Forecasting Unemployment Trends: A Comparative Time Series Analysis of Colombia and the United States

Aye Nyein Thu, Mazhar Bhuyan, Yuqi Yang, Jisup Kwak

2025-04-25

2.2 United States

2.2.1 Dataset Information

The dataset used in this analysis was obtained from the International Labour Organization (ILOSTAT). It contains monthly unemployment indicators from January 2001 to December 2024, disaggregated by age group (15–24 and 25+) and gender (male, female). The data includes both unemployment levels (in thousands) and rates (in percentage terms).

Data Preparation
Four Excel files covering unemployment by age and gender, in both level and rate formats, were imported. The Month variable was converted from "YYYYMM" to standard date format. Variable names were standardized, datasets were merged by country and time, and only U.S. records were retained. All numeric columns were converted and sorted chronologically. No missing values were detected in the dataset.

Summary Statistics

Table 1 shows the summary statistics for each unemployment indicator, while Table 2 displays the first 10 rows of the cleaned dataset used in the time series analysis.

Table 1: Summary Statistics of Unemployment Data in US, 2001–2024

Variable	Mean	SD	Min	Max	N
Age15to24.Per	12.01	3.49	5.7	26.9	288
Age25above.Per	4.76	1.77	2.6	12.8	288
Female.Per	5.56	1.88	2.9	15.7	288
Male.Per	5.95	2.15	3.2	13.3	288
Total.Per	5.77	1.98	3.1	14.4	288
Age15to24.Thou	2584.03	739.48	1232.8	4869.8	288
Age25above.Thou	6368.66	2360.28	3711.0	17686.9	288
Female.Thou	4034.55	1344.29	2243.2	11494.3	288
Male.Thou	4918.15	1735.00	2842.0	11009.8	288
Total.Thou	8952.70	3016.38	5146.1	22504.1	288

Outlier Detection

Outliers were identified using Grubbs' test and a boxplot of the total unemployment rate. A clear outlier occurred in April 2020. **Both the original series and the outlier-removed series will be used for model testing.**

```
## An outlier was detected in April 2020 .
```

Table 2: First 10 Rows of the Cleaned Unemployment Dataset

Month	Age15to24.Per	Age25above.Per	Female.Per	Male.Per	Total.Per	Age15to24.Thou	Age25above.Thou	Female.Thou	Male.Thou	Total.Thou
2001-01-01	10.3	3.6	4.3	5.0	4.7	2250.4	4397.1	2861.2	3786.2	6647.5
2001-02-01	10.3	3.5	4.2	4.9	4.6	2257.5	4265.4	2803.6	3719.3	6522.9
2001-03-01	10.2	3.5	4.2	4.8	4.5	2253.0	4256.2	2817.7	3691.6	6509.3
2001-04-01	9.6	3.2	4.0	4.4	4.2	2095.0	3909.5	2650.4	3354.1	6004.5
2001-05-01	10.0	3.1	4.1	4.2	4.1	2171.4	3729.8	2723.7	3177.6	5901.2
2001-06-01	11.6	3.4	4.8	4.7	4.7	2775.3	4040.4	3204.7	3611.0	6815.7
2001-07-01	10.5	3.5	5.0	4.5	4.7	2584.7	4273.1	3328.2	3529.6	6857.8
2001-08-01	10.7	3.8	5.2	4.6	4.9	2461.3	4555.5	3434.5	3582.3	7016.8
2001-09-01	10.5	3.7	5.0	4.5	4.7	2301.4	4464.5	3325.3	3440.6	6765.9
2001-10-01	11.0	3.9	5.0	5.0	5.0	2424.3	4750.7	3348.3	3826.6	7175.0

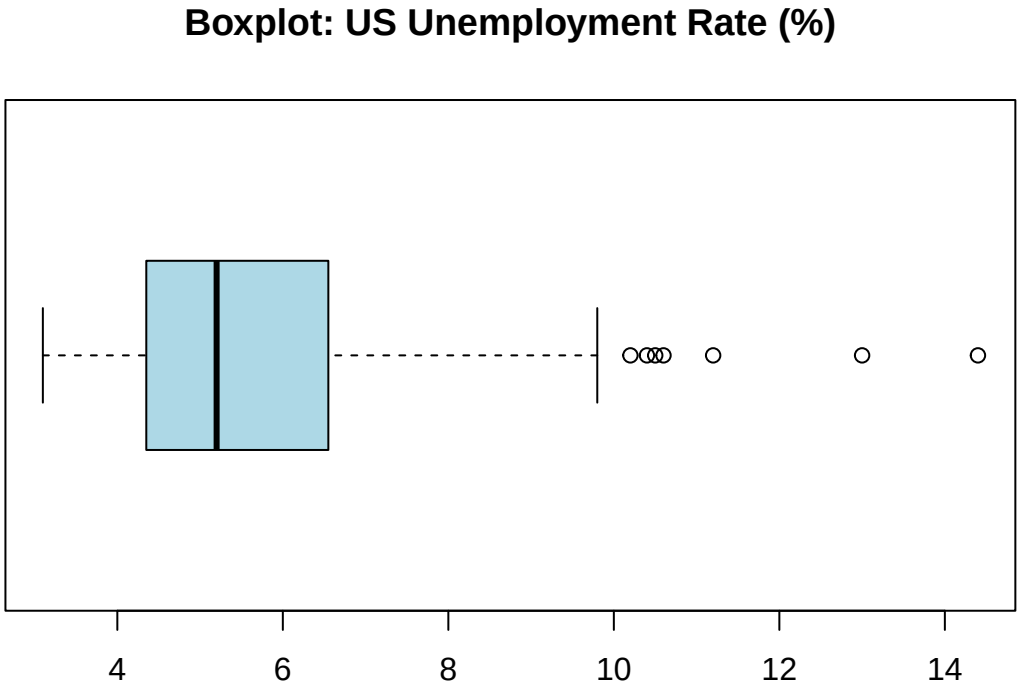

Figure 1: Boxplot of U.S. Unemployment Rate (%)

2.2.2 Time Series Analysis for US (Original Series)

2.2.2.1 Transform into time series and set training and testing windows for US (Original) To facilitate time series forecasting, the cleaned U.S. unemployment dataset was transformed into a monthly multivariate time series object using the `ts()` function. The data span from January 2001 to December 2024, with monthly frequency. To evaluate model performance under realistic forecasting conditions, a 12-month horizon corresponding to the year 2024 was reserved as the testing period. The full dataset was split into two subsets:
A **training set**, consisting of observations from January 2001 to December 2023.
A **testing set**, consisting of observations from January 2024 to December 2024.

Figure 2 shows the unemployment rate for both the training and testing periods.

