BUSINESS FORECASTING PROJECT

Unemployment in the labor force in New York

Srishti Patil

Forecasting Question:

What will be the future of Unemployment in the labor force in New York the coming months and years?

Importance:

This forecasting question is important because it helps understand the trends and patterns in unemployment levels within the labor force over time. By predicting future unemployment rates, governments and policymakers can use this information to plan interventions such as employment programs, welfare policies, and job training initiatives. Economists and businesses can anticipate potential labor shortages or surpluses, which can influence decisions related to hiring, salary adjustments, and resource allocation. Furthermore, forecasting unemployment trends helps in understanding the broader social implications of unemployment, such as its impact on poverty, crime, and societal well-being, enabling more targeted community support. For companies, these predictions are valuable for adjusting hiring strategies, identifying potential skills gaps, and preparing for fluctuations in the availability of labor.

Description of the data:

Variable	Description
FIPS Code	A unique identifier representing geographical areas, typically used in the U.S. to identify counties or cities.
State/Area	The name of the state or area being referred to in the dataset.
Year	The year in which the data was recorded.
Month	The month in which the data was recorded, represented as a number from 1 (January) to 12 (December).
Total Civilian Non-Institutional Population in State/Area	The total civilian population in the state or area, excluding those in institutions.
Total Civilian Labor Force in State/Area	The total number of civilians in the labor force (both employed and unemployed) in the state or area.
Percent (%) of State/Area's Population	The percentage of the state's or area's population that makes up the total civilian labor force.
Total Employment in State/Area	The total number of people employed in the civilian labor force within the state or area.

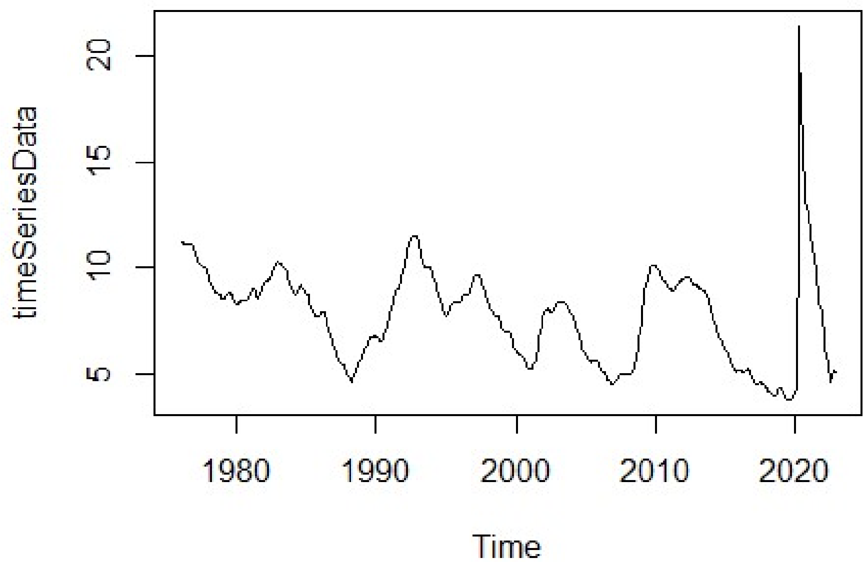
Percent (%) of Labor Force Employed in State/Area	**The percentage of the civilian labor force that is employed.**
Total Unemployment in State/Area	**The total number of unemployed civilians in the labor force in the state or area.**
Percent (%) of Labor Force Unemployed in State/Area	**The percentage of the civilian labor force that is unemployed.**

Accuracy measure and its importance in the forecasted data:

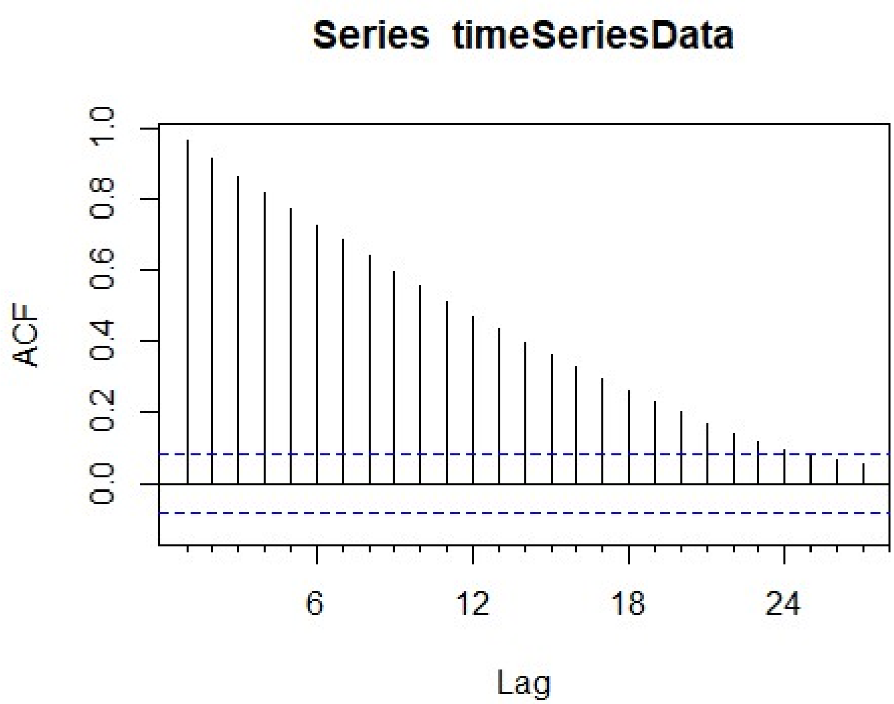
- MAPE provides the forecast error in percentage terms, making it easy to understand how accurate the forecast is, regardless of the unit of measurement.This is particularly helpful when working with unemployment data, as stakeholders (e.g., policymakers, economists, businesses) can easily understand the margin of error in terms of %. For instance, a 5% error in predicting unemployment might be acceptable for some decisions, but not for others (like designing welfare programs).
- MAPE penalizes large errors in a way that is directly proportional to the size of the forecast error relative to the actual value. This means that if the forecast is off by a large amount, MAPE will reflect that higher error more significantly.
- In unemployment forecasting, large errors (e.g., predicting a large drop or increase in unemployment) can have significant real-world consequences, such as misallocation of resources, incorrect policy planning, or misinformed business strategies. By using MAPE, you ensure that your model is sensitive to and corrects large forecast errors.
- A lower MAPE value suggests that the model is accurately capturing the underlying trends in the data. In the case of unemployment forecasting, this means that the model is better at predicting future changes in unemployment levels.
- A lower MAPE means a higher confidence level in the forecasts, which can lead to more informed decisions about labor policies, economic interventions, and employment programs. For example, accurate predictions of unemployment can lead to the better targeting of unemployment benefits or workforce training initiatives.
- If unemployment rates are very small (e.g., < 1%) or very large (e.g., > 15%), the MAPE will normalize the forecast error as a percentage of the actual number, making it easier to assess model accuracy across different contexts. In periods of high unemployment, small absolute errors can translate into large percentage errors, so MAPE is a good way to highlight any discrepancies.

Exploratory Data Analysis:

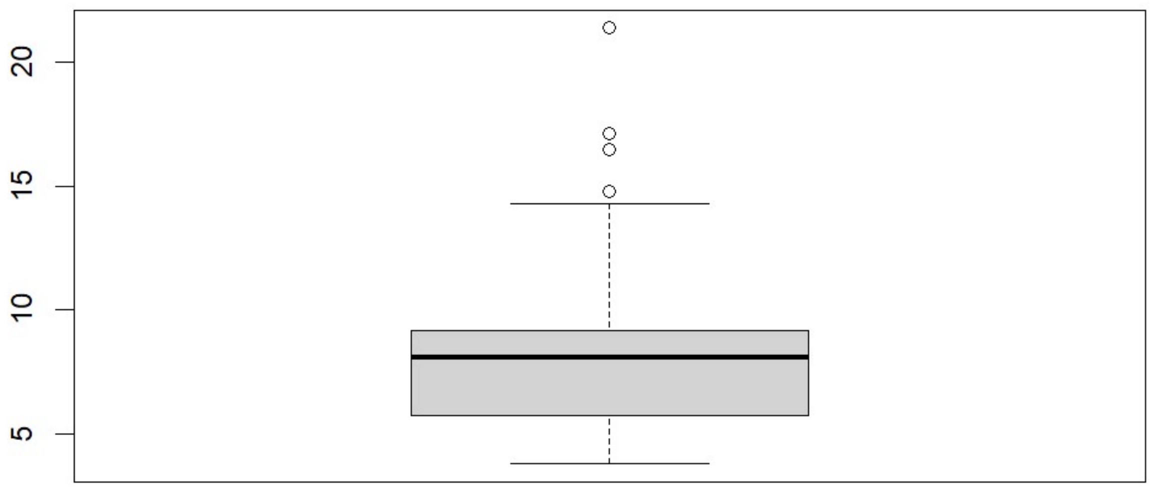
Timeseries data:



ACF:



Boxlpot:



Summary:

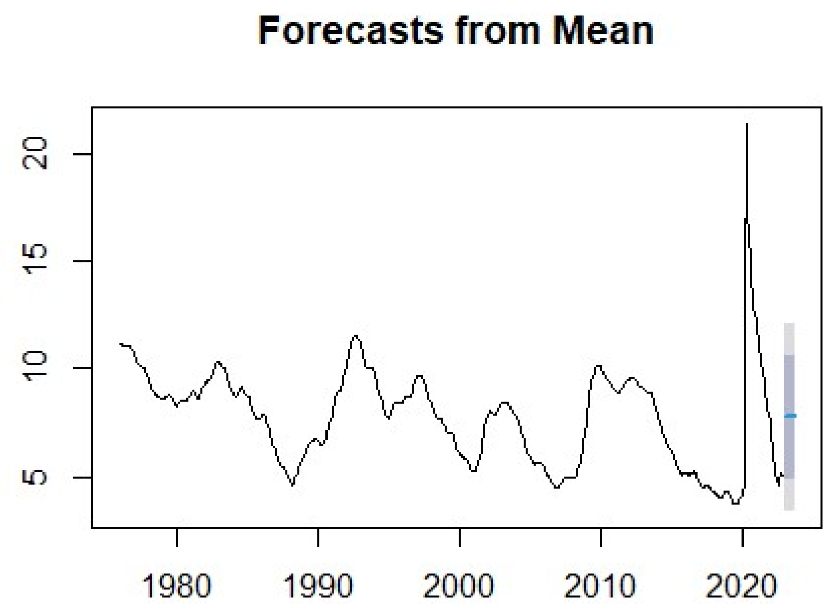
Min. 1st Qu. Median Mean 3rd Qu. Max.

3.8005.775 8.100 7.764 9.200 21.400

- The timeseries plot shows there are cyclical patterns in the unemployment rate in New York over time, with peaks and troughs occurring periodically. A significant spike can be observed around the COVID-19 pandemic (2020), reflecting an economic shock.
- The general trend shows periods of both high and low unemployment, corresponding to economic cycles.
- The autocorrelation function (ACF) plot shows a gradual decline, which suggests strong autocorrelation in the data. This pattern could indicate seasonality or persistent trends in unemployment over time.
- The slow decay of autocorrelation implies non-stationarity, meaning the series has trends or seasonality that need modeling.
- The boxplot shows there are a few outliers, particularly around the spike in unemployment during the COVID-19 pandemic.
- The interquartile range (IQR) is relatively narrow, indicating consistency in the majority of the data.

Insights, residual analysis, Prediction and Accuracy summary from different forecasting methods:

1. **Mean Forecasting**



##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Jan 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Feb 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Mar 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Apr 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## May 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Jun 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Jul 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Aug 2023	7.764007	4.907189	10.62082	3.390606	12.13741
## Sep 2023	7.764007	4.907189	10.62082	3.390606	12.13741

Accuracy:

##	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	-2.984247e-18	2.222627	1.803226	-9.125461	26.7602	1.34275

##	ACF1
## Training set	0.9642393

Accuracy Metrics:

- ME: Near zero (-2.98e-18), indicating unbiased forecasts.
- RMSE: Moderate at 2.22, reflecting overall error magnitude.

- MAE: At 1.80, showing the average deviation from observed values.
- MPE is negative suggesting under-prediction.
- MAPE: High at 26.76%, indicating significant relative errors.
- ACF1: Strong autocorrelation (0.96) suggests forecast residuals are not independent.

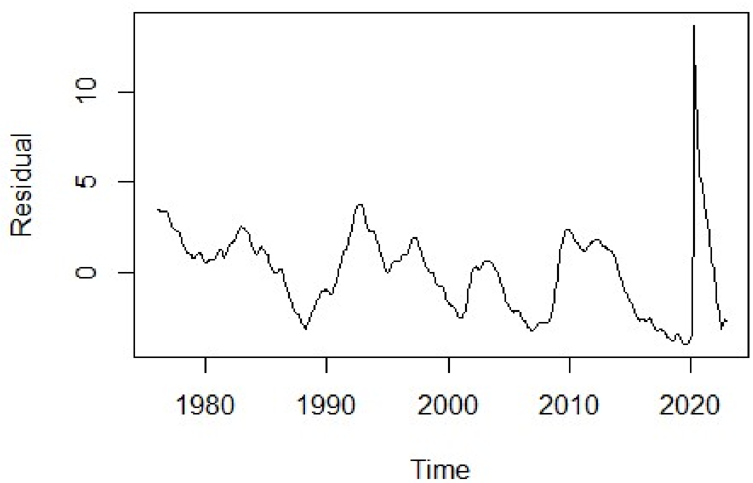
Prediction:

- The point forecast remains consistent at 7.764 across all months.
- Prediction intervals (Lo 80 to Hi 95) widen as confidence decreases, ranging from 4.91 (Lo 80) to 12.14 (Hi 95).

