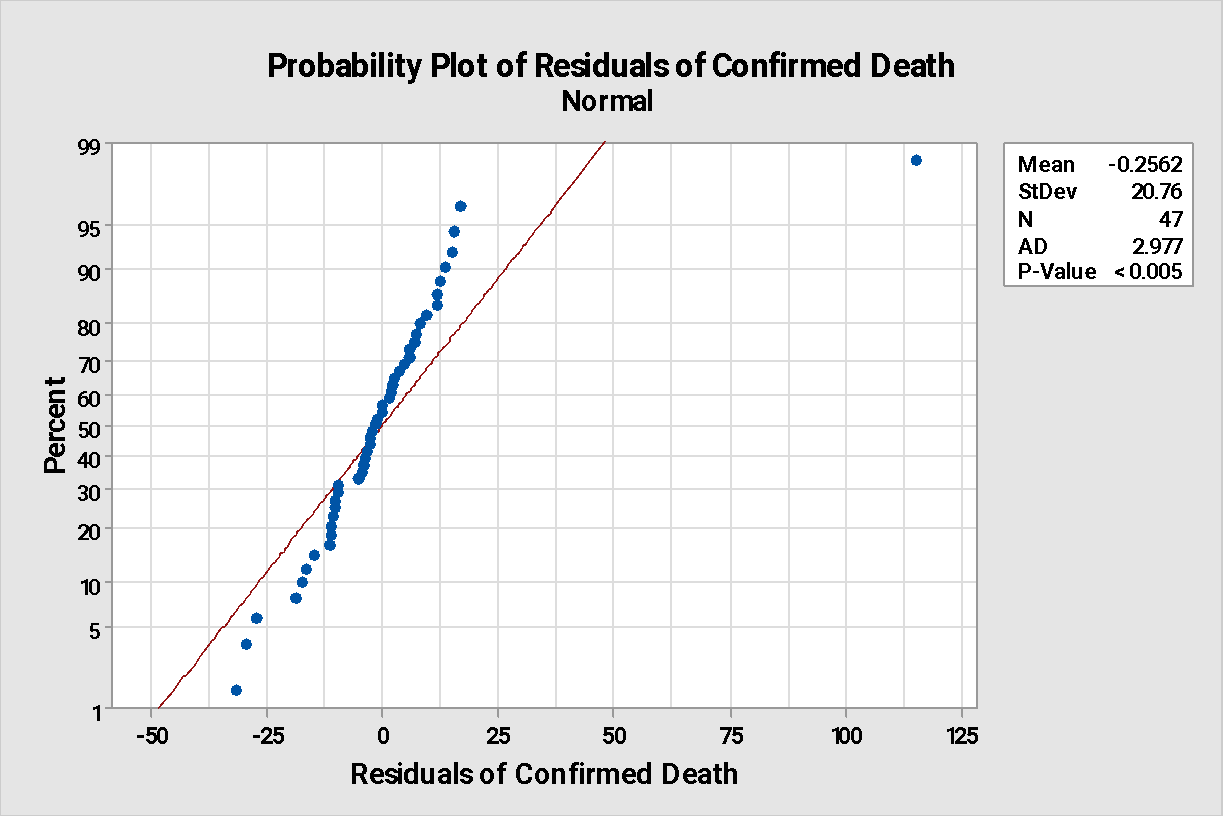
Further to check whether the series has Gaussian white noise. We set the following hypothesis,

H0: Series is normal Vs H1: Series is non-normal

Probability Plot of Residuals of Confirmed Cases

****

Probability Plot of Residuals of Confirmed Death



Interpretation:

1. We see that the p value=0.03<0.05. So we reject the null hypothesis i.e. white noise is not Gaussian noise( series is from a nonnormal distribution).

2. The p value<<0.05 so we reject H0 for the deaths too. The series is from a non-normal distribution.

* Cube root transformation for Covid Cases

After applying the transformation, and fitting an ARIMA model, we sort the residuals and remove the first 3 to Gaussian white noise.