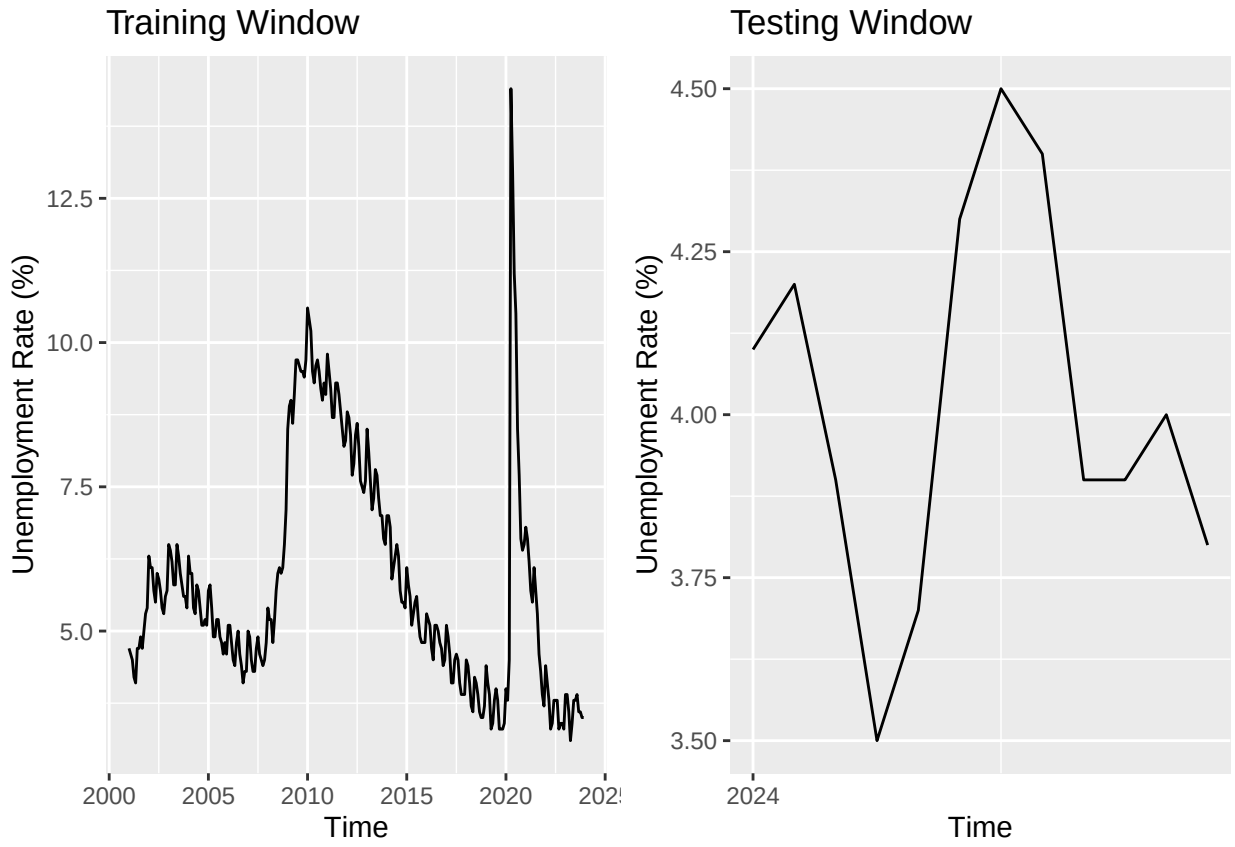

Figure 2: Time series plots of the training and testing sets

To assess the temporal dependence structure of the training data, the autocorrelation function and partial autocorrelation function were examined. As shown in Figure 3, the ACF decays slowly across lags, indicating the presence of non-stationarity and persistent autocorrelation. The PACF shows a cut off at lag 1. These patterns suggest that the series may be non-stationary with strong autocorrelation.

2.2.2.2 Decompose the time series for US (Original) The training series was decomposed into trend, seasonal, and irregular components using classical seasonal decomposition.
Figure 4 shows a strong seasonal pattern and a pronounced outlier around 2020, likely associated with the COVID-19 shock.

To remove the seasonal component, a seasonally adjusted series was generated using the `seasadj()` function. The result is shown in Figure 5.

