TIME SERIES ANALYSIS

Time series data:

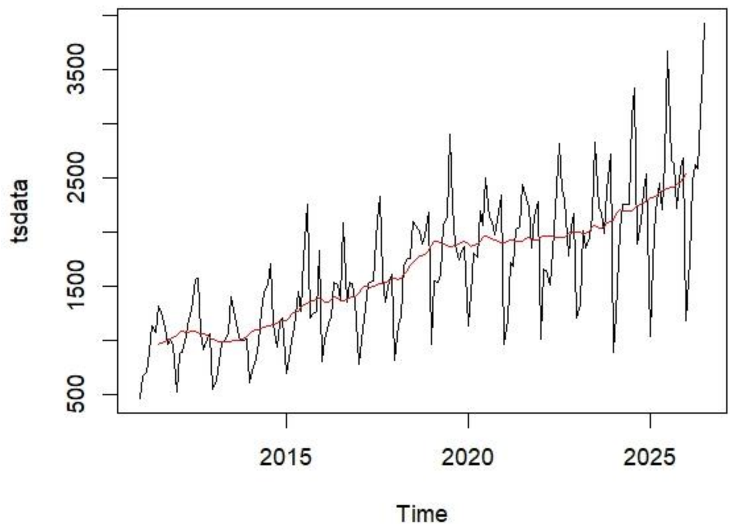

Fig.1

A time series from the years around 2011 through 2025 is shown on the graph. Actual data points with notable fluctuations and an overall upward trend are shown by the black line. By eliminating short-term fluctuations, the red line overlaying the data smooths out the trend component to show the moving average.

Time series with different model fits:

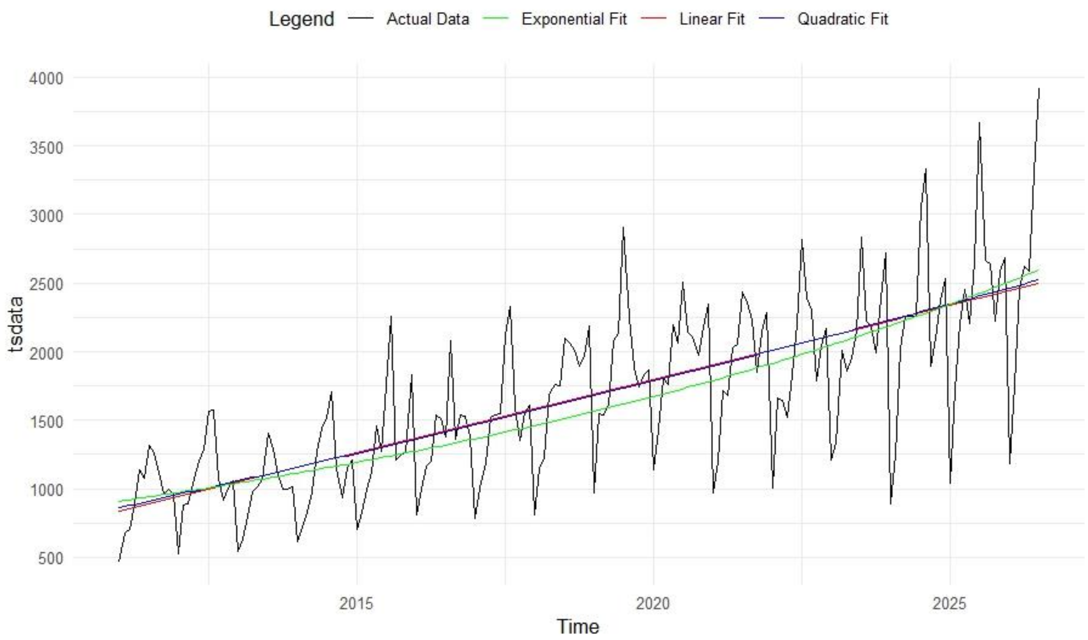

Fig.2

```
> summary(fit_linear)

Call:
lm(formula = tsdata ~ time(tsdata))

Residuals:
     Min       1Q   Median       3Q      Max 
-1335.83  -204.90     3.37   233.49  1428.18 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -2.147e+05  1.424e+04  -15.07   <2e-16 ***
time(tsdata)  1.072e+02  7.056e+00   15.19   <2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 434.1 on 185 degrees of freedom
Multiple R-squared:  0.5551,	Adjusted R-squared:  0.5527 
F-statistic: 230.8 on 1 and 185 DF,  p-value: < 2.2e-16
```

```
> summary(fit_quadratic)

Call:
lm(formula = tsdata ~ time_var + I(time_var^2))

Residuals:
    Min      1Q  Median      3Q     Max 
-1340.9  -206.5     2.0   237.4  1400.7 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.602e+06  7.164e+06   0.363    0.717
time_var      -2.684e+03  7.097e+03  -0.378    0.706
I(time_var^2)  6.912e-01  1.758e+00   0.393    0.695

Residual standard error: 435.1 on 184 degrees of freedom
Multiple R-squared:  0.5554,	Adjusted R-squared:  0.5506 
F-statistic: 114.9 on 2 and 184 DF,  p-value: < 2.2e-16
```

```
> summary(fit_exponential)

Call:
lm(formula = log_tsdata ~ time_var)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.89996 -0.10161  0.02562  0.17715  0.59946 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.293e+02  9.114e+00  -14.18   <2e-16 ***
time_var     6.767e-02  4.515e-03   14.99   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2777 on 185 degrees of freedom
Multiple R-squared:  0.5484,	Adjusted R-squared:  0.5459 
F-statistic: 224.6 on 1 and 185 DF,  p-value: < 2.2e-16
```

Fig.3

- The p-value is highly significant (<0.05) for the Linear and the exponential model. Whereas, it is slightly higher than the threshold for the quadratic model.
- The residual standard error is low in the linear model compared to the other models.
- Considering R-squared, the quadratic model slightly outperforms the linear model, it is not optimal considering its coefficients are not significant.

Considering all factors, the Linear model is the best fit for my data due to its simplicity, and ease of use in the application.

Smoothing the original series:

A breakdown of the time series data into its constituent parts data, trend, seasonal, and remainder is shown in the decomposition graph (Fig.4). The decomposition analysis is useful for choosing a better smoothing method.