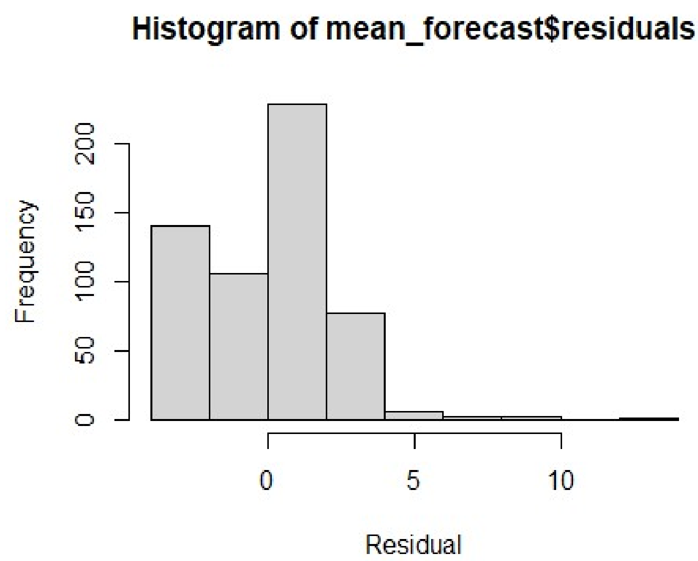
Insights:

The forecasts are consistent but exhibit significant errors and autocorrelation. This highlights potential issues in capt uring variability or model structure.

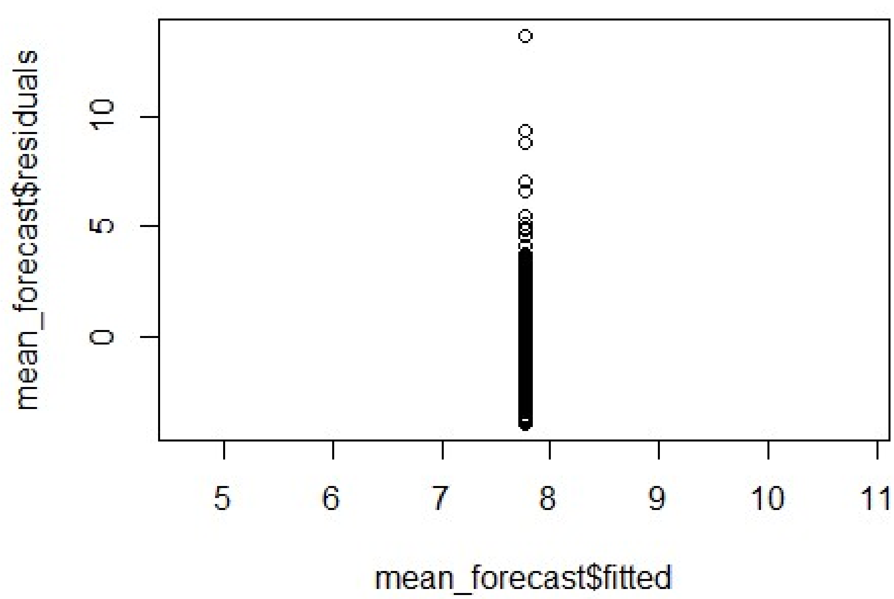
Residual Analysis:

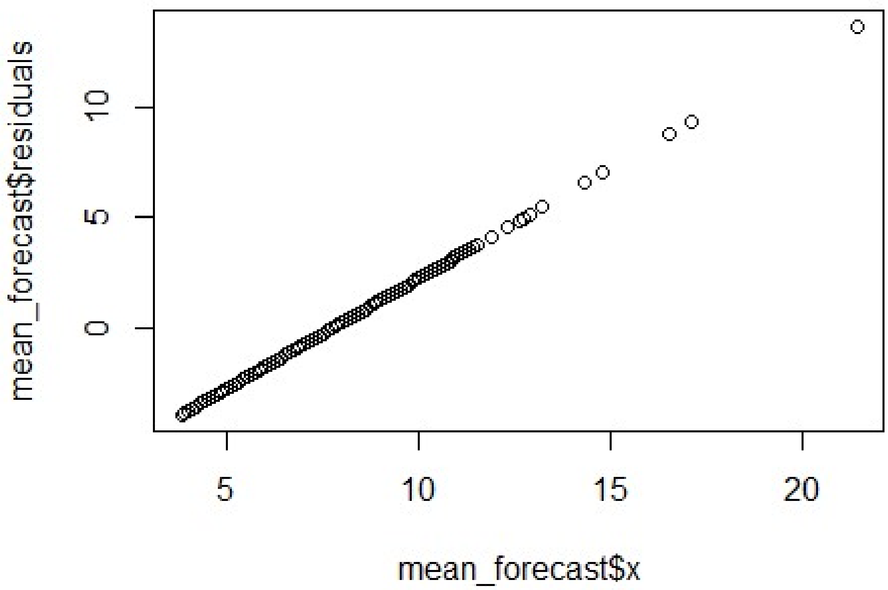


The residuals exhibit a clear pattern over time, suggesting that the model has not fully captured the underlying structure of the data. This indicates potential model misspecification or the presence of unobserved factors influencing the residuals.



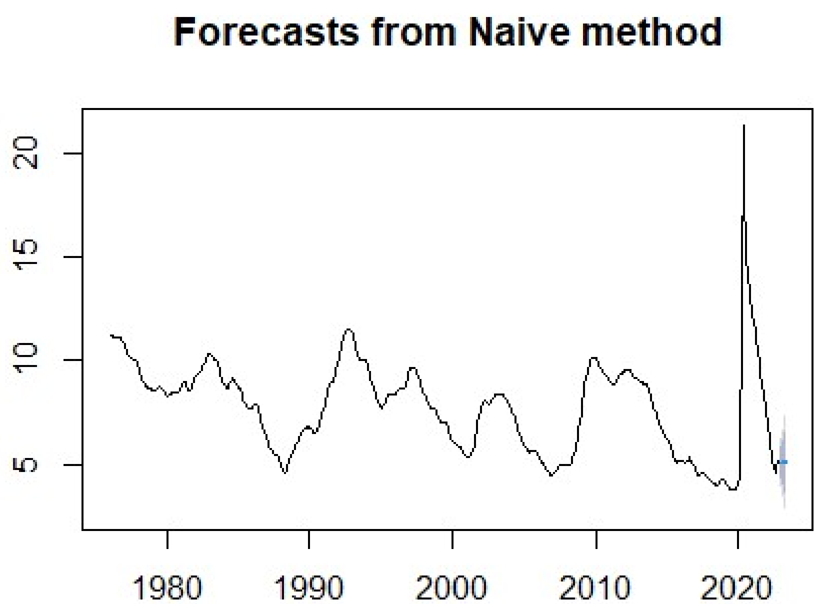
The residuals appear to be right-skewed, indicating that the model may overestimate the values more frequently than it underestimates them. This suggests potential model misspecification or the presence of outliers in the data.





Both The plot suggests that the residuals have a non-constant variance, with larger residuals associated with larger fitted values. This indicates heteroscedasticity.

2. **Naïve Forecast**



##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Jan 2023	5.1	4.374623	5.825377	3.990631	6.209369
## Feb 2023	5.1	4.074161	6.125839	3.531115	6.668885
## Mar 2023	5.1	3.843609	6.356391	3.178516	7.021484
## Apr 2023	5.1	3.649245	6.550755	2.881262	7.318738
## May 2023	5.1	3.478007	6.721993	2.619375	7.580625

Accuracy:

##	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	-0.01083481	0.566015	0.1571936	-0.260081	1.822557	0.1170523

##	ACF1
## Training set	0.2870979

Accuracy Metrics:

- ME: Close to zero (-0.0108), showing minimal bias in predictions.
- RMSE: Low at 0.566, indicating good overall accuracy.
- MAE: Small at 0.157, reflecting small average deviations.
- MPE: Negligible, indicating minimal under-prediction.
- MAPE: Very low (1.82%), showing high relative accuracy.
- ACF1: Moderate autocorrelation (0.287) suggests some dependency in residuals.

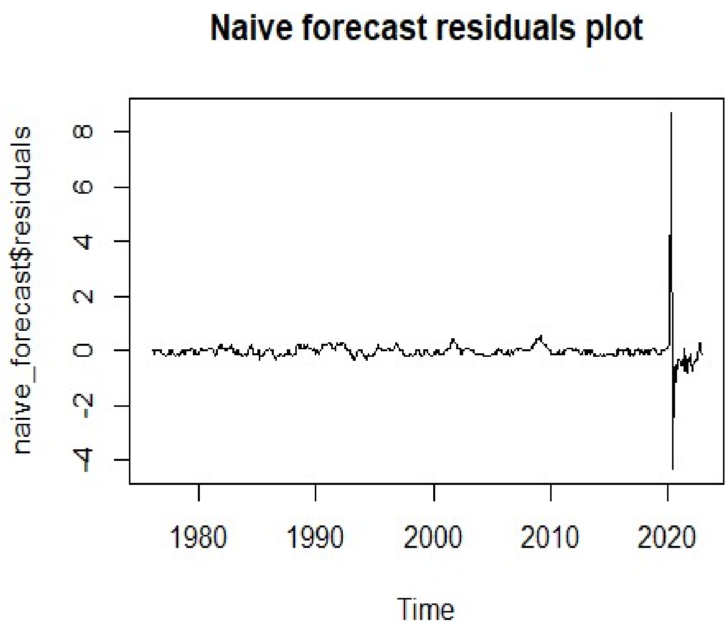
Prediction:

- The Point Forecast is stable at 5.1 across months.
- Prediction intervals widen over time:
    - Lo 80 to Hi 80: From 4.37 to 5.83 (Jan 2023) to 3.48 to 6.72 (May 2023).
    - Lo 95 to Hi 95: From 3.99 to 6.21 (Jan 2023) to 2.62 to 7.58 (May 2023).

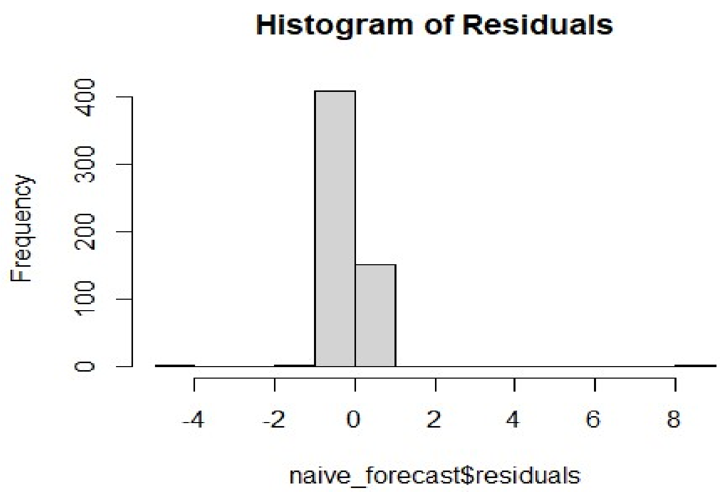
Insights:

This forecasting method demonstrates high accuracy, with small errors and minimal bias. However, the increasing p rediction intervals and moderate residual autocorrelation indicate some uncertainty over longer horizons.

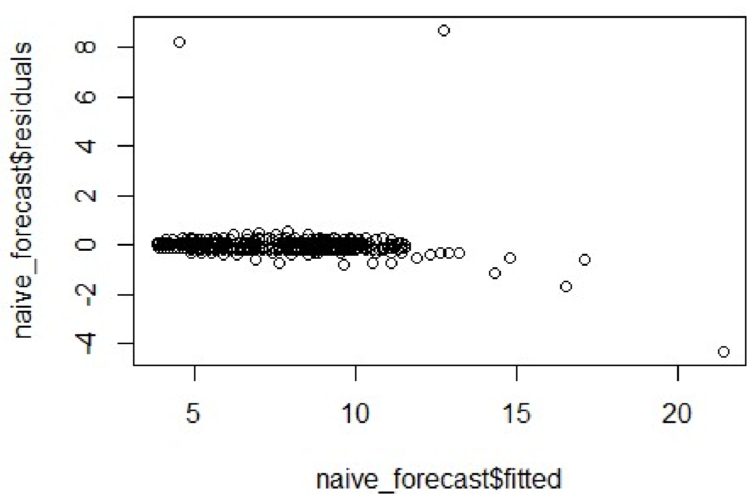
Residual Analysis:

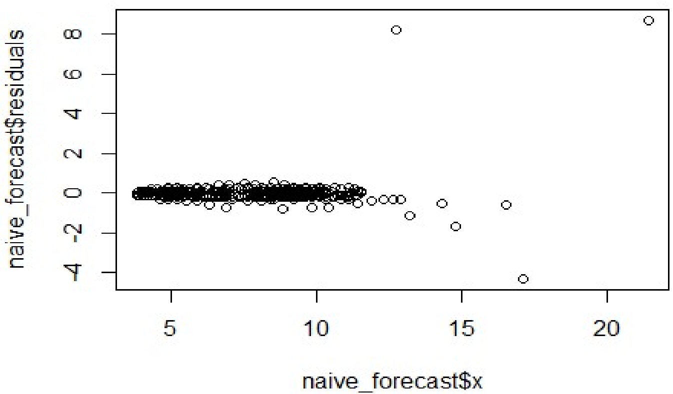


The residual plot shows a non-random pattern, particularly towards the end of the time series. This indicates that the naive forecast model is not capturing the underlying pattern in the data, leading to systematic underestimation or overestimation.

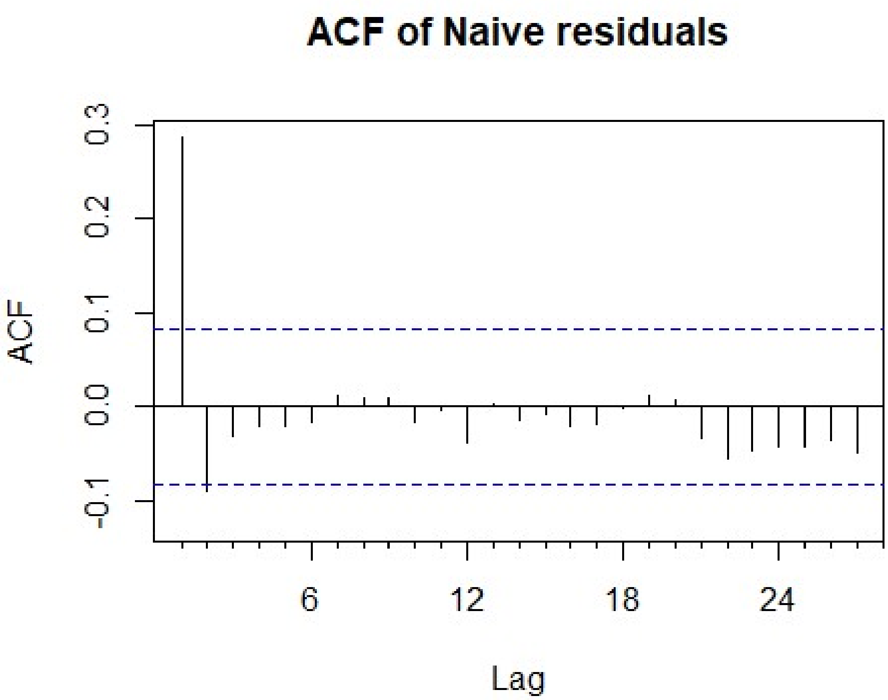


The residuals appear to be slightly right-skewed, indicating that the naive forecast model may overestimate the values more frequently than it underestimates them. This suggests potential model misspecification or the presence of outliers in the data.