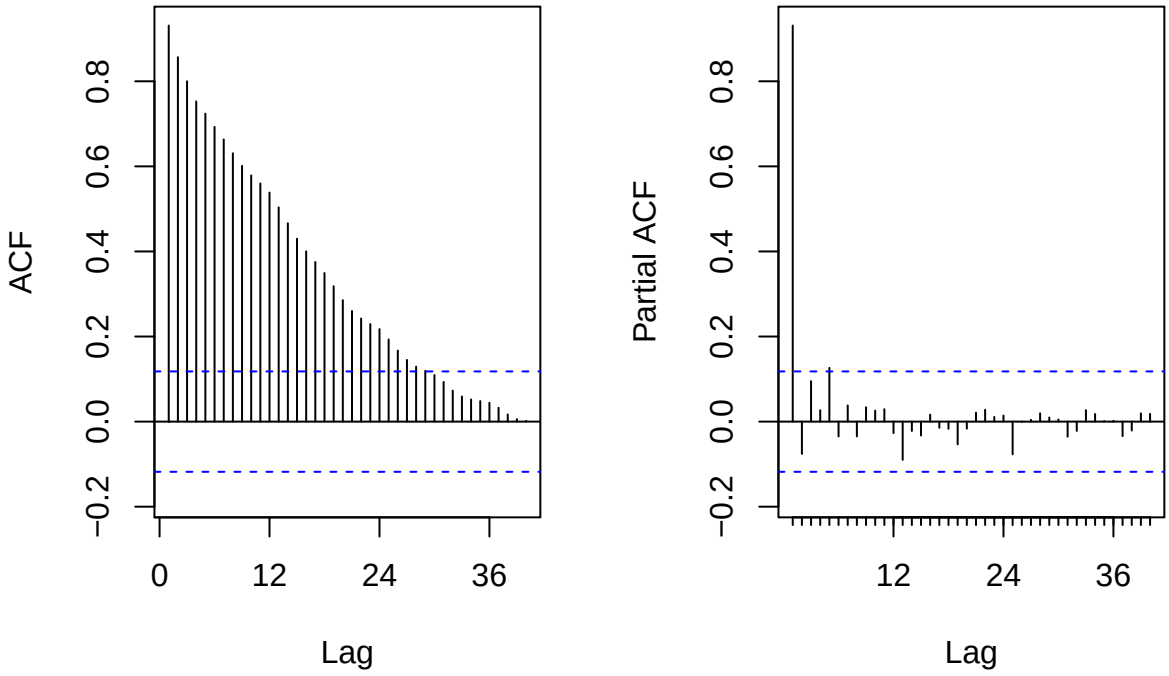

Figure 3: ACF and PACF plot of the training data

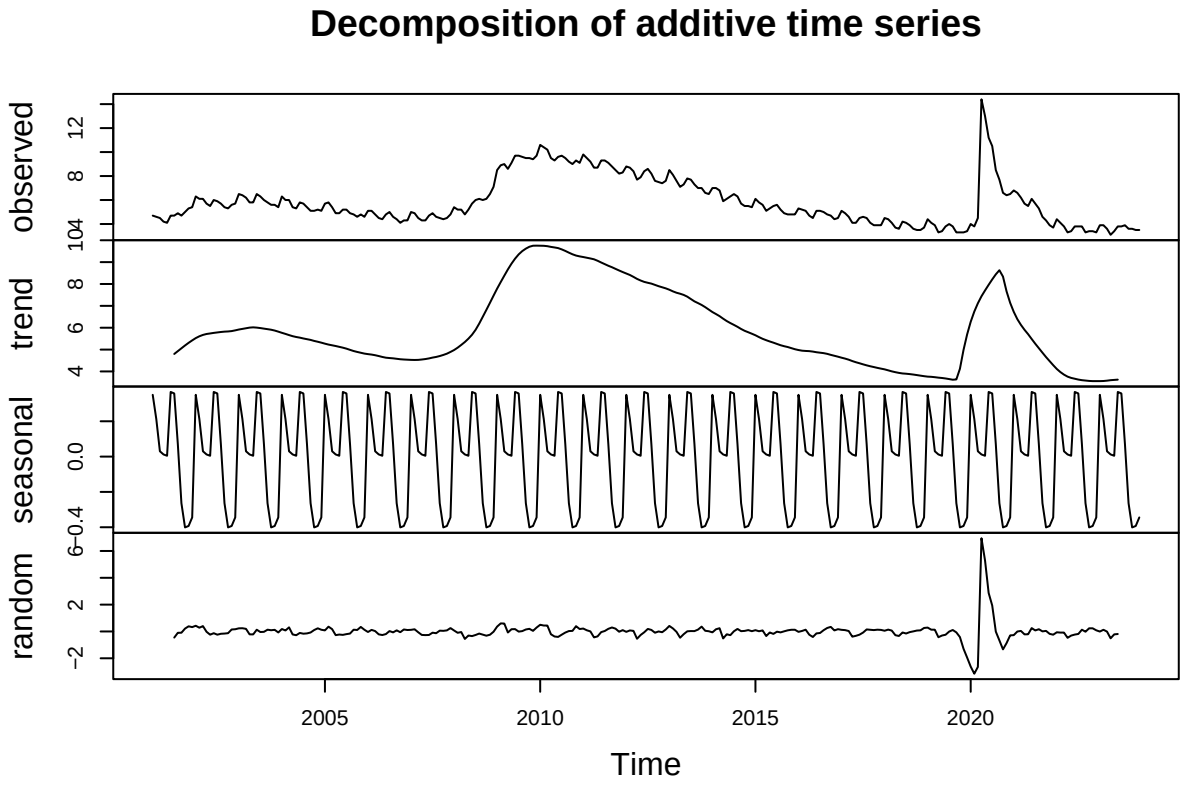

Figure 4: Decomposition of the training set

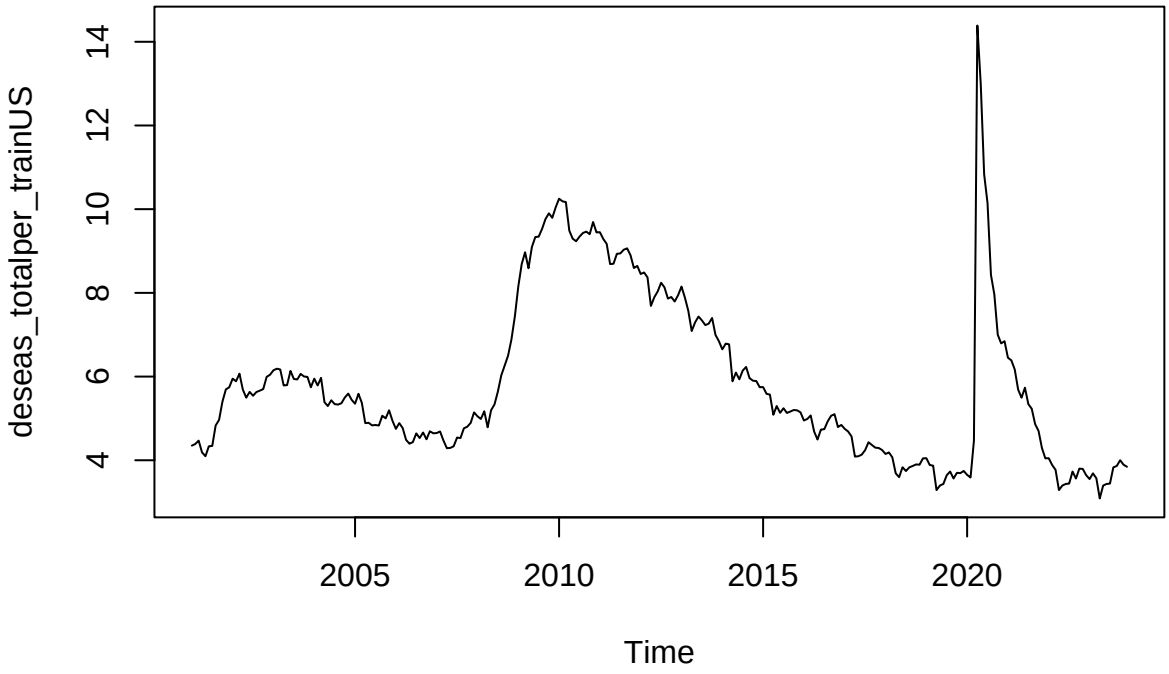

Figure 5: Seasonally adjusted U.S. unemployment rate

To evaluate the stationarity and trend properties of the series, three tests were performed.
For the seasonally adjusted series, the Augmented Dickey-Fuller test yielded a non-significant p-value ($p = 0.43$), indicating weak evidence against the unit root. However, the Mann-Kendall test confirmed a significant decreasing trend (tau = -0.27, $p < 0.001$).
For the original series, similar ADF results were observed ($p = 0.43$), while the seasonal MK test showed statistically significant seasonal patterns in several months (e.g., Season 12, $p < 0.01$), indicating strong seasonality.

The number of differences required to achieve stationarity was evaluated using the `ndiffs()` function. Both the original and seasonally adjusted series were found to require one order of differencing.

```
## Differencing requirement for original series:

## [1] 1

##
## Differencing requirement for deseasoned series:

## [1] 1
```

2.2.2.3 Test Time Series Models for US (Original) Eleven time series models were implemented to forecast the U.S. unemployment rate and evaluate predictive performance in this analysis:

1. **Seasonal Naïve (SNAIVE)** replicates the seasonal pattern from the previous year as a baseline.
2. **Simple Moving Average (SMA)** smooths fluctuations by taking an equal-weighted average over a fixed number of recent observations.
3. **Simple Exponential Smoothing (SES)** applies exponentially decreasing weights to past values, emphasizing more recent data points.
4. **SARIMA** captures both autoregressive and seasonal structures in the original series using `auto.arima()`.
5. **ARIMA** (applied to deseasonalized data) models trend and autoregressive components without explicitly accounting for seasonal patterns.
6. **STL + ETS** combines trend-seasonal decomposition with exponential smoothing using `stlf()`.
7. **ARIMA + Fourier** introduces seasonal flexibility through Fourier terms as external regressors.
8. **TBATS** incorporates Box-Cox transformation, ARMA errors, and trigonometric seasonality.
9. **Neural Network (NNETAR)** captures nonlinear dynamics using lagged inputs in a feedforward network.
10. **State Space Exponential Smoothing (SSES)** selects optimal components automatically.
11. **State Space with BSM** models level, trend, and seasonality stochastically.

Residual diagnostics were conducted using the `checkresiduals()` function, including ACF plots and the Ljung-Box test. Most models generated residuals consistent with white noise, indicating proper specification.

Table 3: Forecast Accuracy for Unemployment Rate (%) Data

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.38333	0.40620	0.38333	9.46470	9.46470	0.28333	1.35284	4.93545
SMA	0.17216	0.33070	0.26108	3.80965	6.29631	0.44396	1.06312	3.31350
SES	0.17215	0.33070	0.26108	3.80953	6.29625	0.44396	1.06311	3.31347
SARIMA	0.51667	0.58878	0.51667	12.42930	12.42930	0.44396	1.85999	6.50904
ARIMA	0.17216	0.33070	0.26108	3.80965	6.29631	0.44396	1.06312	3.31350
ETS	-0.09918	0.54307	0.39400	-2.91283	10.30378	0.55949	1.95874	5.42342
ARIMA_FOURIER	0.40391	0.45420	0.40391	9.80344	9.80344	0.50899	1.46803	5.12882
TBATS	0.32666	0.38544	0.34616	7.88226	8.43937	0.53705	1.26470	4.41240
NNETAR	-5.14902	6.73100	5.17795	-128.59997	129.28874	0.84727	22.63873	68.00987
SSES	1.52521	1.71249	1.52948	38.05395	38.15815	0.73123	5.77401	19.93532
BSM	-0.06016	0.32935	0.22841	-1.82484	5.93544	0.57974	1.18115	3.13239

SMA, SES, ARIMA, SARIMA, STL + ETS, ARIMA + Fourier, TBATS, NNETAR, and SSES all passed residual tests, with Ljung-Box p-values ranging from 0.11 to 0.99.
SNAIVE, as expected for a benchmark, showed strong autocorrelation in residuals and failed the test.
State Space with BSM exhibited the poorest performance in residual checks, with significant autocorrelation ($p < 2.2e\text{-}16$) and visible spikes in the ACF plot, suggesting model misspecification.