ARIMA Model: CC\_Cube\_rt

Final Estimates of Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Coef | SE Coef | T-Value | P-Value |
| MA   1 | 0.823 | 0.152 | 5.43 | 0.000 |
| MA   2 | -0.023 | 0.152 | -0.15 | 0.878 |
| Constant | 0.2361 | 0.0316 | 7.46 | 0.000 |

Residual Sums of Squares

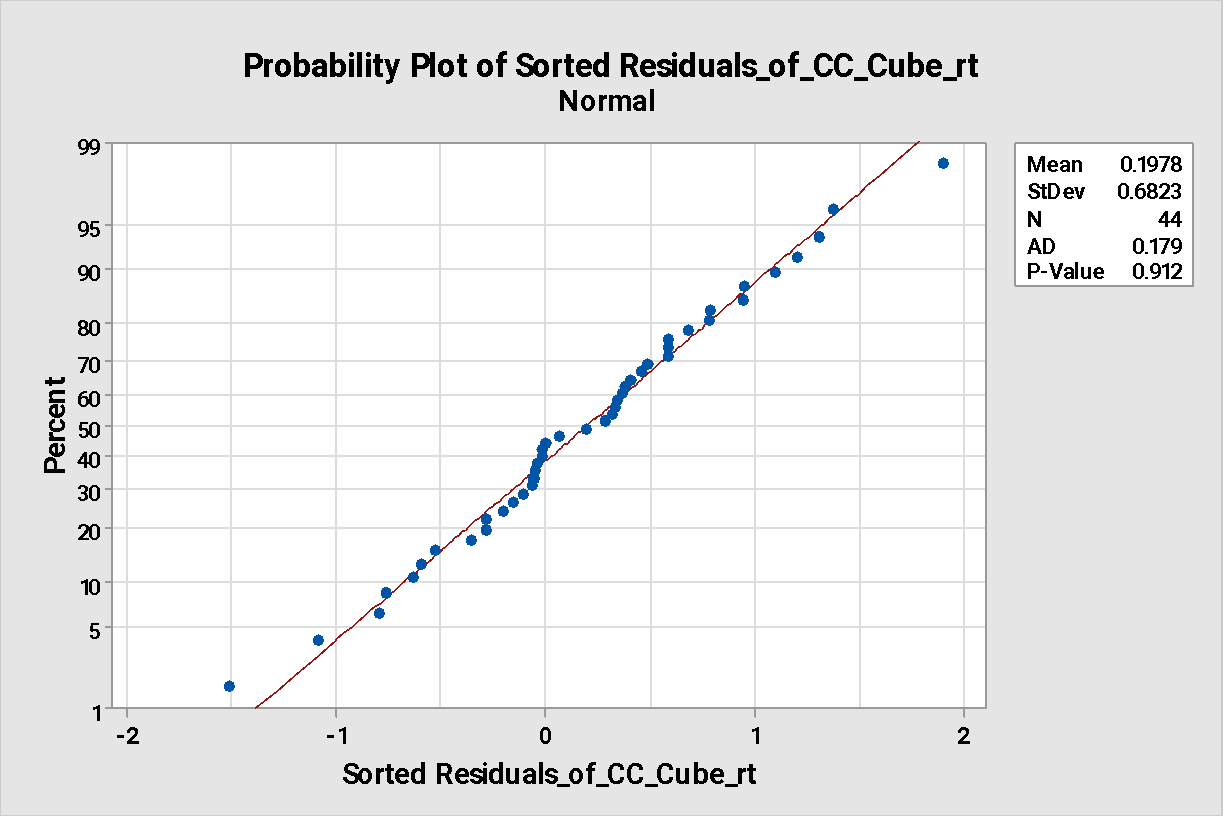
|  |  |  |
| --- | --- | --- |
| DF | SS | MS |
| 44 | 47.7164 | 1.08446 |

*Back forecasts excluded*

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lag | 12 | 24 | 36 | 48 |
| Chi-Square | 6.84 | 24.92 | 31.74 | \* |
| DF | 9 | 21 | 33 | \* |
| P-Value | 0.654 | 0.251 | 0.530 | \* |

Probability Plot of Sorted Residuals\_of\_CC\_Cube\_rt



* Cube root transformation for Covid Death

ARIMA Model: Transformed Deaths

Final Estimates of Parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Coef | SE Coef | T-Value | P-Value |
| AR   1 | -0.58 | 2.74 | -0.21 | 0.835 |
| MA   1 | 0.48 | 2.74 | 0.18 | 0.862 |
| MA   2 | 0.62 | 2.86 | 0.22 | 0.830 |
| Constant | 0.10149 | 0.00304 | 33.43 | 0.000 |