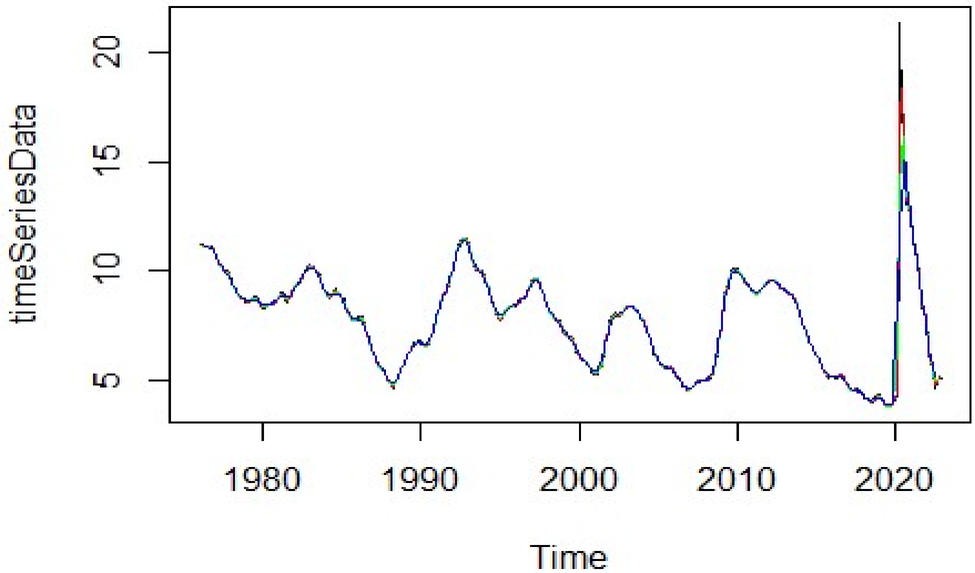


Both the plot suggests that the residuals have a non-constant variance, with larger residuals associated with larger fit ted values. This indicates heteroscedasticity.



The ACF plot of the naive residuals shows significant autocorrelation at lag 1, indicating that the residuals are not independent. This suggests that the naive model is not capturing the underlying structure of the data and may need improvement.

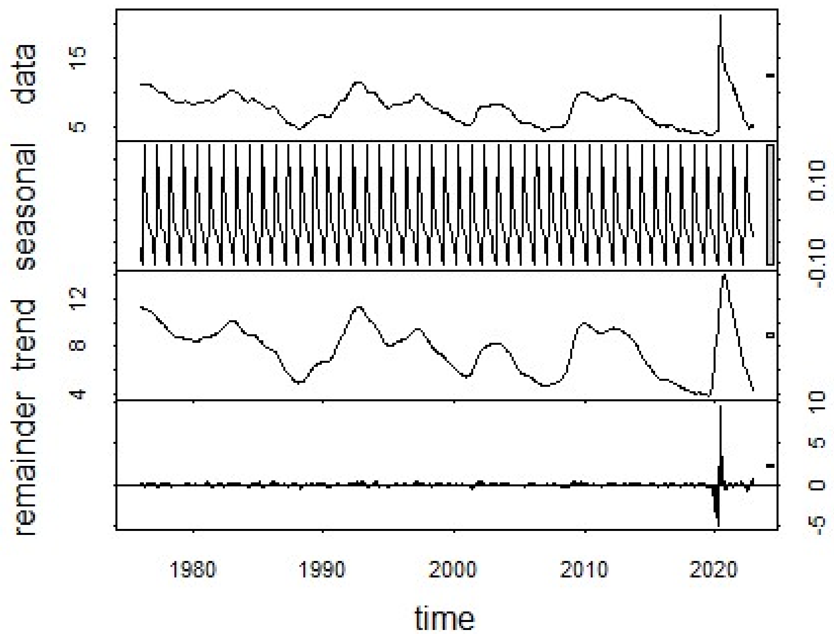
3. ***Moving Averages***



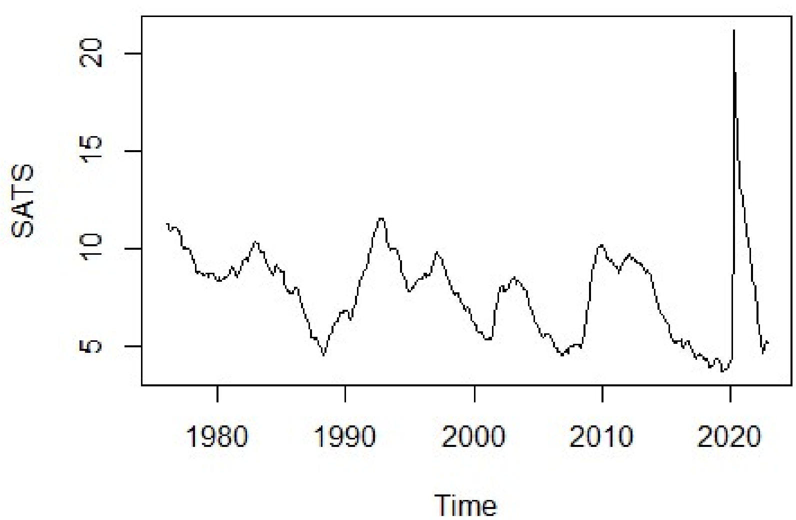
Insights:

The graph showcases a time series dataset with moving averages applied to smooth the data, revealing underlying trends and patterns. The moving averages help reduce noise, making it easier to observe long-term trends and cyclical behavior, as seen in the recurring rises and falls over time. They effectively highlight seasonality by averaging out short-term fluctuations, emphasizing the periodic nature of the data. A notable anomaly around 2020, marked by a sharp spike, stands out, although the moving averages dampen its impact on the overall trend. This smoothing technique allows for a clearer comparison of fluctuations across different time periods, with pre-2000 showing more stability compared to post-2000. Overall, moving averages provide valuable insights into the time series by enhancing trend and seasonality detection while reducing the influence of irregular variations.

4. ***Decomposition***



#seasonal adjustment
SATS <- **seasadj**(stl_decomposition)
plot(SATS)

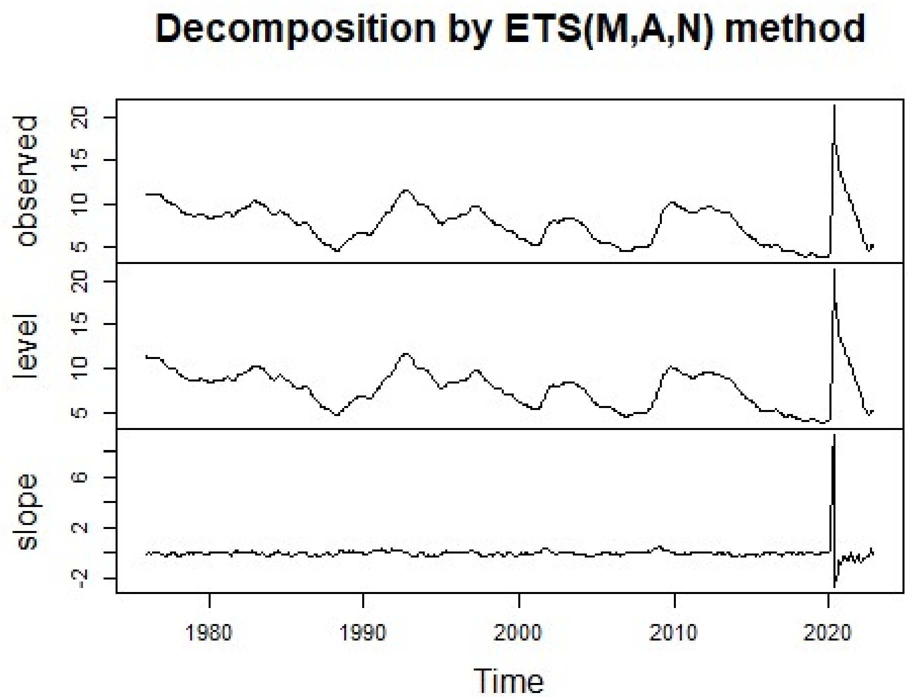


Insights:

The decomposition plot provides a breakdown of the time series data into its components: Data, Seasonal, Trend, and Remainder.

- The data exhibits a fluctuating pattern with periodic peaks, suggesting that it follows a seasonal pattern with significant variability over time. There's a noticeable spike towards the end, which could be an anomaly or special event.
- The seasonal component shows a clear, repeating pattern with a high degree of periodicity, indicating that the data has regular seasonal fluctuations. The consistent up-and-down pattern reflects strong seasonal effects that repeat over time.
- The trend component indicates an overall increasing or decreasing tendency over time, with some irregularity. The trend isn't perfectly linear, but it suggests a long-term direction (increasing or decreasing) while capturing underlying shifts in the data over the years. There is also a noticeable change or shift in the trend around the early 2000s.
- The remainder component represents the residuals, which ideally should show randomness if the decomposition captures all underlying patterns.
- In this plot, there are some spikes and outliers, especially in the later years, which suggests that there may be additional unexplained or irregular patterns in the data that the model hasn't captured. This could imply noise or unmodeled factors affecting the series.

5. ***Exponential Smoothing Methods***



ets_forecast

```
## ETS(M,A,N)
##
## Call:
## ets(y = timeSeriesData)
##
##   Smoothing parameters:
##     alpha = 0.8979
##     beta  = 0.8979
##
##   Initial states:
##     l = 11.2626
##     b = -0.1155
##
##   sigma:  0.0758
##
##      AIC     AICc      BIC
## 2931.286 2931.393 2952.961
```

Accuracy:

```
##                       ME      RMSE       MAE        MPE     MAPE       MASE
## Training set -5.77345e-07 0.6731741 0.1116492 0.07488355 1.202059 0.08313824
```

```
##                 ACF1
## Training set -0.08833626
```

Accuracy:

- ***ME: -0.0000005773*** *indicating almost no bias.*
- ***RMSE: 0.6732*** *small average error magnitude.*
- ***MAE: 0.1116*** *low overall error.*
- ***MPE: 0.0749%*** *very small bias in percentage terms.*
- ***MAPE: 1.2021%*** *high accuracy with minimal percentage errors.*
- ***ACF1: -0.0883*** *indicating no significant autocorrelation in residuals.*

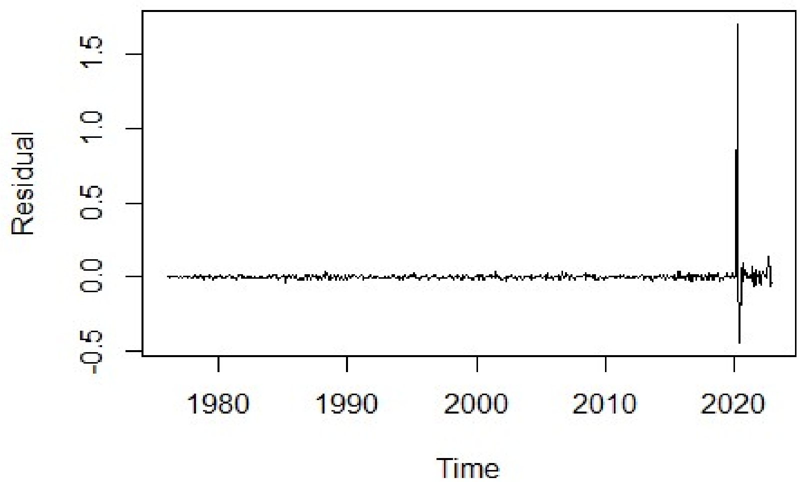
Model Overview

- ***ETS(M,A,N)****: Exponential Smoothing Model with Multiplicative Errors, Additive Trend, and No Seasonalit y.*
- ***Smoothing Parameters****:*
  - *Alpha (level):* ***0.8979***
  - *Beta (trend):* ***0.8979***
- ***Initial States****:*
  - *Level (l):* ***11.2626***
  - *Trend (b):* ***-0.1155***
- ***Model Metrics****:*
  - ***AIC****: 2931.286*
  - ***AICc****: 2931.393*
  - ***BIC****: 2952.961*

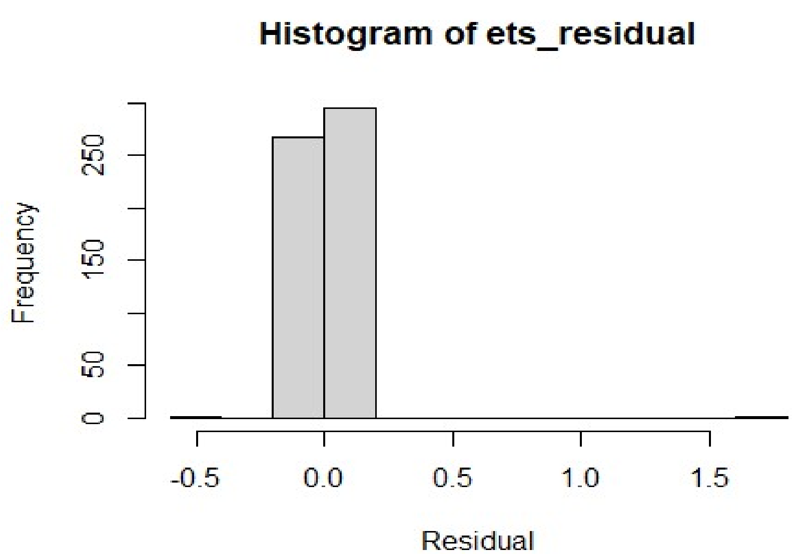
Insights:

The ETS(M,A,N) model provides highly accurate forecasts with minimal bias and errors. The RMSE and MAPE val ues indicate excellent prediction performance, and the residuals show no significant autocorrelation, confirming the model's suitability for the data.