2.2.2.4: Performance check for US (Original) To assess out-of-sample performance, forecasts were generated for the 12-month testing period in 2024 and compared to observed unemployment rates. Model accuracy was evaluated using six standard metrics: Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), autocorrelation at lag 1 (ACF1), and Theil's U statistic. A composite score was calculated as the average of RMSE and MAPE.

The results are summarized in Table 3. Based on the average of RMSE and MAPE, the State Space with BSM delivered the best forecast accuracy, followed by the SES and ARIMA models.

For visual comparison, Figure 6 displays the test set alongside forecasted values from selected models. SMA and SSES were excluded due to technical issues. Notably, the NNETAR model failed to capture the sharp spike observed in the actual 2024 data.

```
## The best model by Average is: BSM
```

2.2.2.5 Forecast for 2025 with the best three models for US (Original) To forecast U.S. unemployment trends for 2025, the three best-performing models based on out-of-sample accuracy, **Basic Structural Model (BSM), Simple Exponential Smoothing (SES), and ARIMA**, were retrained using the full dataset from January 2001 to December 2024. Each model was then used to generate 12-month ahead forecasts for the year 2025.

Figure 7 presents the overlay of all three forecasted series against the historical unemployment trend. While all models project a relatively stable unemployment rate in 2025, the BSM model shows slightly more variation. The SES and ARIMA models produce smoother forecasts.

2.2.3 Time Series Analysis for US (Outliers-removed Series)

While the original unemployment series revealed meaningful seasonal and trend components, it also included extreme fluctuations, particularly during the COVID-19 period. To ensure model robustness, an outlier-

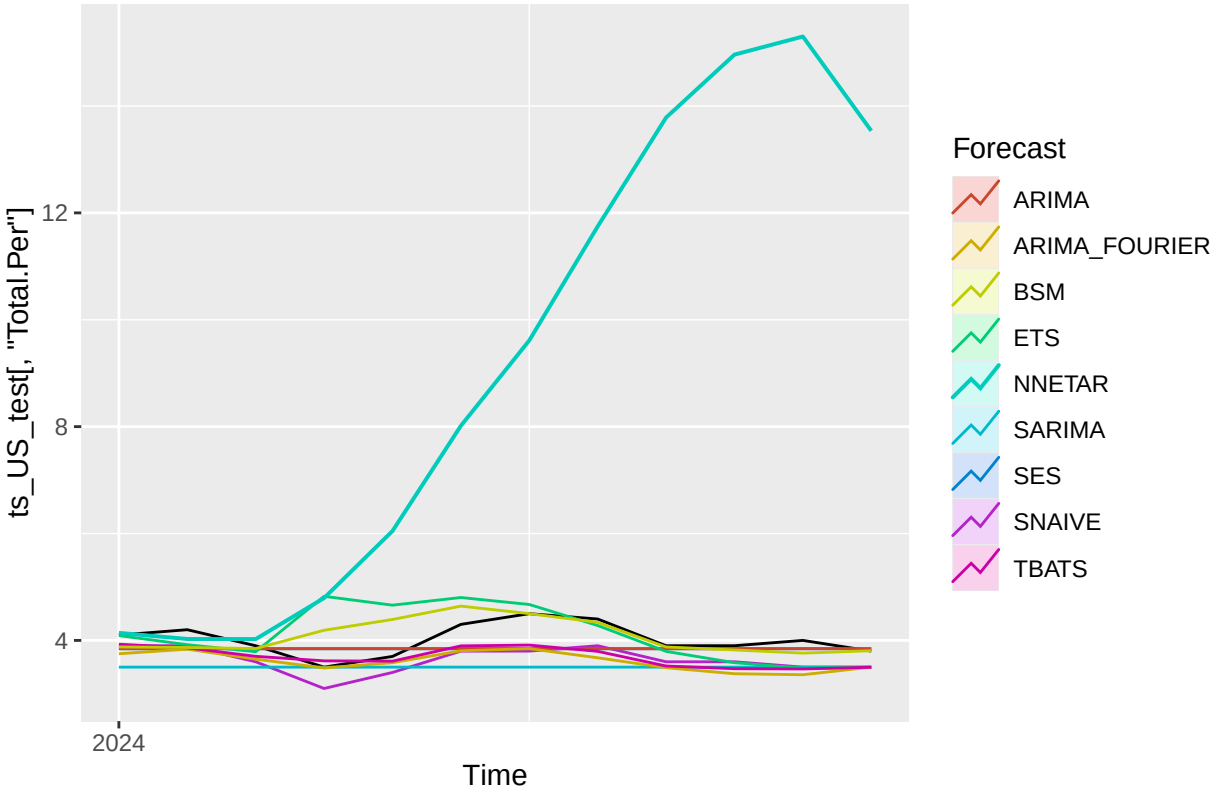

Figure 6: Forecast Comparison Across Models

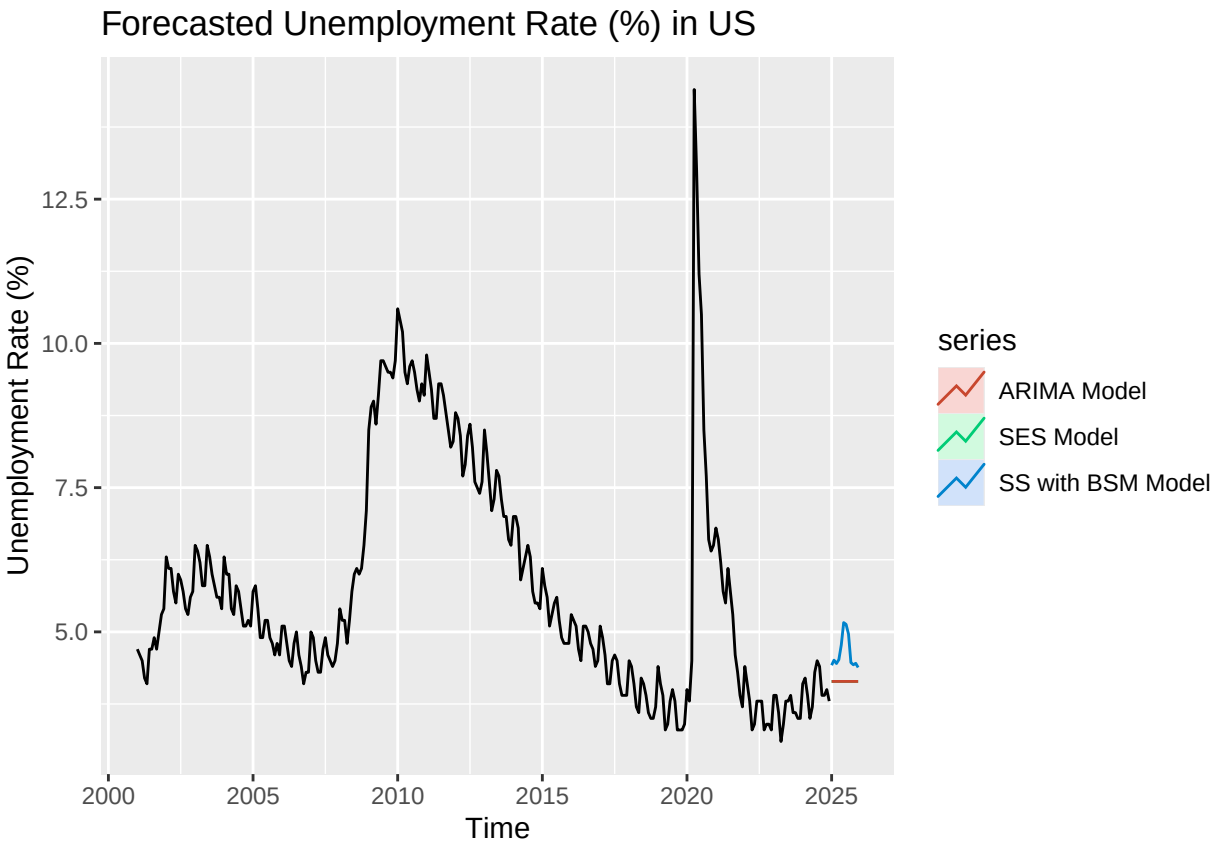

Figure 7: Forecasts U.S. Unemployment Rate in 2025 from Top Three Models

adjusted version of the series was created and analyzed using the same framework as the original. This section presents the results based on the outliers-removed dataset.