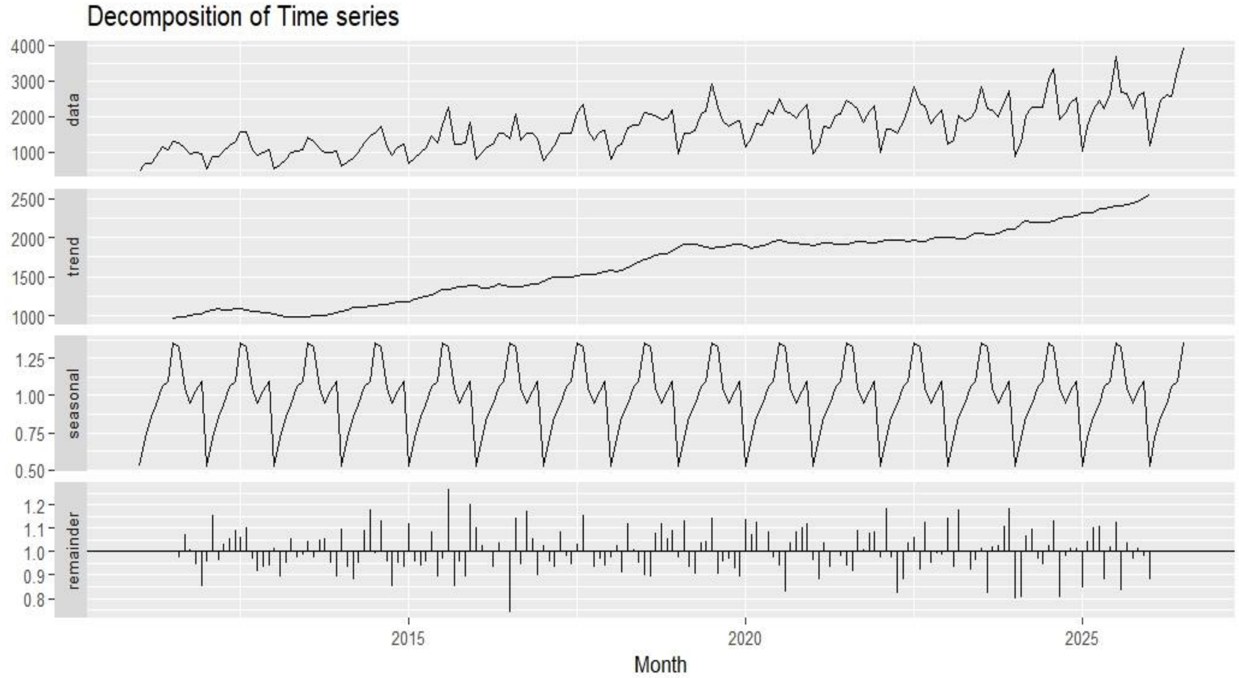

Fig.4

- The raw time series data, which shows both trend and seasonal variations, is shown in the top row.
- The second row shows trend in a gradual upward movement.
- The seasonal component is shown in the third row, which displays a constant pattern repeating throughout the time series.
- The bottom row represents the remainder of the time series. The remainder should ideally resemble white noise, indicating that the majority of the systematic information in the time series has been modeled.

The plot indicates that seasonality and trend are both present. The Holt-Winters smoothing is especially well suited because it includes components for both trend and seasonality. When seasonal variations are roughly constant throughout the series, an additive model is usually applied, and when seasonal variations vary in proportion to the series level, a multiplicative model is applied.

My dataset will be smoothed using the Holt-Winters additive method because the variations are essentially constant throughout the series.

Forecast from Holt-Winters method:

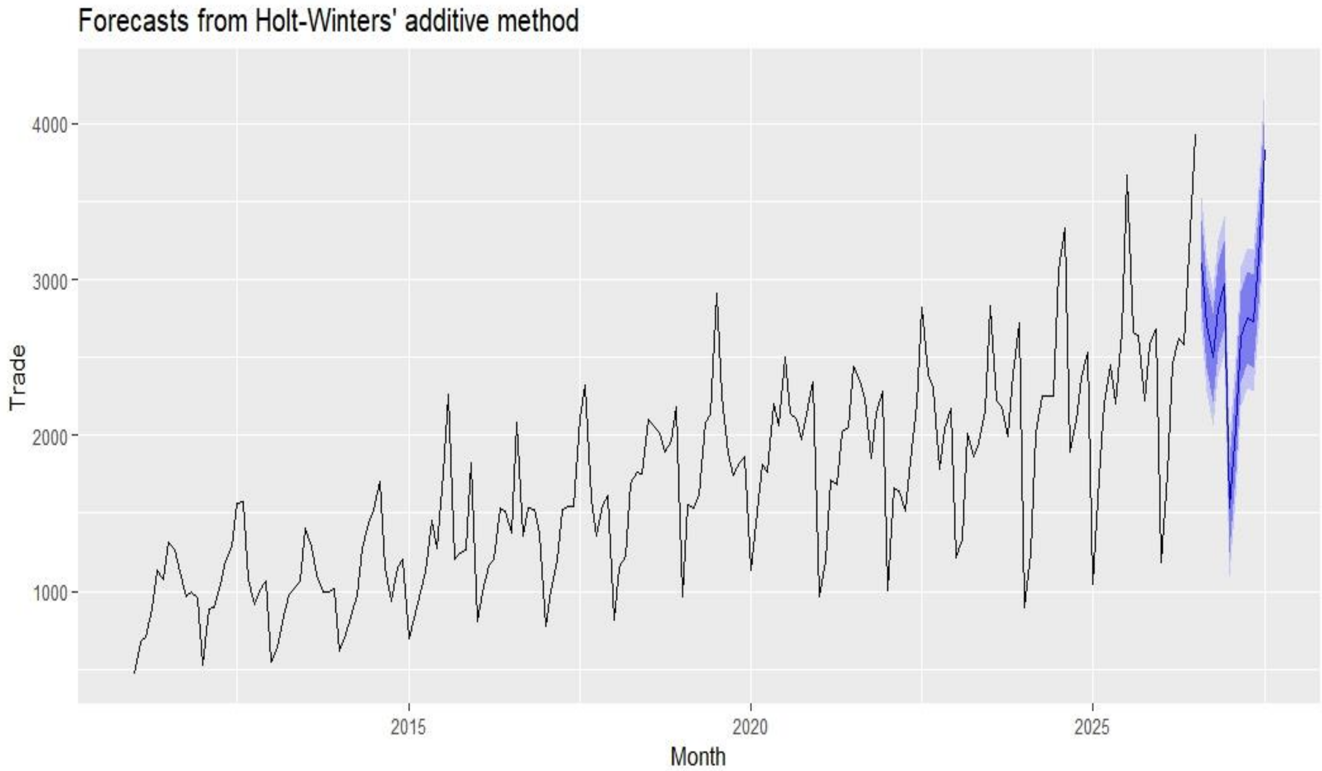

Fig.5

- The graph presented in Fig.5 illustrates the results of applying the Holt-Winters additive method to a time series dataset, which is used in forecasting scenarios that exhibit both trend and seasonality. The black line represents the time series data.
- The future values as predicted by the forecasting model are represented by the blue lines and shaded areas. The point forecasts are represented by the darker blue line in the center of the shaded area. The lighter blue-shaded areas surrounding this line represent the range that is anticipated to contain upcoming data points.
- The forecasts generated by this method extend from the end of the observed data into the future, providing actionable insights for planning and decision-making. The forecast shows approximately one year beyond the last observed data point.

X11 Decomposition:

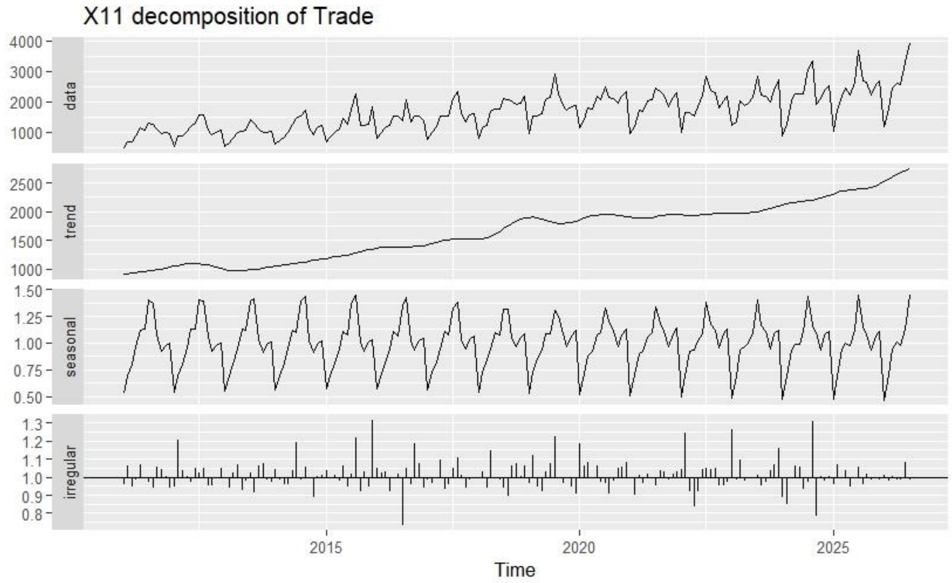

Fig.6

- The seasonality is adjusted in the x11 decomposition compared to the normal multiplicative decomposition shown in Fig.4.
- The X11 decomposition is specifically applied to adjust for seasonality, which helps us analyze data without the repeating seasonal effects.

Seasonally adjusted plot:

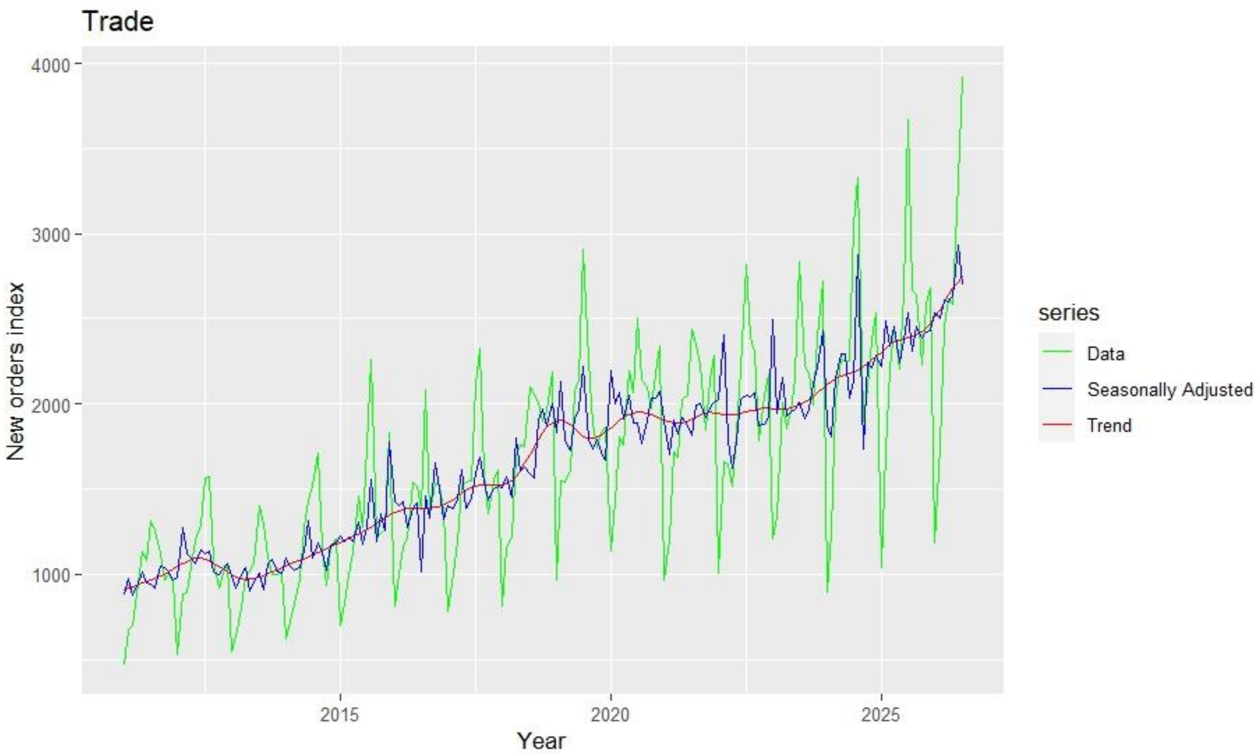

Fig.7

Original Data (Green Line): The green line displays the raw fluctuations over time as represented by the actual recorded data points.

Seasonally Adjusted Data (Blue Line): By correcting the data for seasonal influences, non-seasonal trends and cyclical patterns can be identified and analyzed more clearly.

Trend(Red Line): The long-term trend within the time series is smoothed and represented by the red line. A statistical decomposition process is used to extract this trend from the original data, eliminating both irregular movements and seasonal variations. The general direction over time is represented simply by the trend line.

Monthly Seasonal Patterns:

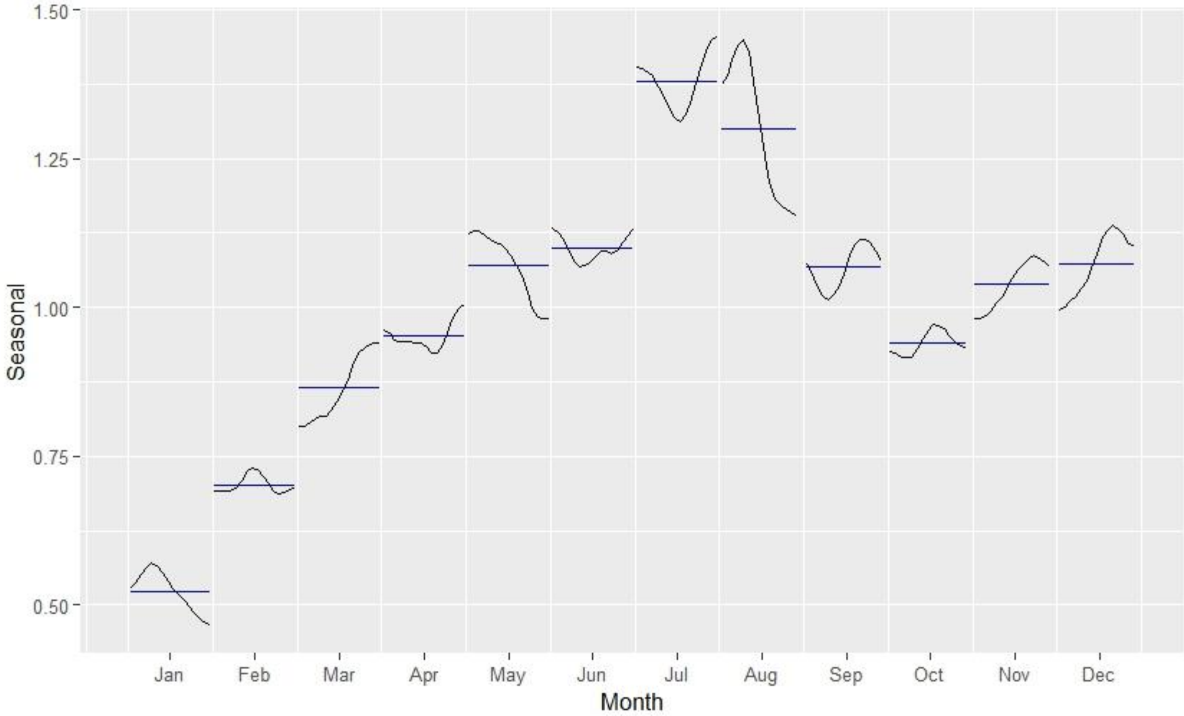

Fig.8

- The graph highlights monthly deviations from the mean and shows the seasonal component of a time series.
- If we notice in Fig.8, In May August, and November, there appear to be a few notable highs and lows. These might be a sign of extraordinary events impacting the data during those months. This is helpful in determining recurring seasonal effects.

Forecasting:

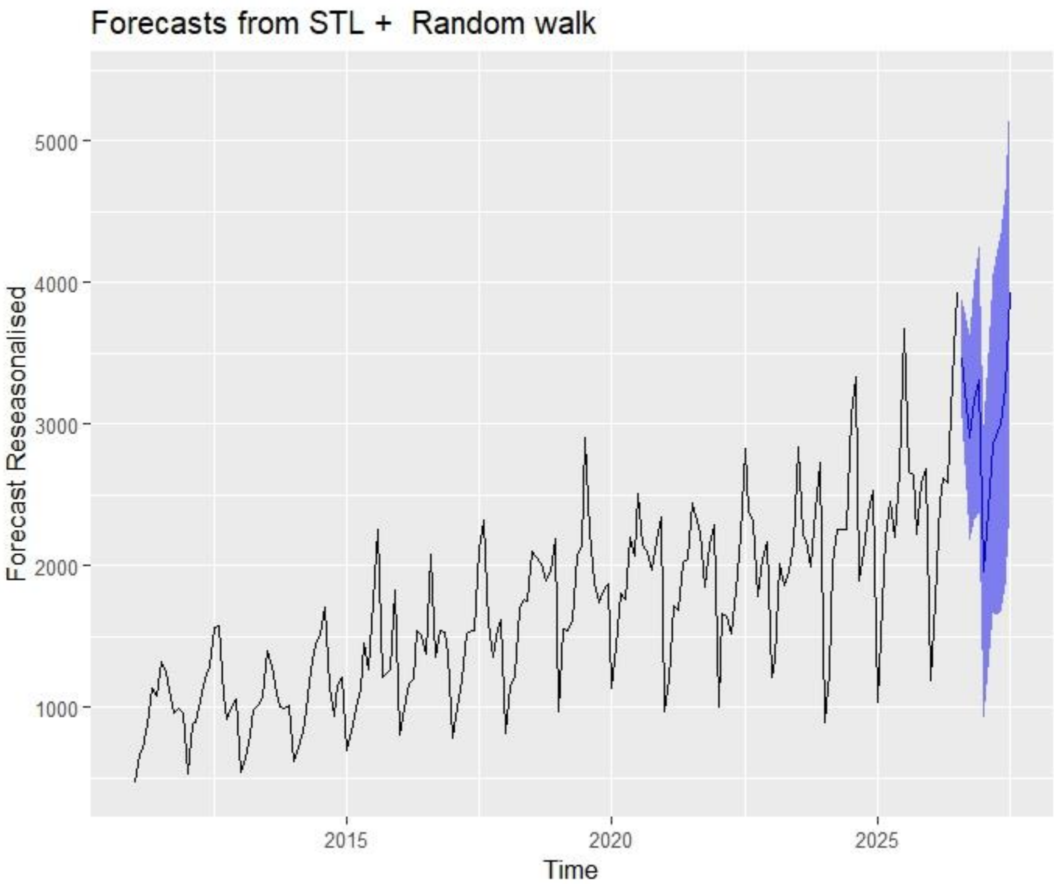

Fig.9

The actual observed values in the time series up to the final known data point are represented by the black line.

The forecast period is shown in the section to the right, which is shaded in blue. The forecast is represented by the blue line at the center. The range where future data points would be located is indicated by the shaded area.

The term "Forecasts Reseasonalised" on the Y-axis indicates that, after the STL decomposition and naive forecasting, the seasonal component was added back to the final forecasted values to preserve the seasonality pattern of the original data.