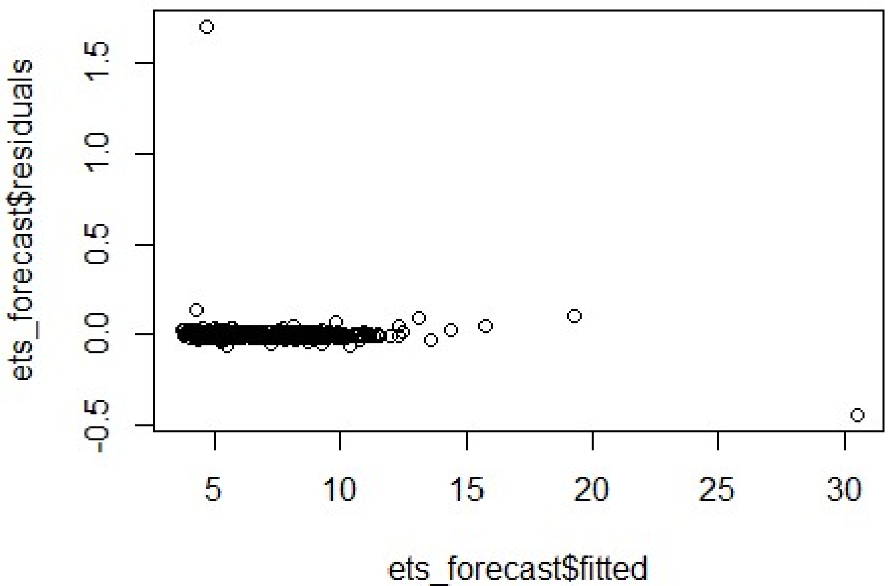
Residual Analysis:

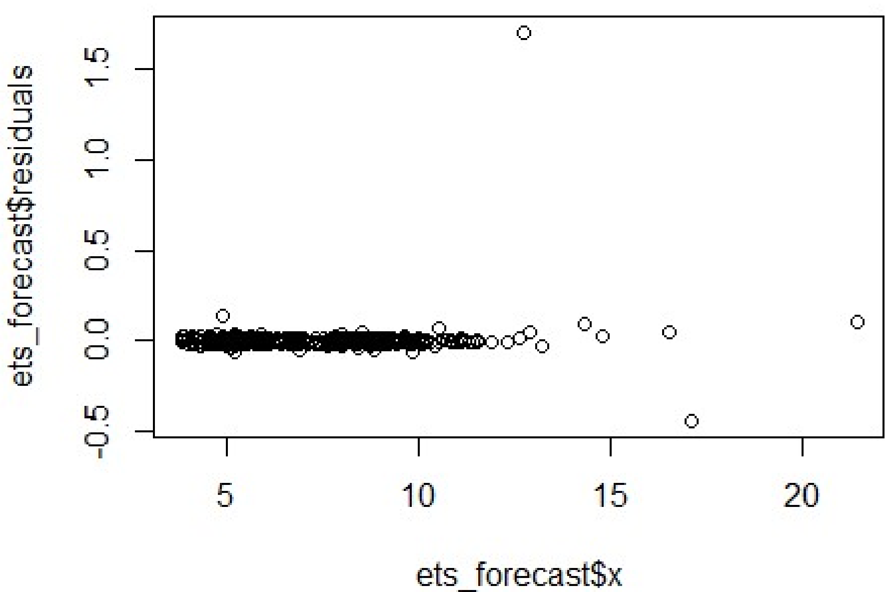


The residual plot shows a non-random pattern, particularly towards the end of the time series. This indicates that the model is not capturing the underlying pattern in the data, leading to systematic underestimation or overestimation.



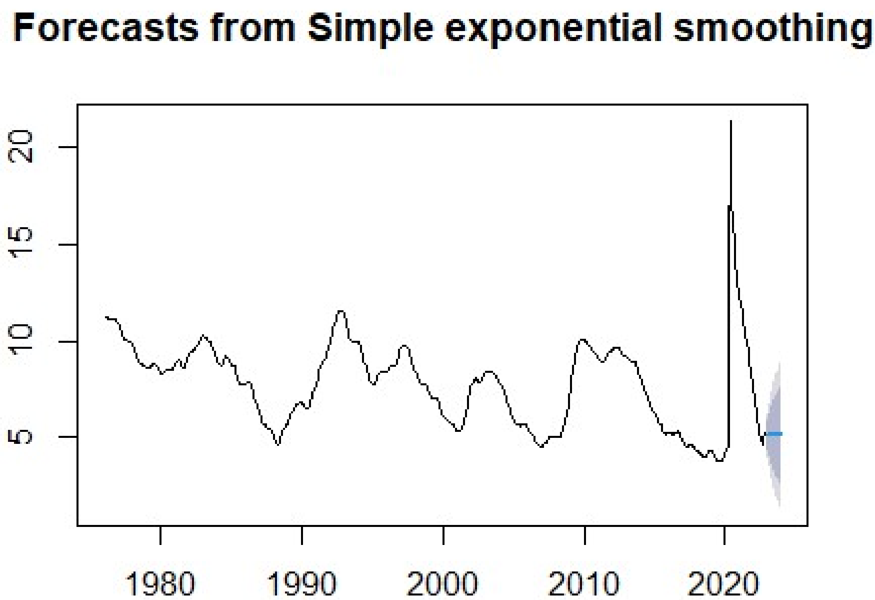
The residuals appear to be slightly right-skewed, indicating that the model may overestimate the values more freque ntly than it underestimates them. This suggests potential model misspecification or the presence of outliers in the dat a.





Both the plot suggests that the residuals have a non-constant variance, with larger residuals associated with larger fit ted values. This indicates heteroscedasticity.

6. ***Simple exponential smoothing***




```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2023        5.10001 4.373967 5.826053 3.989622 6.210398
## Feb 2023        5.10001 4.073281 6.126739 3.529763 6.670257
## Mar 2023        5.10001 3.842550 6.357470 3.176891 7.023129
## Apr 2023        5.10001 3.648032 6.551988 2.879401 7.320619
## May 2023        5.10001 3.476658 6.723362 2.617307 7.582713
## Jun 2023        5.10001 3.321722 6.878298 2.380354 7.819666
## Jul 2023        5.10001 3.179244 7.020776 2.162453 8.037567
## Aug 2023        5.10001 3.046629 7.153391 1.959635 8.240385
## Sep 2023        5.10001 2.922073 7.277947 1.769144 8.430876
## Oct 2023        5.10001 2.804266 7.395754 1.588972 8.611048
## Nov 2023        5.10001 2.692215 7.507805 1.417606 8.782414
## Dec 2023        5.10001 2.585153 7.614867 1.253867 8.946153
```

Accuracy:

```
##                      ME      RMSE       MAE        MPE     MAPE      MASE
## Training set -0.01081135 0.5655293 0.1569322 -0.2596054 1.819525 0.1168576
##                   ACF1
## Training set 0.2871729
```

Accuracy:

1. **ME: -0.0108** (very close to zero), showing minimal bias.
2. **RMSE: 0.5655**, indicating strong accuracy.
3. **MAE: 0.157**, reflecting low average deviations.

4. **MPE: -0.26%**, suggesting minor under-prediction.
5. **MAPE: 1.82%**, showing high prediction accuracy relative to observed values.
6. **ACF1: 0.287**, indicating moderate dependency in residuals.

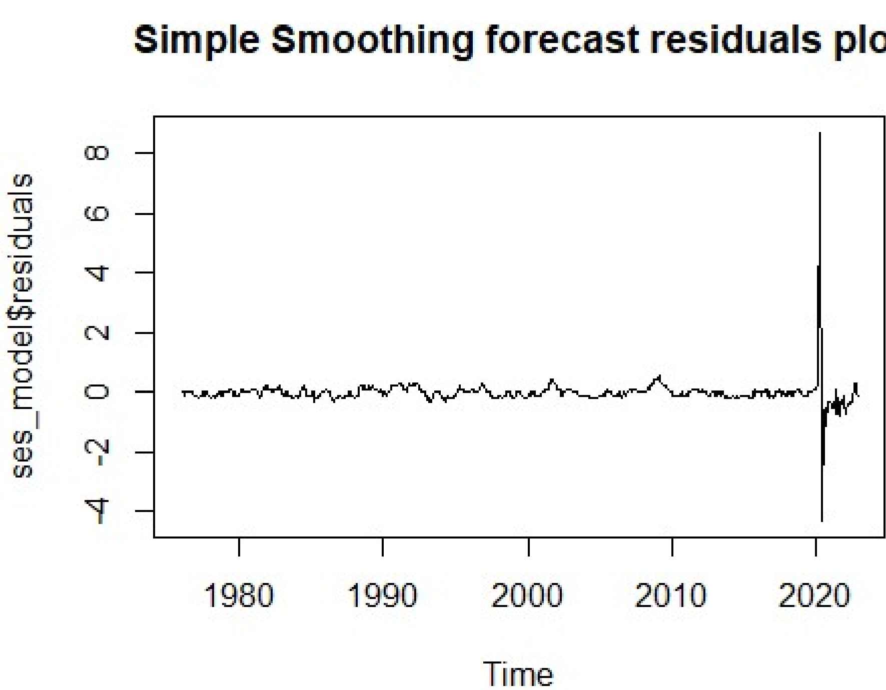
Prediction:

1. **Point Forecast:** Constant at **5.10001** across all months.
2. **Prediction Intervals:**
    - **80% Interval** Widens over time from **[4.37, 5.83] (Jan 2023)** to **[2.80, 7.40] (Oct 2023)**.
    - **95% Interval** Expands more significantly from **[3.99, 6.21] (Jan 2023)** to **[1.59, 8.61] (Oct 2023)**.
    - Prediction uncertainty grows for longer time horizons.

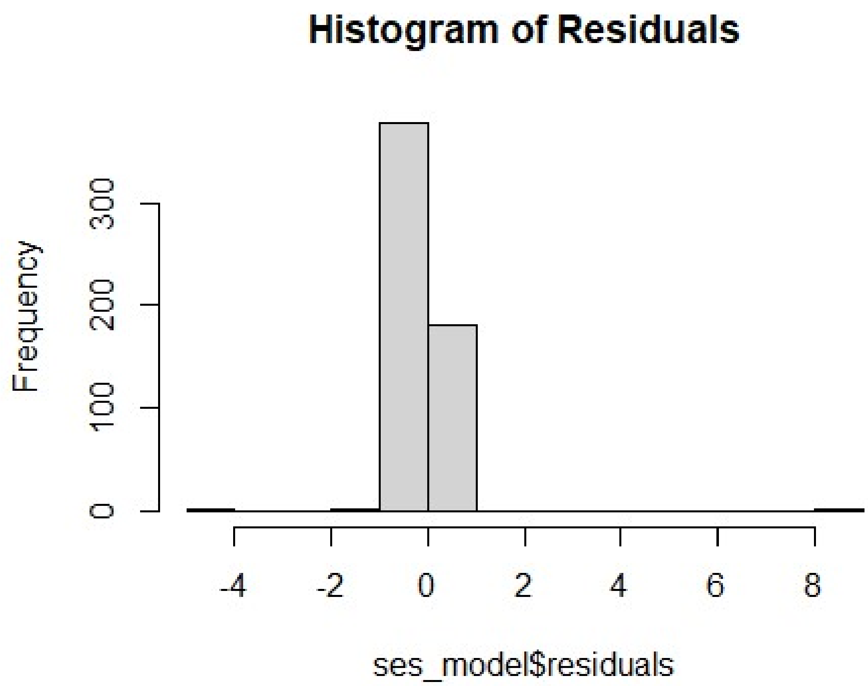
Insights:

The forecasting model delivers high accuracy with minimal bias and small errors. However, the increasing prediction intervals highlight growing uncertainty over time, and residual autocorrelation suggests some improvements may still be possible in capturing dependencies.

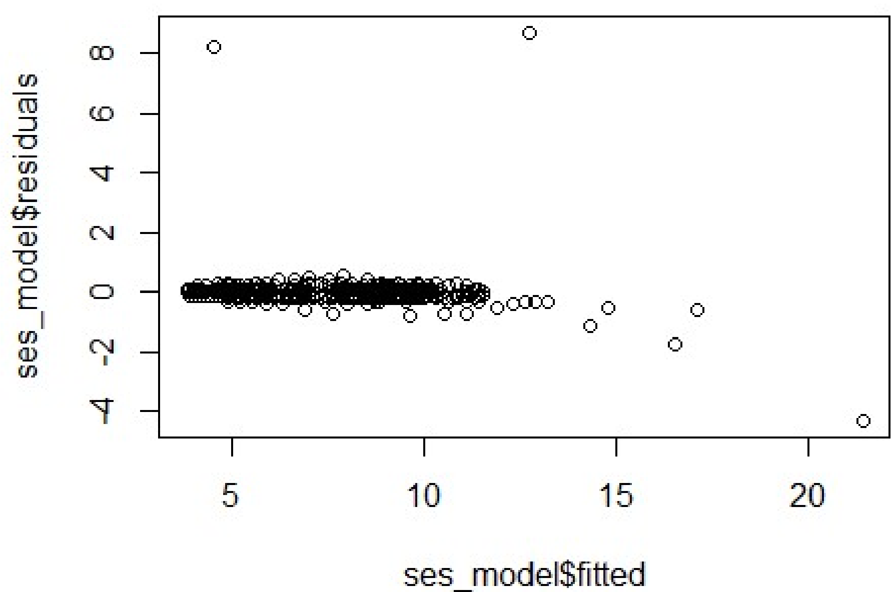
Residuals:

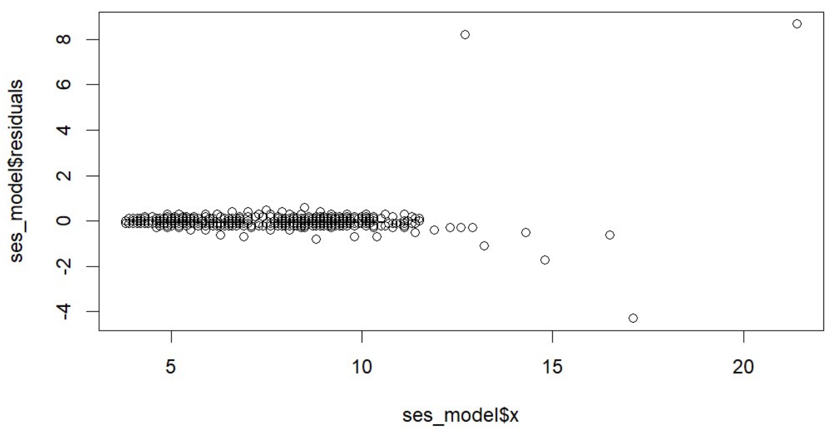


The residual plot shows a non-random pattern, particularly towards the end of the time series. This indicates that the simple smoothing model is not capturing the underlying pattern in the data, leading to systematic underestimation or overestimation.