2.2.3.1 Transform into time series and set training and testing windows for US (Outliers) To address extreme deviations, the original unemployment rate series was smoothed using the `tsclean()` function. As shown in Figure 8, the adjusted series closely follows the original trajectory, except for a visibly reduced spike in early 2020.

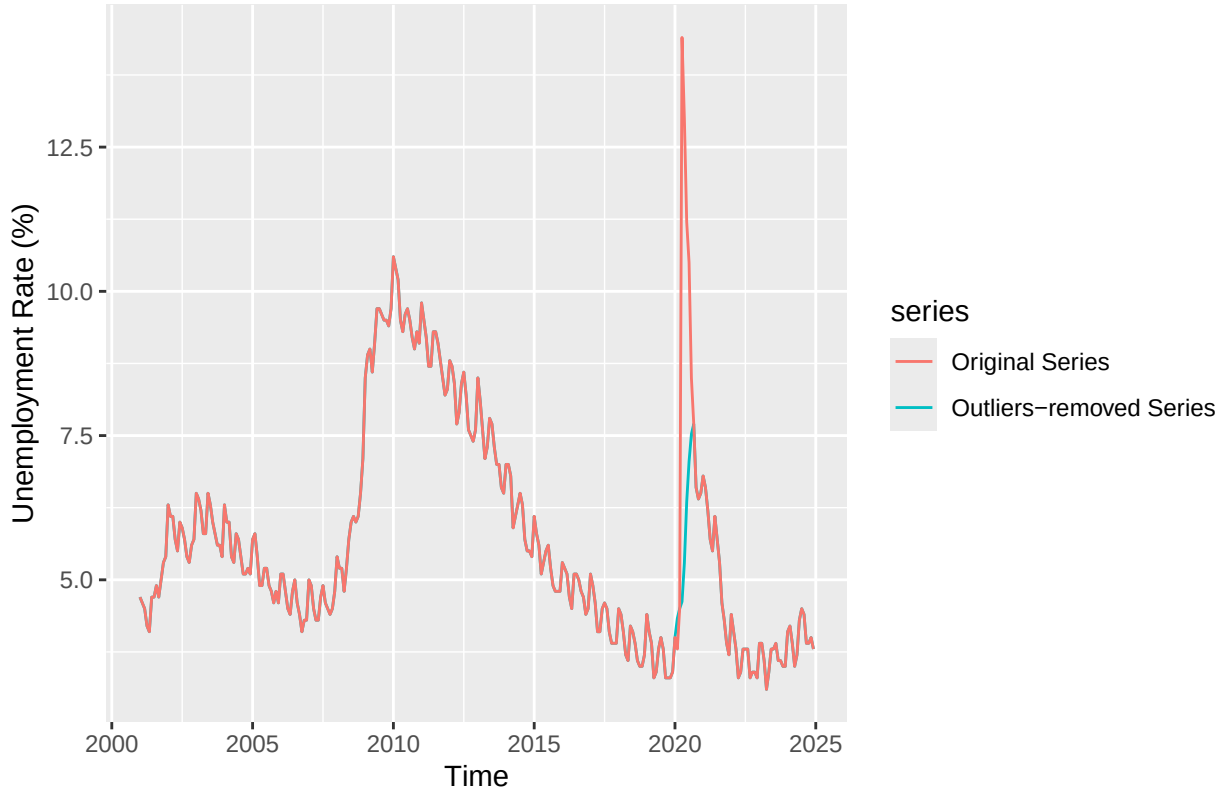

Figure 8: Comparison of Original and Outliers-removed Series (US)

Following the same procedure as before, the outliers-removed series was split into a training set (January 2001 to December 2023) and a testing set (January to December 2024). These segments are illustrated in Figure 9.

Temporal dependence was assessed using ACF and PACF plots. As shown in Figure 10, the outliers-removed series retains strong autocorrelation, with a slow ACF decay and a sharp PACF cut-off at lag 1, indicating the underlying dependency structure remains intact.

2.2.3.2 Decompose the time series for US (Outliers) The outliers-removed training set was decomposed into trend, seasonal, and irregular components using classical additive decomposition. As shown in Figure 11, the seasonal pattern remains strong, and the COVID-related spike appears in the irregular component, though its magnitude is attenuated.

To remove the seasonal effect, a seasonally adjusted series was created using the `seasadj()`. The deseasonalized series is plotted in Figure 12.

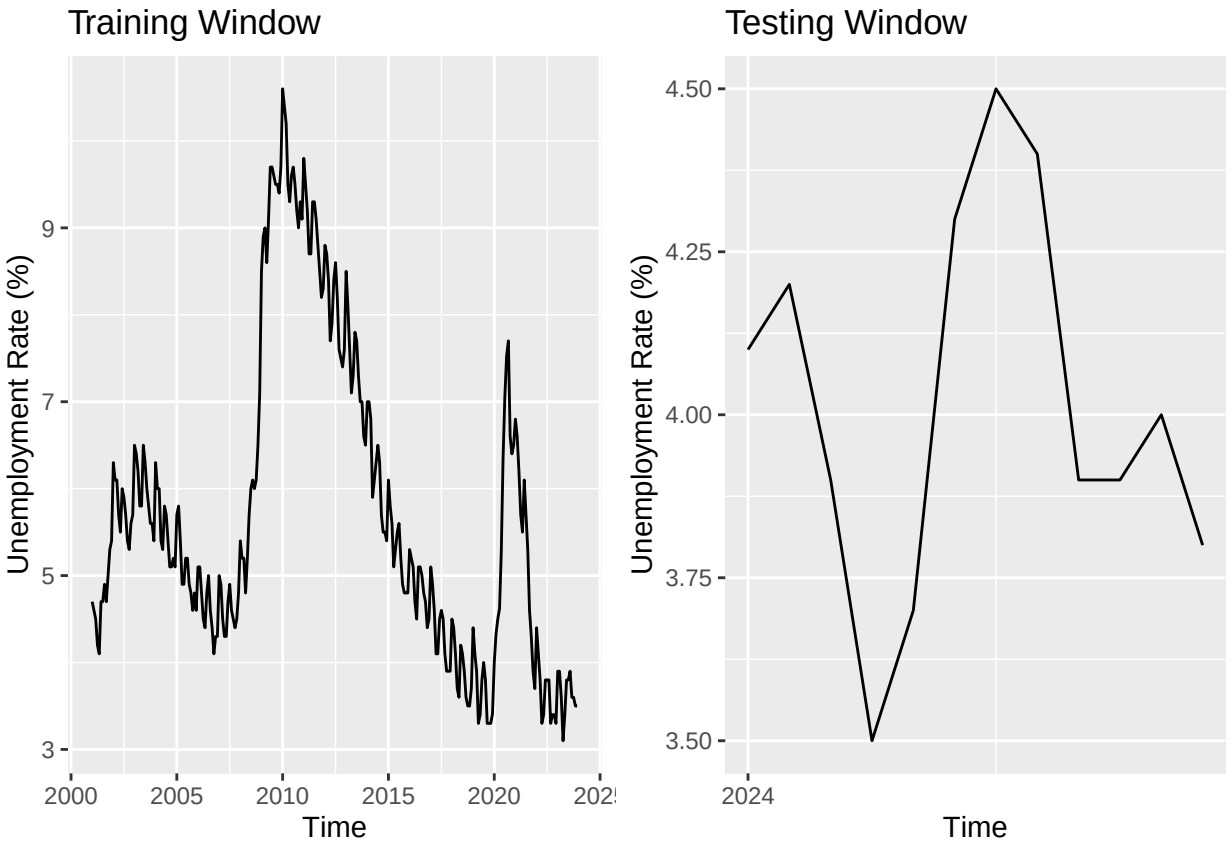

Figure 9: Time series plots of the training and testing sets (Outliers-Removed Series)