Residual Sums of Squares

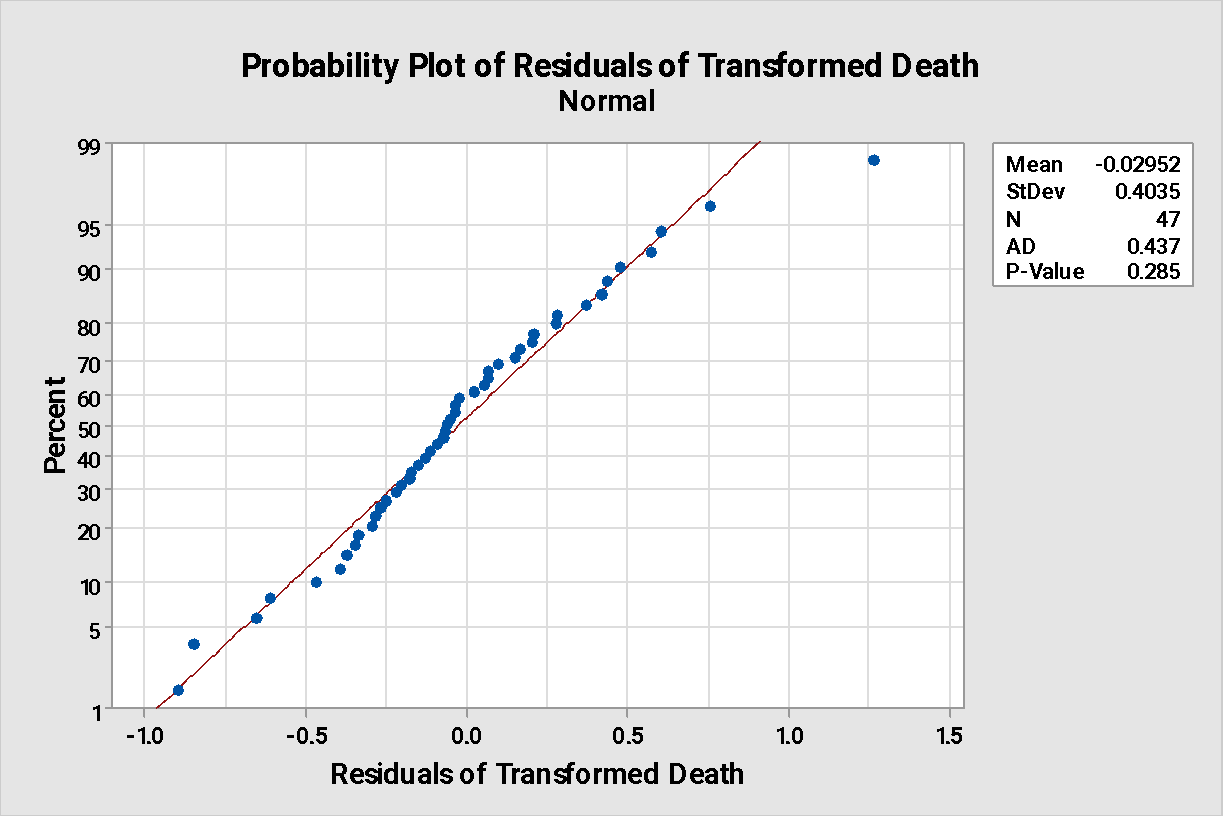
|  |  |  |
| --- | --- | --- |
| DF | SS | MS |
| 43 | 7.52857 | 0.175083 |

*Back forecasts excluded*

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lag | 12 | 24 | 36 | 48 |
| Chi-Square | 9.75 | 28.80 | 39.92 | \* |
| DF | 8 | 20 | 32 | \* |
| P-Value | 0.283 | 0.092 | 0.159 | \* |

Probability Plot of Residuals of Transformed Death



Interpretation:

* It is observed that the p-value >0.05, so we accept the null hypothesis. Hence, the white noise is Gaussian white noise. Also from the ARIMA model, we observe that the MSE value, 1.08446 is the least among all models of the Covid cases series.
* Similarly ,we observe, p-value > 0.05 i.e. We accept H0. And this model also has the least MSE value, 0.175083 among all the other models for deaths.

Therefore, both series are weakly stationary and have Gaussian white noise.

* **Conclusion: What tools do you recommend for analysis with justification**

The time series analysis is done, and the trend components have been smoothed out. Also the series is been transformed by using cube root transformation. The autocorrelation structure of residuals and squared residuals can be examined. If squared residuals shows autocorrelation then ARCH (Autoregressive Conditionally Heteroscedastic) model can be used.