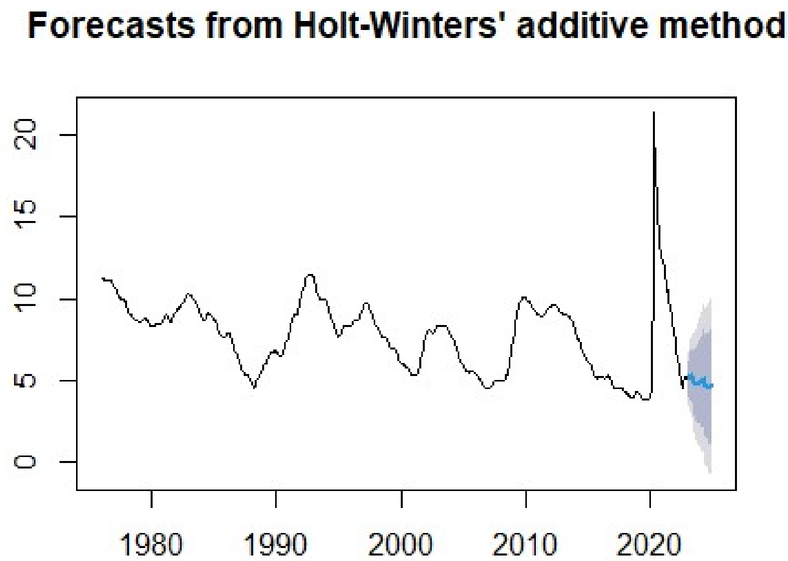
The residuals appear to be slightly right-skewed, indicating that the simple smoothing model may overestimate the v alues more frequently than it underestimates them. This suggests potential model misspecification or the presence of outliers in the data.





Both the plot suggests that the residuals have a non-constant variance, with larger residuals associated with larger fit
ted values. This indicates heteroscedasticity.

7. ***HoltWinters***



##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Jan 2023	5.167662	4.412097	5.923227	4.01212553	6.323199
## Feb 2023	5.142006	4.093601	6.190411	3.53860945	6.745403
## Mar 2023	5.105368	3.829618	6.381118	3.15427656	7.056459
## Apr 2023	5.356503	3.888159	6.824847	3.11086505	7.602141
## May 2023	5.354360	3.715873	6.992847	2.84851044	7.860210
## Jun 2023	4.891182	3.098598	6.683766	2.14966157	7.632702
## Jul 2023	4.878536	2.944064	6.813009	1.92001585	7.837057
## Aug 2023	4.785936	2.719267	6.852605	1.62523807	7.946634
## Sep 2023	4.754629	2.563699	6.945558	1.40389155	8.105366
## Oct 2023	4.740390	2.431856	7.048925	1.20979196	8.270989
## Nov 2023	4.796287	2.375832	7.216741	1.09452087	8.498053
## Dec 2023	4.910405	2.382960	7.437851	1.04501068	8.775800
## Jan 2024	4.978068	2.339841	7.616294	0.94324816	9.012887
## Feb 2024	4.952411	2.215654	7.689169	0.76690198	9.137921
## Mar 2024	4.915773	2.083891	7.747655	0.58478316	9.246762
## Apr 2024	5.166908	2.242976	8.090840	0.69513991	9.638676
## May 2024	5.164765	2.151575	8.177955	0.55648885	9.773041
## Jun 2024	4.701587	1.601690	7.801483	-0.03929578	9.442470
## Jul 2024	4.688941	1.504681	7.873202	-0.18096481	9.558847
## Aug 2024	4.596341	1.329878	7.862804	-0.39928322	9.591965
## Sep 2024	4.565034	1.218370	7.911698	-0.55324707	9.683314
## Oct 2024	4.550795	1.125792	7.975799	-0.68729512	9.788886
## Nov 2024	4.606692	1.105085	8.108298	-0.74855302	9.961937
## Dec 2024	4.720810	1.144226	8.297395	-0.74910388	10.190724

Accuracy:

```
##                     ME      RMSE       MAE        MPE     MAPE      MASE
## Training set 0.005347043 0.5811476 0.1874316 -0.03797169 2.228681 0.1395687
##                  ACF1
## Training set 0.3138823
```

Accuracy:

- **ME:** 0.0053 no significant bias.
- **RMSE:** 0.5811 moderate error.
- **MAE:** 0.1874 good precision.
- **MAPE:** 2.23% excellent accuracy.
- **MASE:** 0.1396 better than naive forecasting.
- **ACF1:** 0.3139 low residual correlation, model fits well.

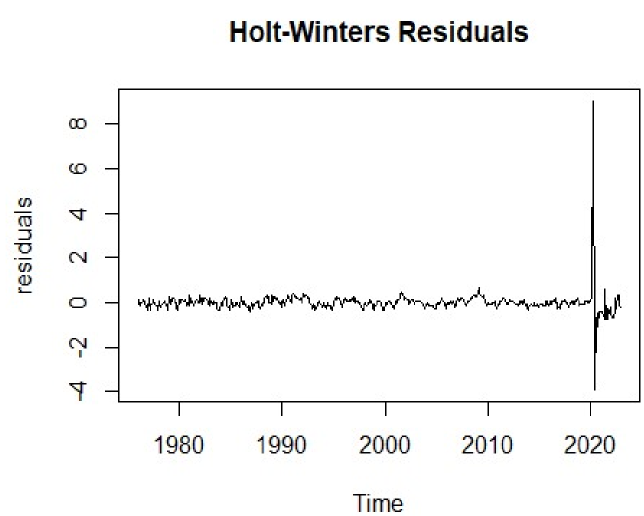
Predictions:

- Point forecasts range from **4.74** to **5.36** over the period.
- Confidence intervals (80% and 95%) show a varying range, with wider uncertainty as we move into later months.

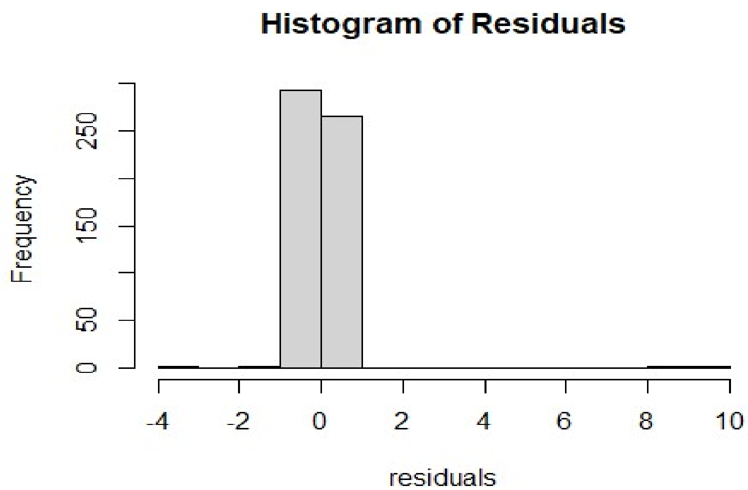
Insights:

The model provides accurate and reliable forecasts with minimal bias and good precision. Prediction uncertainty increases in later months.

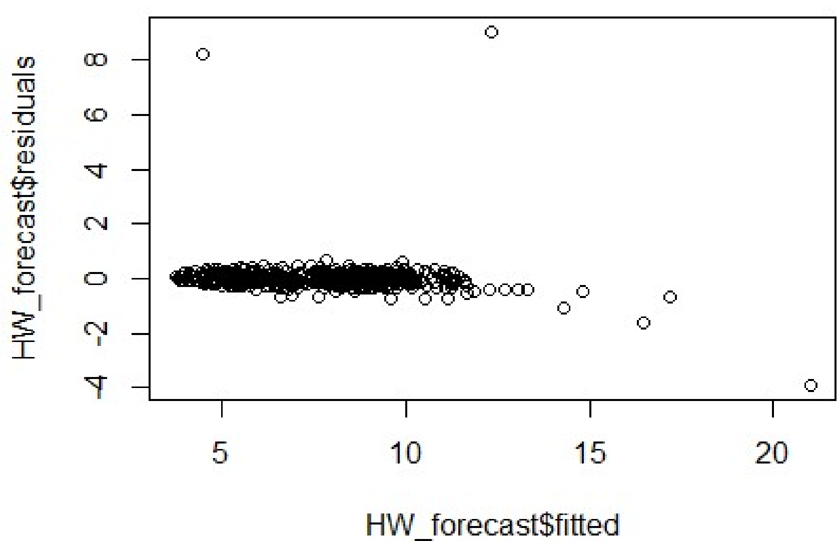
Residuals:



The residual plot shows a non-random pattern, particularly towards the end of the time series. This indicates that the simple smoothing model is not capturing the underlying pattern in the data, leading to systematic underestimation or overestimation.



The residuals appear to be slightly right-skewed, indicating that the simple smoothing model may overestimate the v alues more frequently than it underestimates them. This suggests potential model misspecification or the presence of outliers in the data.



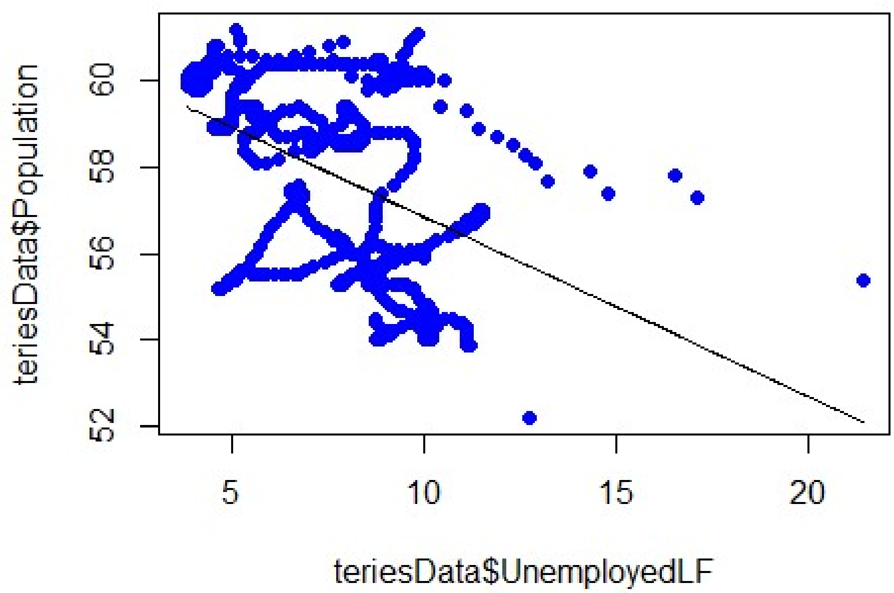
The plot suggests that the residuals have a non-constant variance, with larger residuals associated with larger fitted v alues. This indicates heteroscedasticity.

Accuracy comparison:

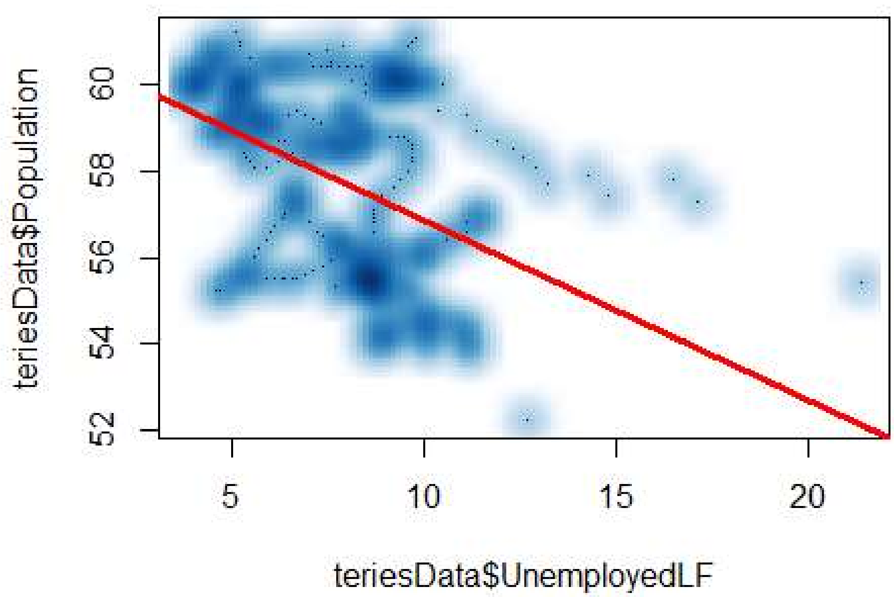
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Mean Forecast	-2.984247e-18	2.222627	1.803226	-9.125461	26.7602	1.34275	0.9642393
Naïve Forecast	-0.01083481	0.566015	0.1571936	-0.260081	1.822557	0.1170523	0.2870979
SES Model	-0.01081135	0.5655293	0.1569322	-0.2596054	1.819525	0.1168576	0.2871729
ETS Forecast	-5.77345e-07	0.6731741	0.1116492	0.07488355	1.202059	0.08313824	-0.08833626
HW Forecast	0.005347043	0.5811476	0.1874316	-0.03797169	2.228681	0.1395687	0.3138823

The best forecasting model according to MAPE (Mean Absolute Percentage Error) is the ETS Forecast with a MAPE value of 1.202059.