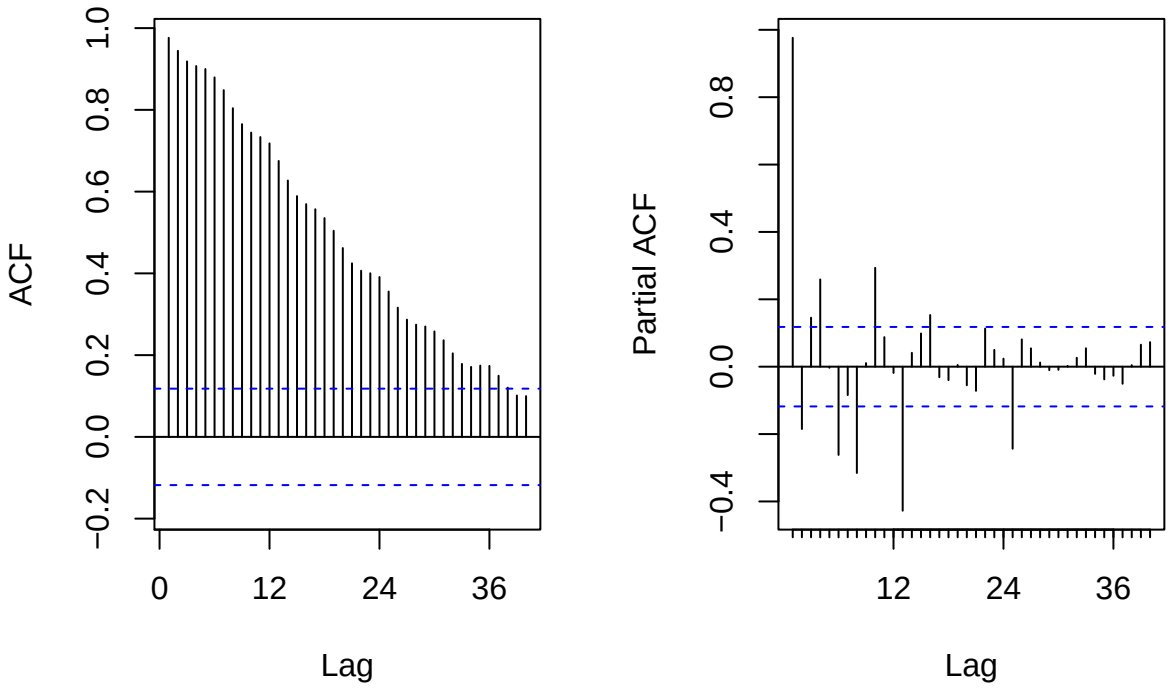

Figure 10: ACF and PACF plot of the training data (Outliers-Removed Series)

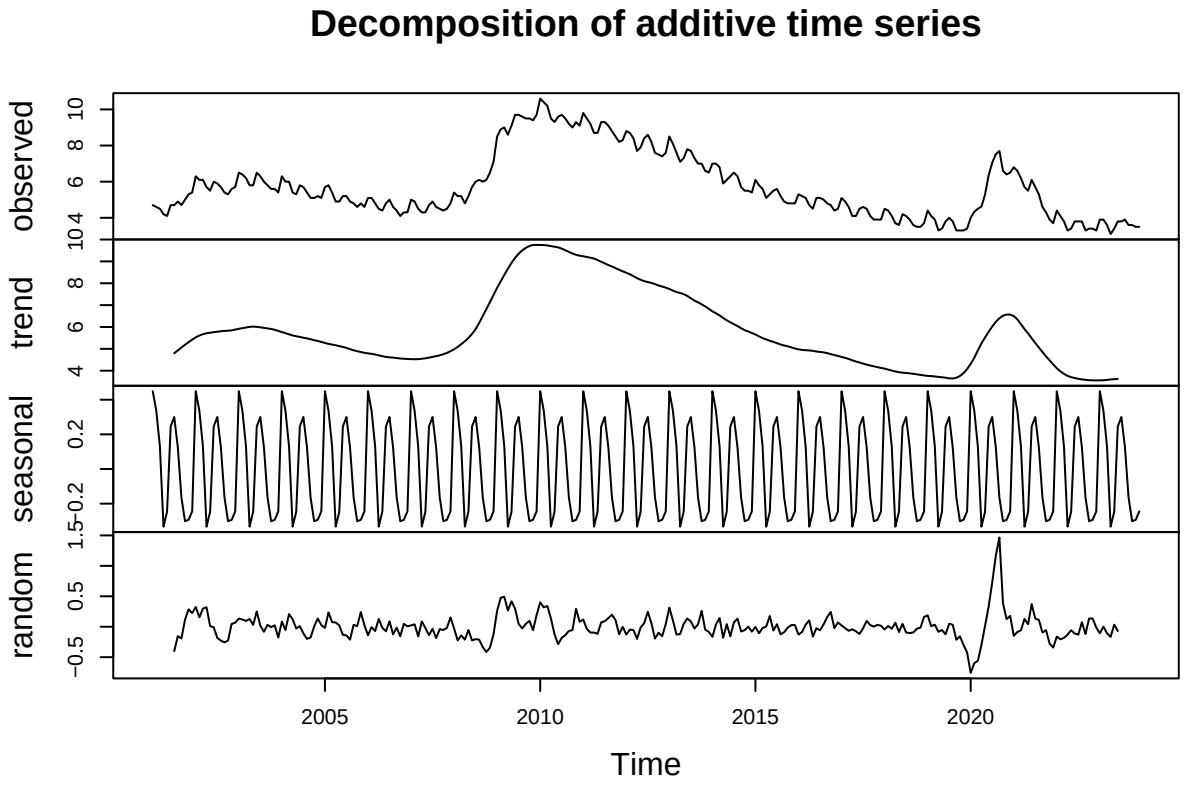

Figure 11: Decomposition of the training set (Outliers-Removed Series)