Regression:



teriesData$Population
60
58
56
54
52
5
10
15
20
teriesData$UnemployedLF



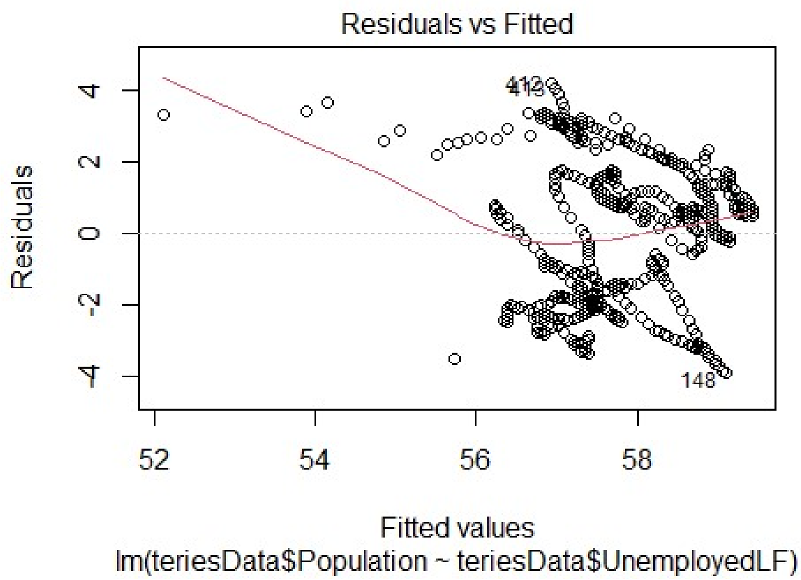
teriesData$Population
60
58
56
54
52
5
10
15
20
teriesData$UnemployedLF

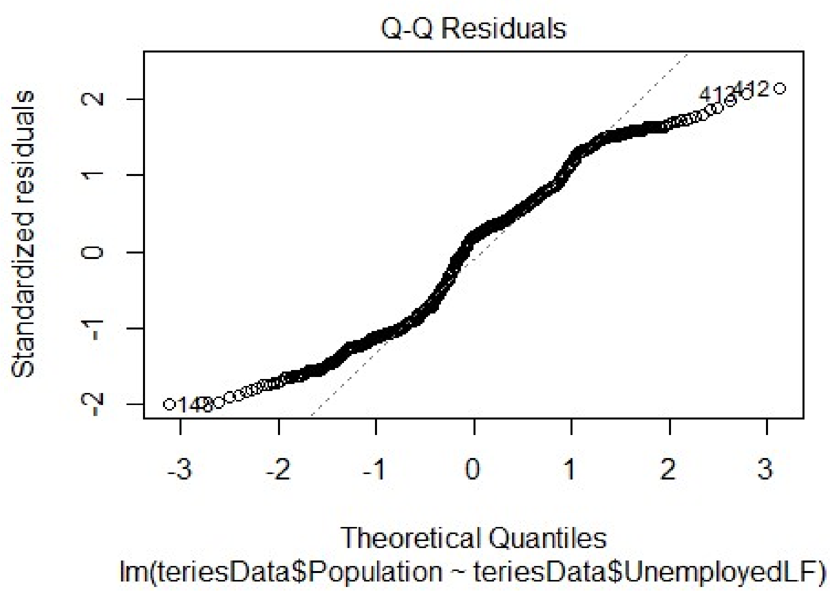
Insights:

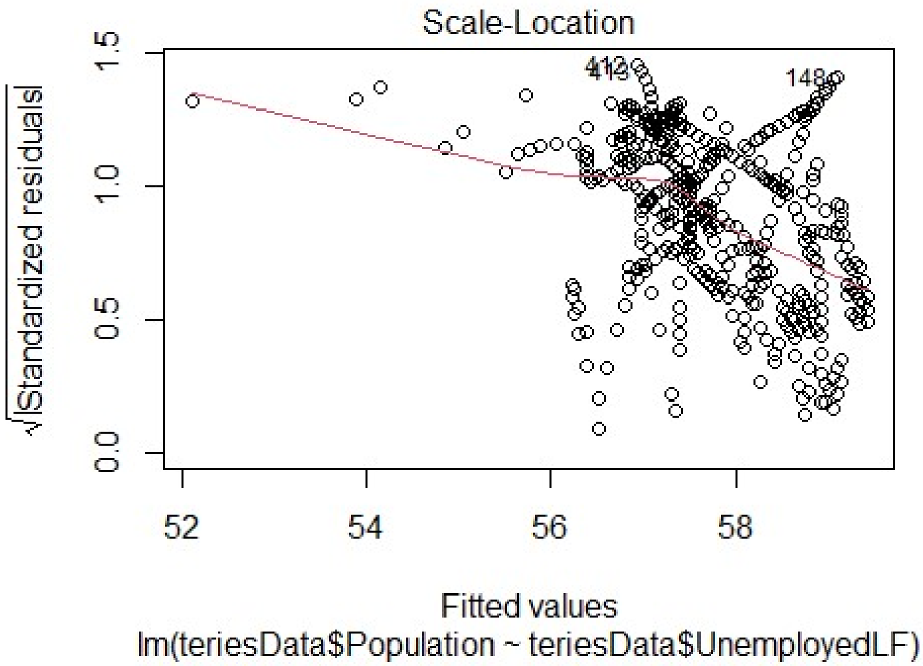
The regression model predicts Population based on Unemployed Labor Force (UnemployedLF).

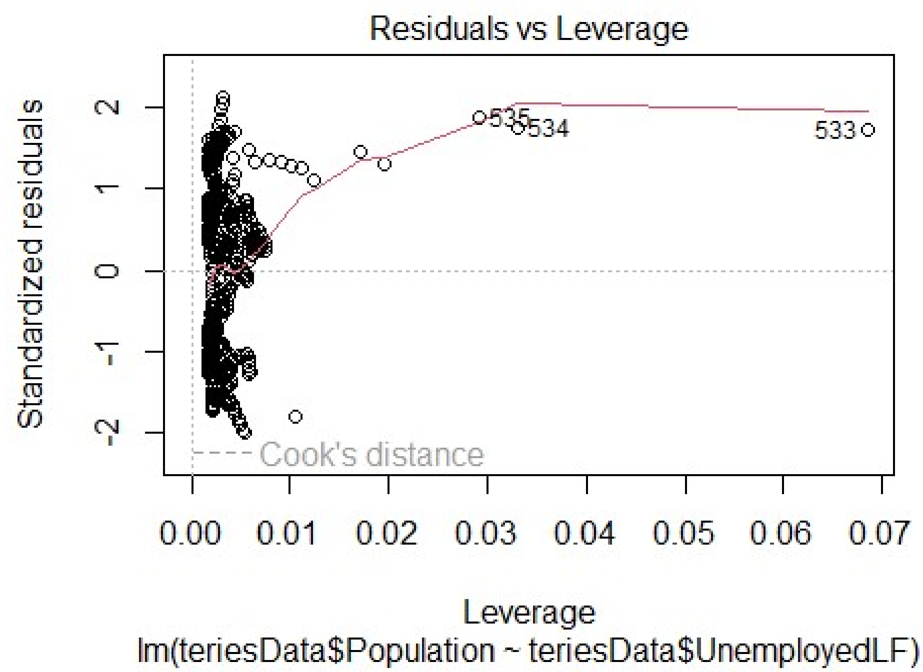
For every unit increase in the unemployed labor force, the population decreases by 0.4157 million (or approximately 416,000 people). This negative relationship suggests that higher unemployment correlates with a decline in populati on, potentially due to out-migration or other factors.

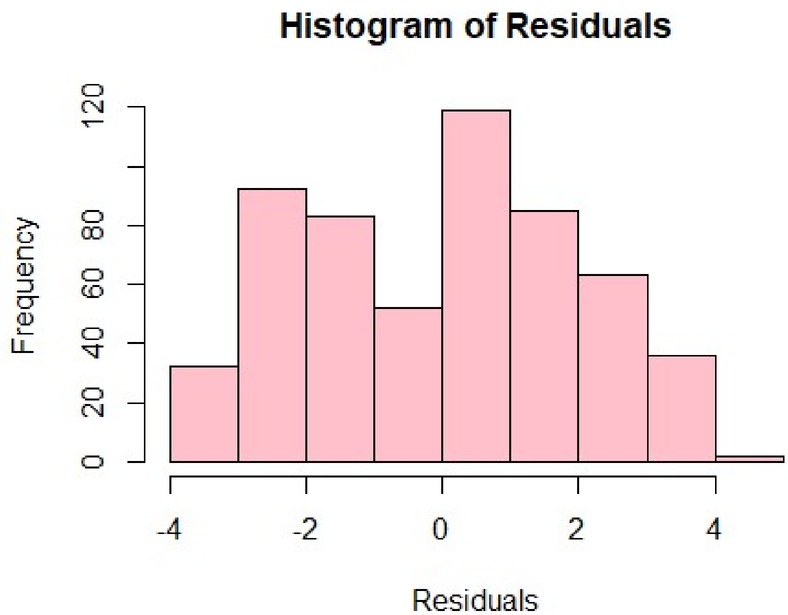
Residual analysis*:*

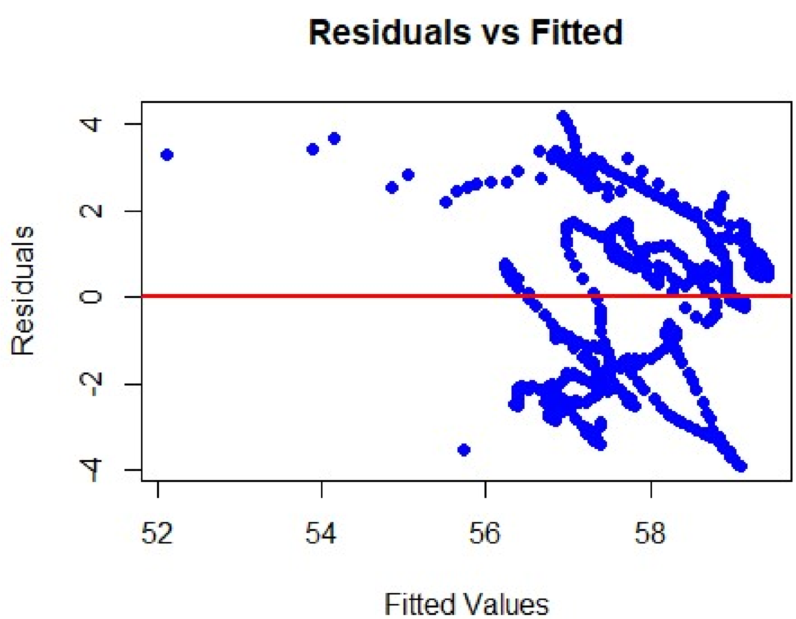





Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
52
54
56
58
Fitted values
lm(teriesData$Population ~ teriesData$UnemployedLF)
412
413
148


Residuals vs Leverage
Standardized residuals
-2
-1
0
1
2
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Leverage
lm(teriesData$Population ~ teriesData$UnemployedLF)
535
534
533
Cook's distance


Histogram of Residuals
Frequency
0
20
40
60
80
120
-4
-2
0
2
4
Residuals


Residuals vs Fitted
Residuals
4
2
0
-2
-4
52
54
56
58
Fitted Values

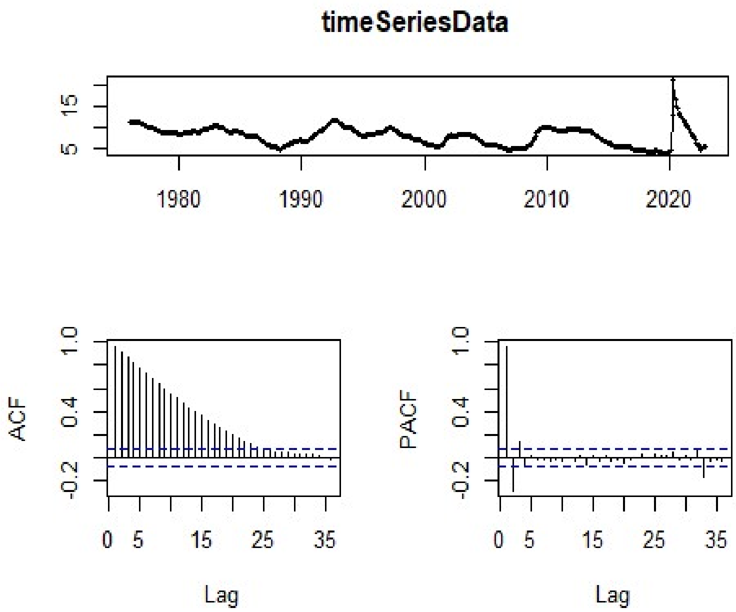
1. Histogram of Residuals:

- Residuals are roughly symmetric and centered around zero.
- No extreme bias, but slight deviations from normality are observed.

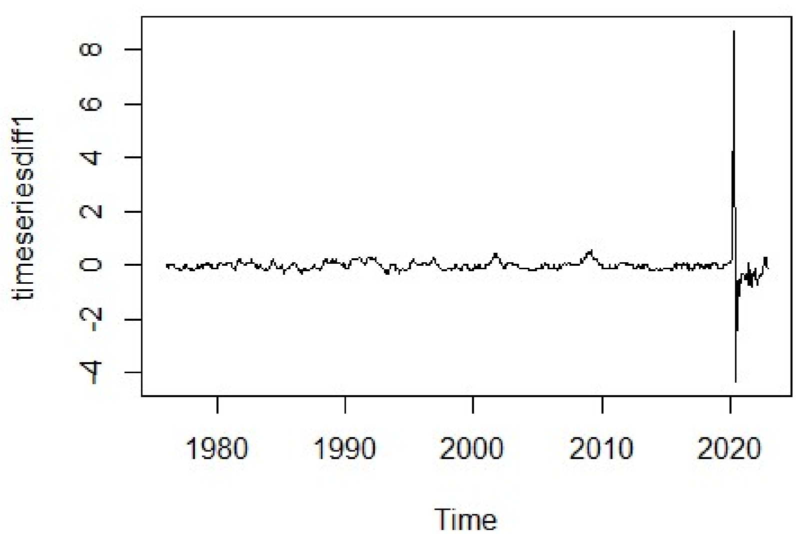
2. Residuals vs. Fitted Plot:

- Residuals show no strong patterns, indicating good model fit.
- Slight clustering may suggest minor issues like autocorrelation or model misspecification

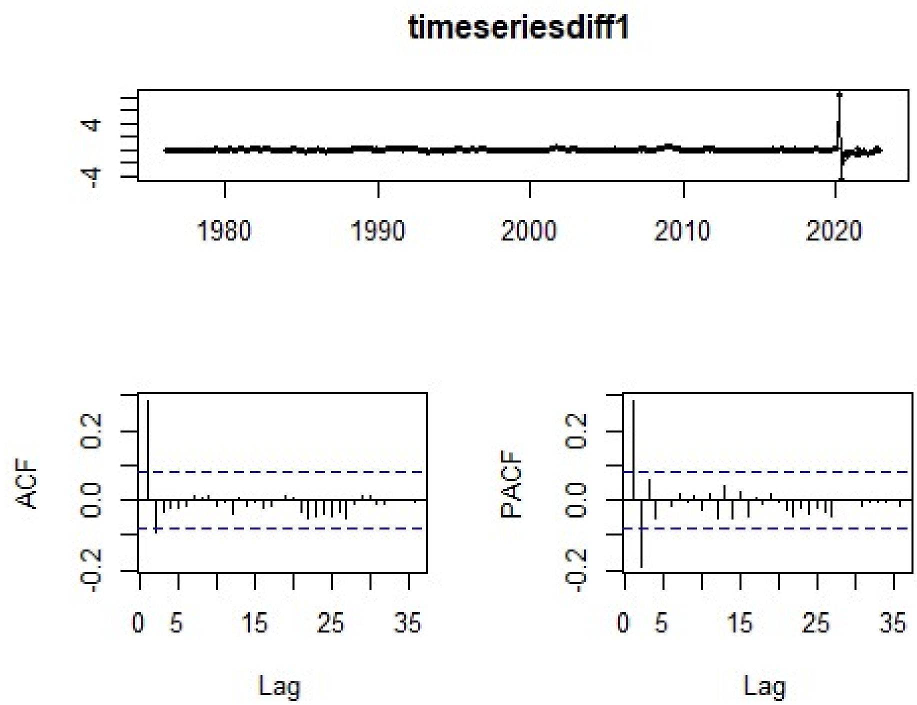
ARIMA



timeSeriesData
15
5
1980
1990
2000
2010
2020
ACF
1.0
0.4
-0.2
0
5
15
25
35
Lag
PACF
1.0
0.4
-0.2
0
5
15
25
35
Lag



timeseriesdiff1
8
6
4
2
0
-2
-4
1980
1990
2000
2010
2020
Time




```
##
## Fitting models using approximations to speed things up...
##
## ARIMA(0,1,0)                           : 960.0246
## ARIMA(0,1,0)           with drift      : 961.8325
## ARIMA(0,1,0)(0,0,1)[12]                : 961.1424
## ARIMA(0,1,0)(0,0,1)[12] with drift     : 962.9491
## ARIMA(0,1,0)(0,0,2)[12]                : 962.0794
## ARIMA(0,1,0)(0,0,2)[12] with drift     : 963.899
## ARIMA(0,1,0)(1,0,0)[12]                : 973.2252
## ARIMA(0,1,0)(1,0,0)[12] with drift     : 975.0543
## ARIMA(0,1,0)(1,0,1)[12]                : Inf
## ARIMA(0,1,0)(1,0,1)[12] with drift     : Inf
## ARIMA(0,1,0)(1,0,2)[12]                : Inf
## ARIMA(0,1,0)(1,0,2)[12] with drift     : Inf
## ARIMA(0,1,0)(2,0,0)[12]                : 985.9717
## ARIMA(0,1,0)(2,0,0)[12] with drift     : 987.8799
## ARIMA(0,1,0)(2,0,1)[12]                : 986.9373
## ARIMA(0,1,0)(2,0,1)[12] with drift     : 988.9378
## ARIMA(0,1,0)(2,0,2)[12]                : Inf
## ARIMA(0,1,0)(2,0,2)[12] with drift     : Inf
## ARIMA(0,1,1)                           : 894.0672
## ARIMA(0,1,1)           with drift      : 895.9671
## ARIMA(0,1,1)(0,0,1)[12]                : 894.1153
## ARIMA(0,1,1)(0,0,1)[12] with drift     : 896.0161
## ARIMA(0,1,1)(0,0,2)[12]                : 895.8706
## ARIMA(0,1,1)(0,0,2)[12] with drift     : 897.7801
## ARIMA(0,1,1)(1,0,0)[12]                : 906.2337
```

```
## ARIMA(0,1,1)(1,0,0)[12] with drift : 908.1463
## ARIMA(0,1,1)(1,0,1)[12] : 906.8622
## ARIMA(0,1,1)(1,0,1)[12] with drift : 908.839
## ARIMA(0,1,1)(1,0,2)[12] : 908.8894
## ARIMA(0,1,1)(1,0,2)[12] with drift : 910.8758
## ARIMA(0,1,1)(2,0,0)[12] : 919.9774
## ARIMA(0,1,1)(2,0,0)[12] with drift : 921.9362
## ARIMA(0,1,1)(2,0,1)[12] : 921.4894
## ARIMA(0,1,1)(2,0,1)[12] with drift : 923.511
## ARIMA(0,1,1)(2,0,2)[12] : 923.0491
## ARIMA(0,1,1)(2,0,2)[12] with drift : 925.0799
## ARIMA(0,1,2) : 892.0526
## ARIMA(0,1,2) with drift : 893.9338
## ARIMA(0,1,2)(0,0,1)[12] : 891.7343
## ARIMA(0,1,2)(0,0,1)[12] with drift : 893.6132
## ARIMA(0,1,2)(0,0,2)[12] : 893.3616
## ARIMA(0,1,2)(0,0,2)[12] with drift : 895.2489
## ARIMA(0,1,2)(1,0,0)[12] : 903.8808
## ARIMA(0,1,2)(1,0,0)[12] with drift : 905.7752
## ARIMA(0,1,2)(1,0,1)[12] : 903.9959
## ARIMA(0,1,2)(1,0,1)[12] with drift : 905.9593
## ARIMA(0,1,2)(1,0,2)[12] : 906.0365
## ARIMA(0,1,2)(1,0,2)[12] with drift : 908.0089
## ARIMA(0,1,2)(2,0,0)[12] : 917.4214
## ARIMA(0,1,2)(2,0,0)[12] with drift : 919.3711
## ARIMA(0,1,2)(2,0,1)[12] : 918.643
## ARIMA(0,1,2)(2,0,1)[12] with drift : 920.6694
## ARIMA(0,1,3) : 893.7933
## ARIMA(0,1,3) with drift : 895.6744
## ARIMA(0,1,3)(0,0,1)[12] : 893.4985
## ARIMA(0,1,3)(0,0,1)[12] with drift : 895.377
## ARIMA(0,1,3)(0,0,2)[12] : 895.0655
## ARIMA(0,1,3)(0,0,2)[12] with drift : 896.9519
## ARIMA(0,1,3)(1,0,0)[12] : 905.6472
## ARIMA(0,1,3)(1,0,0)[12] with drift : 907.5427
## ARIMA(0,1,3)(1,0,1)[12] : 905.5836
## ARIMA(0,1,3)(1,0,1)[12] with drift : 907.5485
## ARIMA(0,1,3)(2,0,0)[12] : 919.1067
## ARIMA(0,1,3)(2,0,0)[12] with drift : 921.0594
## ARIMA(0,1,4) : 895.5312
## ARIMA(0,1,4) with drift : 897.4115
## ARIMA(0,1,4)(0,0,1)[12] : 895.1975
## ARIMA(0,1,4)(0,0,1)[12] with drift : 897.074
## ARIMA(0,1,4)(1,0,0)[12] : 907.3473
## ARIMA(0,1,4)(1,0,0)[12] with drift : 909.2422
## ARIMA(0,1,5) : 897.5711
## ARIMA(0,1,5) with drift : 899.4595
## ARIMA(1,1,0) : 914.5168
## ARIMA(1,1,0) with drift : 916.4224
## ARIMA(1,1,0)(0,0,1)[12] : 915.3602
## ARIMA(1,1,0)(0,0,1)[12] with drift : 917.2681
## ARIMA(1,1,0)(0,0,2)[12] : 917.1153
## ARIMA(1,1,0)(0,0,2)[12] with drift : 919.0319
## ARIMA(1,1,0)(1,0,0)[12] : 927.4415
## ARIMA(1,1,0)(1,0,0)[12] with drift : 929.3666
## ARIMA(1,1,0)(1,0,1)[12] : 928.2795
```

```
## ARIMA(1,1,0)(1,0,1)[12] with drift         : 930.2787
## ARIMA(1,1,0)(1,0,2)[12]                    : 930.3119
## ARIMA(1,1,0)(1,0,2)[12] with drift         : 932.317
## ARIMA(1,1,0)(2,0,0)[12]                    : 941.1191
## ARIMA(1,1,0)(2,0,0)[12] with drift         : 943.0937
## ARIMA(1,1,0)(2,0,1)[12]                    : 942.9254
## ARIMA(1,1,0)(2,0,1)[12] with drift         : 944.9454
## ARIMA(1,1,0)(2,0,2)[12]                    : 944.6535
## ARIMA(1,1,0)(2,0,2)[12] with drift         : 946.6872
## ARIMA(1,1,1)                               : 893.5276
## ARIMA(1,1,1)           with drift          : 895.4193
## ARIMA(1,1,1)(0,0,1)[12]                    : 893.2422
## ARIMA(1,1,1)(0,0,1)[12] with drift         : 895.1329
## ARIMA(1,1,1)(0,0,2)[12]                    : 894.9324
## ARIMA(1,1,1)(0,0,2)[12] with drift         : 896.832
## ARIMA(1,1,1)(1,0,0)[12]                    : 905.4126
## ARIMA(1,1,1)(1,0,0)[12] with drift         : 907.3143
## ARIMA(1,1,1)(1,0,1)[12]                    : 905.7752
## ARIMA(1,1,1)(1,0,1)[12] with drift         : 907.7515
## ARIMA(1,1,1)(1,0,2)[12]                    : 907.8137
## ARIMA(1,1,1)(1,0,2)[12] with drift         : 909.7994
## ARIMA(1,1,1)(2,0,0)[12]                    : 918.9024
## ARIMA(1,1,1)(2,0,0)[12] with drift         : 920.8645
## ARIMA(1,1,1)(2,0,1)[12]                    : 920.3272
## ARIMA(1,1,1)(2,0,1)[12] with drift         : 922.3561
## ARIMA(1,1,2)                               : 894.5802
## ARIMA(1,1,2)           with drift          : 896.4525
## ARIMA(1,1,2)(0,0,1)[12]                    : 894.2767
## ARIMA(1,1,2)(0,0,1)[12] with drift         : 896.1463
## ARIMA(1,1,2)(0,0,2)[12]                    : 895.803
## ARIMA(1,1,2)(0,0,2)[12] with drift         : 897.6799
## ARIMA(1,1,2)(1,0,0)[12]                    : 906.3757
## ARIMA(1,1,2)(1,0,0)[12] with drift         : 908.2745
## ARIMA(1,1,2)(1,0,1)[12]                    : 906.0839
## ARIMA(1,1,2)(1,0,1)[12] with drift         : 908.0717
## ARIMA(1,1,2)(2,0,0)[12]                    : 919.6214
## ARIMA(1,1,2)(2,0,0)[12] with drift         : 921.5871
## ARIMA(1,1,3)                               : Inf
## ARIMA(1,1,3)           with drift          : Inf
## ARIMA(1,1,3)(0,0,1)[12]                    : 896.1911
## ARIMA(1,1,3)(0,0,1)[12] with drift         : Inf
## ARIMA(1,1,3)(1,0,0)[12]                    : Inf
## ARIMA(1,1,3)(1,0,0)[12] with drift         : Inf
## ARIMA(1,1,4)                               : Inf
## ARIMA(1,1,4)           with drift          : Inf
## ARIMA(2,1,0)                               : 897.208
## ARIMA(2,1,0)           with drift          : 899.0678
## ARIMA(2,1,0)(0,0,1)[12]                    : 897.3147
## ARIMA(2,1,0)(0,0,1)[12] with drift         : 899.1724
## ARIMA(2,1,0)(0,0,2)[12]                    : 898.6726
## ARIMA(2,1,0)(0,0,2)[12] with drift         : 900.5397
## ARIMA(2,1,0)(1,0,0)[12]                    : 909.3592
## ARIMA(2,1,0)(1,0,0)[12] with drift         : 911.2506
## ARIMA(2,1,0)(1,0,1)[12]                    : 909.2401
## ARIMA(2,1,0)(1,0,1)[12] with drift         : 911.2296
## ARIMA(2,1,0)(1,0,2)[12]                    : 911.2365
```

```
##  ARIMA(2,1,0)(1,0,2)[12] with drift         : 913.2263
##  ARIMA(2,1,0)(2,0,0)[12]                    : 922.6056
##  ARIMA(2,1,0)(2,0,0)[12] with drift         : 924.5583
##  ARIMA(2,1,0)(2,0,1)[12]                    : 923.7297
##  ARIMA(2,1,0)(2,0,1)[12] with drift         : 925.7543
##  ARIMA(2,1,1)                               : 895.9943
##  ARIMA(2,1,1)            with drift         : 897.8775
##  ARIMA(2,1,1)(0,0,1)[12]                    : 895.7265
##  ARIMA(2,1,1)(0,0,1)[12] with drift         : 897.6086
##  ARIMA(2,1,1)(0,0,2)[12]                    : 897.3212
##  ARIMA(2,1,1)(0,0,2)[12] with drift         : 899.212
##  ARIMA(2,1,1)(1,0,0)[12]                    : 907.888
##  ARIMA(2,1,1)(1,0,0)[12] with drift         : 909.7948
##  ARIMA(2,1,1)(1,0,1)[12]                    : 908.1549
##  ARIMA(2,1,1)(1,0,1)[12] with drift         : 910.1483
##  ARIMA(2,1,1)(2,0,0)[12]                    : 921.3316
##  ARIMA(2,1,1)(2,0,0)[12] with drift         : 923.2967
##  ARIMA(2,1,2)                               : Inf
##  ARIMA(2,1,2)            with drift         : Inf
##  ARIMA(2,1,2)(0,0,1)[12]                    : 897.1767
##  ARIMA(2,1,2)(0,0,1)[12] with drift         : 899.0489
##  ARIMA(2,1,2)(1,0,0)[12]                    : Inf
##  ARIMA(2,1,2)(1,0,0)[12] with drift         : Inf
##  ARIMA(2,1,3)                               : Inf
##  ARIMA(2,1,3)            with drift         : Inf
##  ARIMA(3,1,0)                               : 898.1608
##  ARIMA(3,1,0)            with drift         : 900.0537
##  ARIMA(3,1,0)(0,0,1)[12]                    : 898.0867
##  ARIMA(3,1,0)(0,0,1)[12] with drift         : 899.9797
##  ARIMA(3,1,0)(0,0,2)[12]                    : 899.5978
##  ARIMA(3,1,0)(0,0,2)[12] with drift         : 901.4994
##  ARIMA(3,1,0)(1,0,0)[12]                    : 910.1209
##  ARIMA(3,1,0)(1,0,0)[12] with drift         : 912.0462
##  ARIMA(3,1,0)(1,0,1)[12]                    : 910.3674
##  ARIMA(3,1,0)(1,0,1)[12] with drift         : 912.3765
##  ARIMA(3,1,0)(2,0,0)[12]                    : 923.4915
##  ARIMA(3,1,0)(2,0,0)[12] with drift         : 925.4721
##  ARIMA(3,1,1)                               : 895.8086
##  ARIMA(3,1,1)            with drift         : 897.6591
##  ARIMA(3,1,1)(0,0,1)[12]                    : 896.6225
##  ARIMA(3,1,1)(0,0,1)[12] with drift         : 898.4834
##  ARIMA(3,1,1)(1,0,0)[12]                    : 908.7043
##  ARIMA(3,1,1)(1,0,0)[12] with drift         : 910.5918
##  ARIMA(3,1,2)                               : 897.6315
##  ARIMA(3,1,2)            with drift         : 899.456
##  ARIMA(4,1,0)                               : 899.739
##  ARIMA(4,1,0)            with drift         : 901.6195
##  ARIMA(4,1,0)(0,0,1)[12]                    : 899.4891
##  ARIMA(4,1,0)(0,0,1)[12] with drift         : 901.3679
##  ARIMA(4,1,0)(1,0,0)[12]                    : 911.5294
##  ARIMA(4,1,0)(1,0,0)[12] with drift         : 913.4491
##  ARIMA(4,1,1)                               : Inf
##  ARIMA(4,1,1)            with drift         : Inf
##  ARIMA(5,1,0)                               : 902.7872
##  ARIMA(5,1,0)            with drift         : 904.6767
##
```

```
##
##  Best model: ARIMA(0,1,2)(0,0,1)[12]
```

```
## Series: timeSeriesData
## ARIMA(0,1,2)(0,0,1)[12]
##
## Coefficients:
##          ma1      ma2     sma1
##       0.3587  -0.0919  -0.0655
## s.e.  0.0425   0.0438   0.0432
##
## sigma^2 = 0.2822:  log likelihood = -441.38
## AIC=890.76   AICc=890.84   BIC=908.1
```