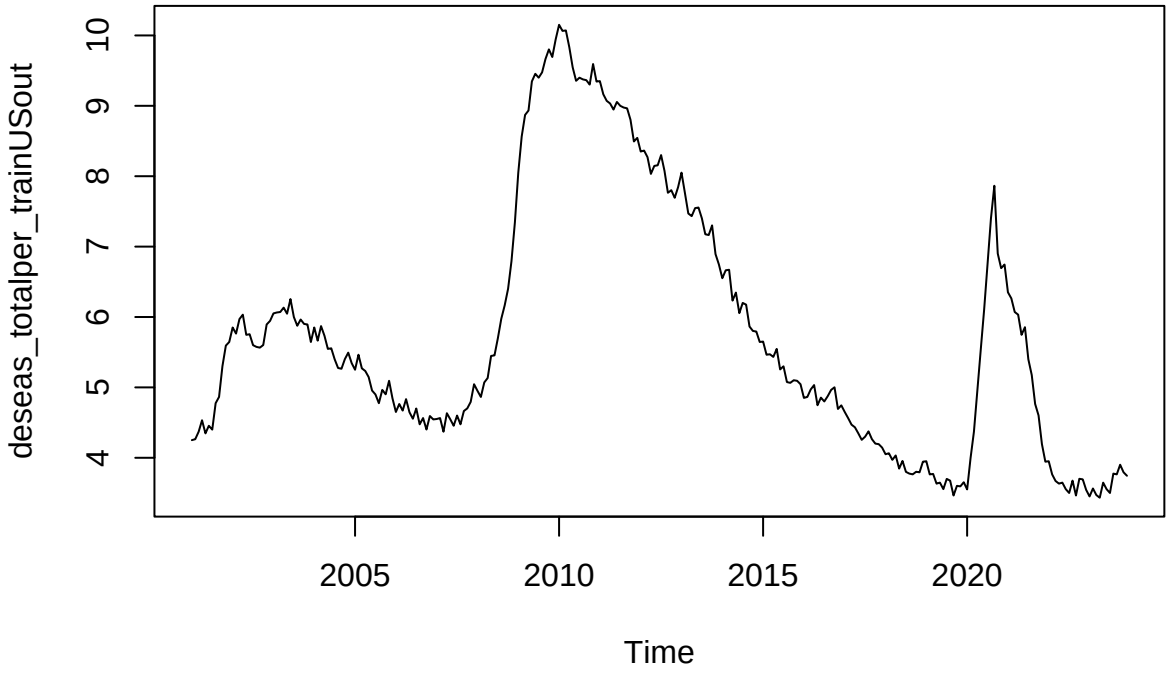

Figure 12: Seasonally adjusted U.S. unemployment rate (Outliers-Removed Series)

Stationarity and trend were evaluated using three tests:
The ADF test produced non-significant p-values for both original ($p = 0.5664$) and deseasoned ($p = 0.4728$) series, suggesting the presence of unit roots.
The Mann-Kendall test on the deseasoned series showed a significant downward trend (tau = -0.305, $p < 0.001$).
The Seasonal Mann-Kendall test confirmed significant seasonal trends, particularly in January, February, March, and December ($p < 0.05$).
These results indicate that, after outlier adjustment, the series retains both trend and seasonality.

The `ndiffs()` function indicated that both the original and deseasoned series required one order of differencing, consistent with findings from the unadjusted series.

```
## Differencing requirement for outliers-removed original series:

## [1] 1

##
## Differencing requirement for outliers-removed deseasoned series:

## [1] 1
```

2.2.3.3: Test Time Series Models for US (Outliers) To evaluate model robustness, the same eleven forecasting models were applied to the outliers-removed series. Each model was assessed using residual diagnostics based on the `checkresiduals()` function, which includes ACF plots and the Ljung-Box test.

1. **SNAIVE**: Residuals exhibited strong seasonal autocorrelation and failed the white noise test ($p <$ 2.2e-16), confirming model inadequacy.

2. **SMA(1)**: Smoothed the series but left significant autocorrelation ($p < 0.001$), suggesting underfitting.

3. **SES**: Produced visually acceptable fits, but residuals showed serial dependence ($p <$ 1e-15), indicating insufficient complexity.

4. **SARIMA(4,1,0)(1,1,1)[12]**: Captured short- and long-term structure but failed the Ljung-Box test ($p = 0.005$), with mild residual autocorrelation.

5. **ARIMA(1,1,2)** (on deseasoned series): Achieved the best residual behavior among all models, with white-noise residuals confirmed ($p = 0.037$).

6. **STL + ETS**: Forecasts were reasonable, but residuals failed diagnostic tests ($p =$ 1e-5), particularly during volatile periods.

7. **ARIMA + Fourier**: Despite incorporating seasonal terms via Fourier regressors, residuals remained autocorrelated ($p =$ 1.4e-08), indicating incomplete seasonal modeling.

8. **TBATS**: Aimed at complex seasonal dynamics, but residuals showed substantial autocorrelation, especially around the COVID-19 period.

9. **NNETAR**: Captured nonlinearities but exhibited residual serial dependence ($p = 0.00015$), suggesting overfitting or limited generalization.

10. **SSES (Exponential Smoothing)**: Residuals remained autocorrelated, confirming that exponential smoothing alone was insufficient for structural irregularities.

Table 4: Forecast Accuracy for Unemployment Rate (%) Data

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.38333	0.40620	0.38333	9.46470	9.46470	0.28333	1.35284	4.93545
SMA	0.27186	0.39196	0.32013	6.30418	7.67176	0.44396	1.24293	4.03186
SES	0.27185	0.39195	0.32013	6.30406	7.67168	0.44396	1.24292	4.03182
SARIMA	0.29883	0.35509	0.30781	7.39350	7.61242	0.58243	1.20095	3.98375
ARIMA	0.35322	0.45308	0.38245	8.34064	9.17576	0.46153	1.44837	4.81442
ETS	0.08226	0.13550	0.10280	2.00098	2.52297	0.12960	0.46055	1.32923
ARIMA_FOURIER	0.50968	0.55033	0.50968	12.49002	12.49002	0.55406	1.82006	6.52017
TBATS	0.44741	0.49729	0.44741	10.95395	10.95395	0.62666	1.65645	5.72562
NNETAR	0.09033	0.13552	0.10614	2.11667	2.54494	0.19699	0.44444	1.34023
SSES	0.41563	0.47631	0.42271	10.27258	10.44526	0.65134	1.60031	5.46079
BSM	0.05951	0.12113	0.10407	1.37766	2.53923	0.04463	0.39014	1.33018

11. **State Space with BSM**: Residuals showed clear autocorrelation ($p < 2.2e\text{-}16$), with spikes in ACF plots, indicating poor fit despite the model's theoretical flexibility.

2.2.3.4 Performance check for US (Outliers) To evaluate forecast performance, all eleven models were assessed against the 2024 testing set. Forecast accuracy was measured using multiple criteria, including RMSE, MAE, MAPE, and an average score calculated as the mean of RMSE and MAPE.

Table 4 summarizes the results. The NNETAR model achieved the highest overall performance, with the lowest average error (1.25235), best RMSE (0.12675), and lowest MAPE (2.38%). These results suggest that NNETAR effectively captured both linear and nonlinear dynamics in the cleaned series.

The ETS and BSM models also performed well, with average errors below 1.34 and MAPE values around 2.5%. Their low ACF1 values (0.13 and 0.04, respectively) indicate minimal residual autocorrelation and well-specified forecasts.

In contrast, baseline models such as SNAIVE, SMA, and SES showed higher average errors (> 4.0), highlighting their limited adaptability to the adjusted series. Models like ARIMA + Fourier, TBATS, and SSES also underperformed, with average errors exceeding 5 and persistent autocorrelation, suggesting poor fit even after outlier removal.

Figure 12 displays forecast trajectories alongside actual unemployment data in 2024. The top models (NNETAR, ETS, BSM) closely tracked observed trends, while models like SNAIVE, SES, and TBATS showed visible deviations, consistent with the quantitative rankings.

```
## The best model by Average is: ETS
```

2.2.3.5 Forecast for 2025 with the best three models for US (Outliers) To forecast the U.S. unemployment rate for 2025, the three models with the best performance in the previous section were selected: **State Space with BSM, NNETAR, and ETS**. Each model was retrained using the full outliers-removed dataset from 2001 to 2024, and forecasts were generated for the 12 months of 2025.