The best ARIMA model for the time series data was identifies as: ARIMA(0,1,2)(0,0,1)[12].

Explanations:

Model Components are ARIMA(0,1,2) which states:

p=0: No autoregressive terms.

d=1: Data is differenced once to achieve stationarity.

q=2: Two moving average terms.

Seasonal Component are (0,0,1)[12] which states:

P=0: No seasonal autoregressive terms.

D=0: No seasonal differencing.

Q=1: One seasonal MA term.

[12]: Indicates seasonality occurs with a periodicity of 12.

Coefficients

The coefficients of the model are:

ma1=0.358

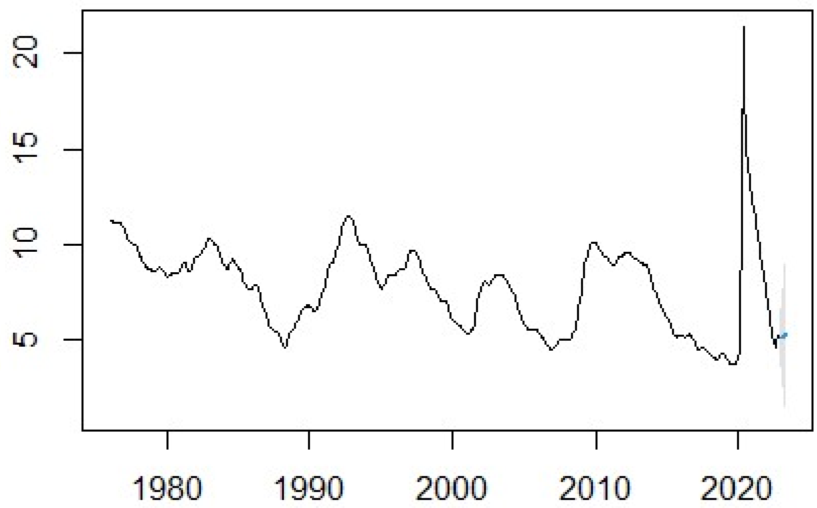
7: First-order moving average parameter.

ma2=−0.0919: Second-order moving average parameter.

sma1=−0.0655: Seasonal moving average parameter.

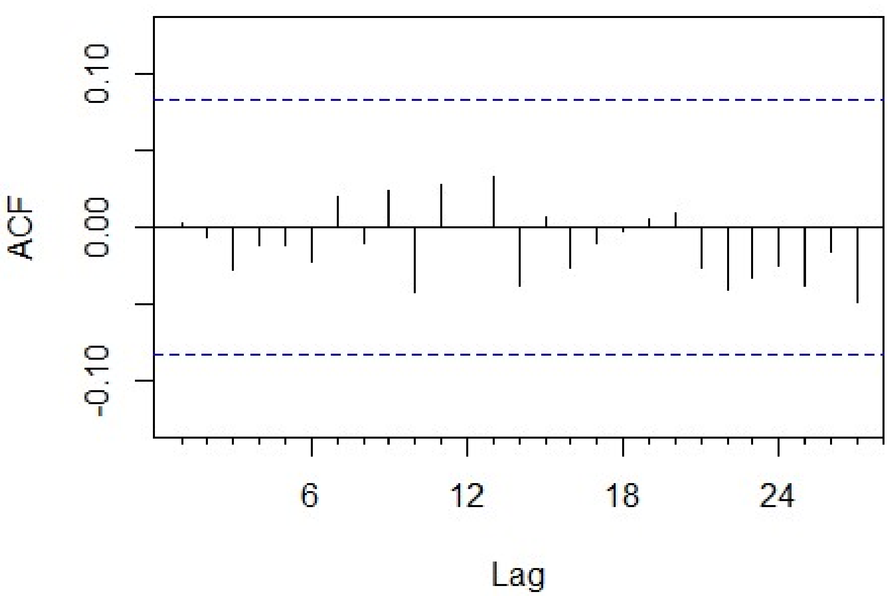
The output states that the model accounts for short-term trends using the two MA terms and one seasonal MA term a nd the seasonal patterns are identified based on the periodicity of 12.

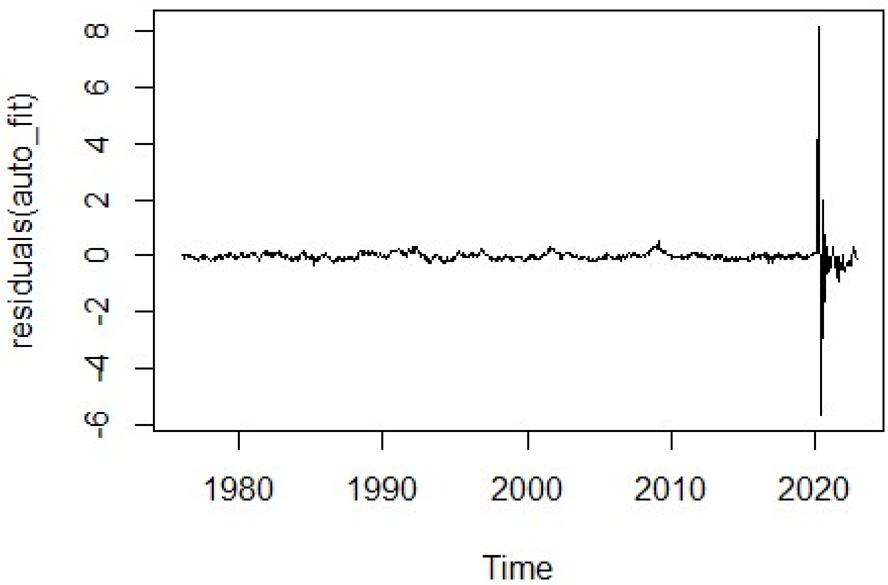
Forecasts from ARIMA(0,1,2)(0,0,1)[12]

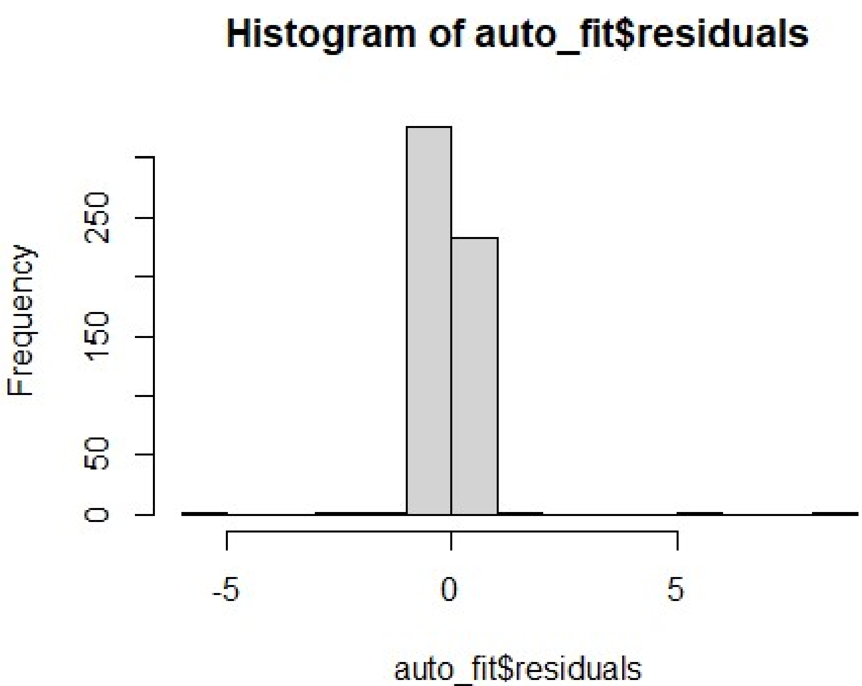


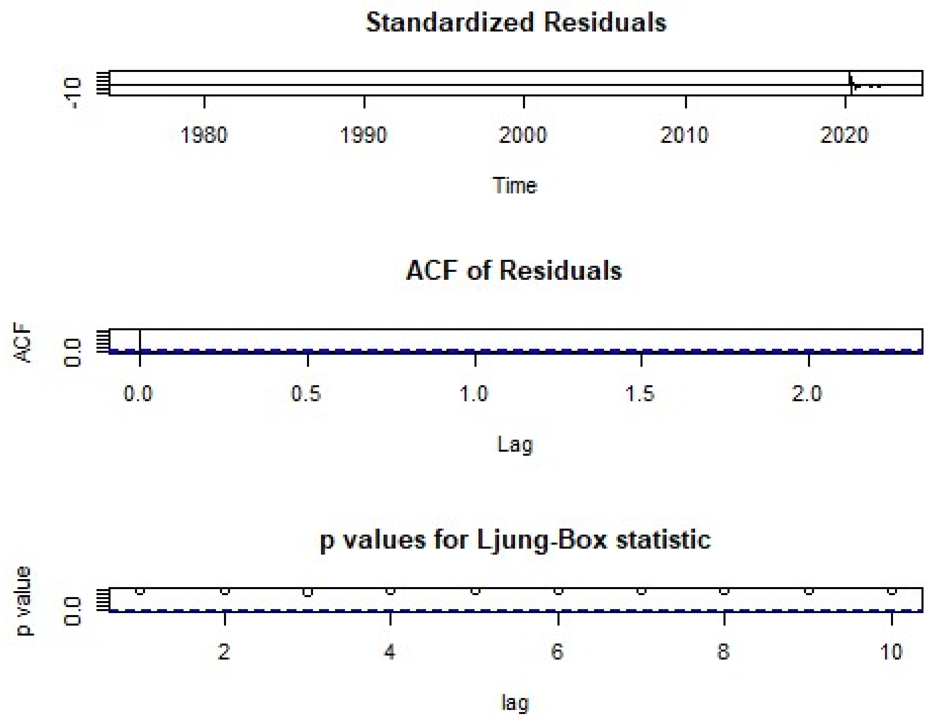
Residual Analysis:

Series auto_fit$residuals




residuals(auto_fit)
8
6
4
2
0
-2
-4
-6
1980
1990
2000
2010
2020
Time


Histogram of auto_fit$residuals
Frequency
0
50
150
250
-5
0
5
auto_fit$residuals



- The residuals show no significant autocorrelation (from the ACF plot), indicating that the model effectively captures the dynamics of the data.
- The residuals are mostly random, with no visible trends or patterns in the time series plot.
- The histogram indicates a near-normal distribution of errors.
- The residuals' variance increases toward the later time periods, hinting at possible heteroscedasticity or the presence of outliers in specific periods.

Decision based on analysis:

- **Overall Unemployment Outlook**:

  Unemployment in New York is forecasted to continue following a cyclical pattern, with periods of higher and lower unemployment likely due to seasonal factors. However, the long-term trend component will need to be closely monitored to detect any structural shifts in the labor market, such as policy changes or economic disruptions.

- **Policy and Planning Implications**:

  For policymakers and businesses, understanding these seasonal fluctuations will be critical in planning for peak unemployment periods and addressing labor market challenges.

- **Forecasting Strategy**:

  The current ETS(M,A,N) and Holt-Winters models appear to be effective for predicting unemployment. It is advisable to continue using these models while refining them periodically, especially to account for outliers or any extraordinary economic events that may disrupt the seasonal and trend patterns.

Ideas to improve the forecasts

- Continue monitoring the trend closely for any signs of significant long-term changes in the unemployment rate.
- Include key economic variables like GDP growth, inflation rates, industrial production, and consumer confidence indices, as they significantly impact unemployment trends.
- Account for changes in population size, age structure, and migration patterns, which influence labor market dynamics.
- Consider the effects of new labor laws, minimum wage adjustments, or unemployment benefits, as these can alter the labor market structure.
- Use robust statistical techniques to identify and correct for outliers caused by unusual events, such as natural disasters or sudden policy shifts.
- Use weekly or daily data (if available) to provide more granular insights into unemployment fluctuations.
- Enhance visualization techniques to identify patterns and anomalies more effectively, aiding in better interpretation and decision-making.