Figure 14 presents the forecasted unemployment trajectories. All three models effectively capture the underlying trend and seasonal fluctuations. The projected unemployment rates fall within a range of 3.5% to 5.0% across the year.

NNETAR shows more responsive short-term movements but slightly underestimates the unemployment rate compared to the other models. The ETS model provides a smoother trend-following projection. The BSM model, which incorporates structural components, appears more responsive to recent data shifts. While the models differ in sensitivity and volatility, their trajectories remain consistent, which supports confidence in the robustness of the forecasts.

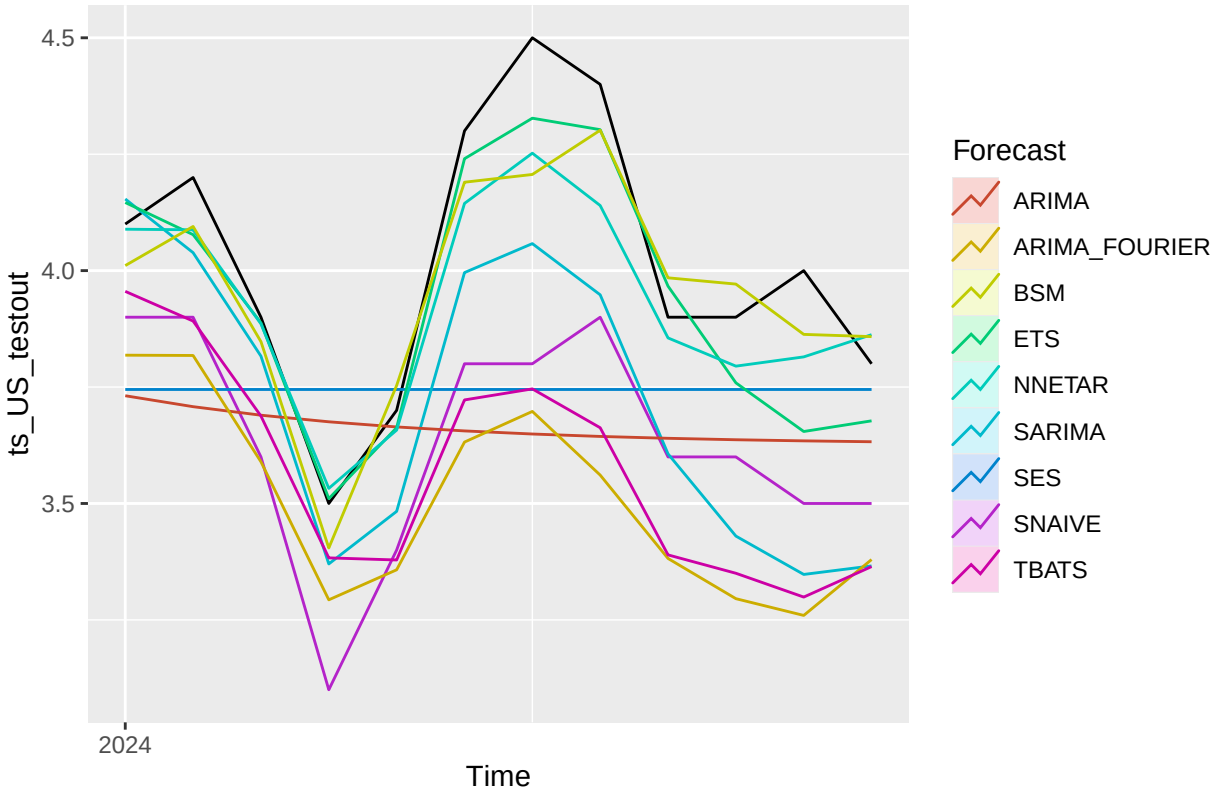

Figure 13: Forecast Comparison Across Models (Outliers Removed)

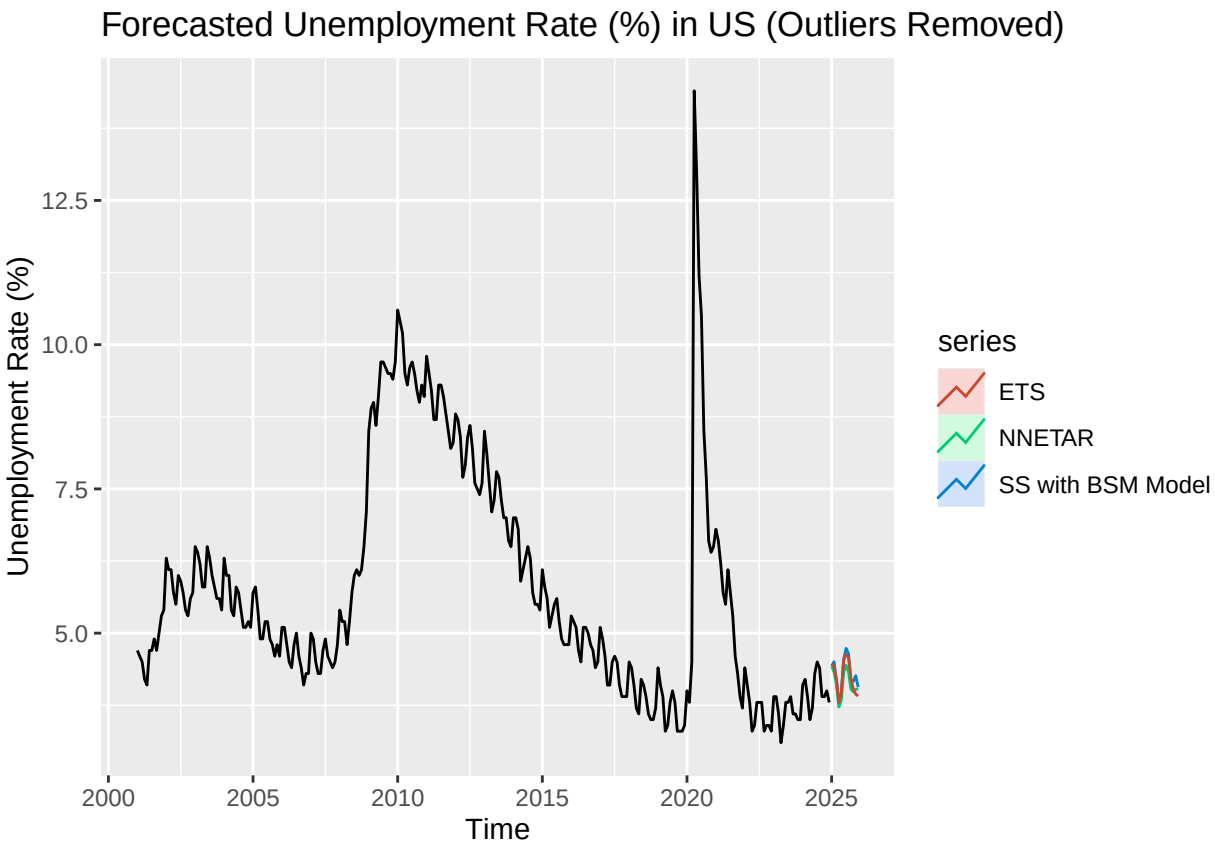

Figure 14: Forecasts from Top Three Models (2025, US, Outliers-Removed)

2.2.3.6 The average of 3 forecasts To provide a unified prediction, the forecasts from the three selected models were averaged. The combined forecast captures both seasonality and the recent upward trend in U.S. unemployment.

As shown in Figure 15, the combined forecast aligns closely with observed values from recent years and projects a stable outlook for 2025. The monthly unemployment rate is expected to range from 3.75% to 4.54%, with a yearly average of 4.17%. This is very close to the ILO's projection of 4.3%, suggesting that the combined model estimate is both credible and well-calibrated.

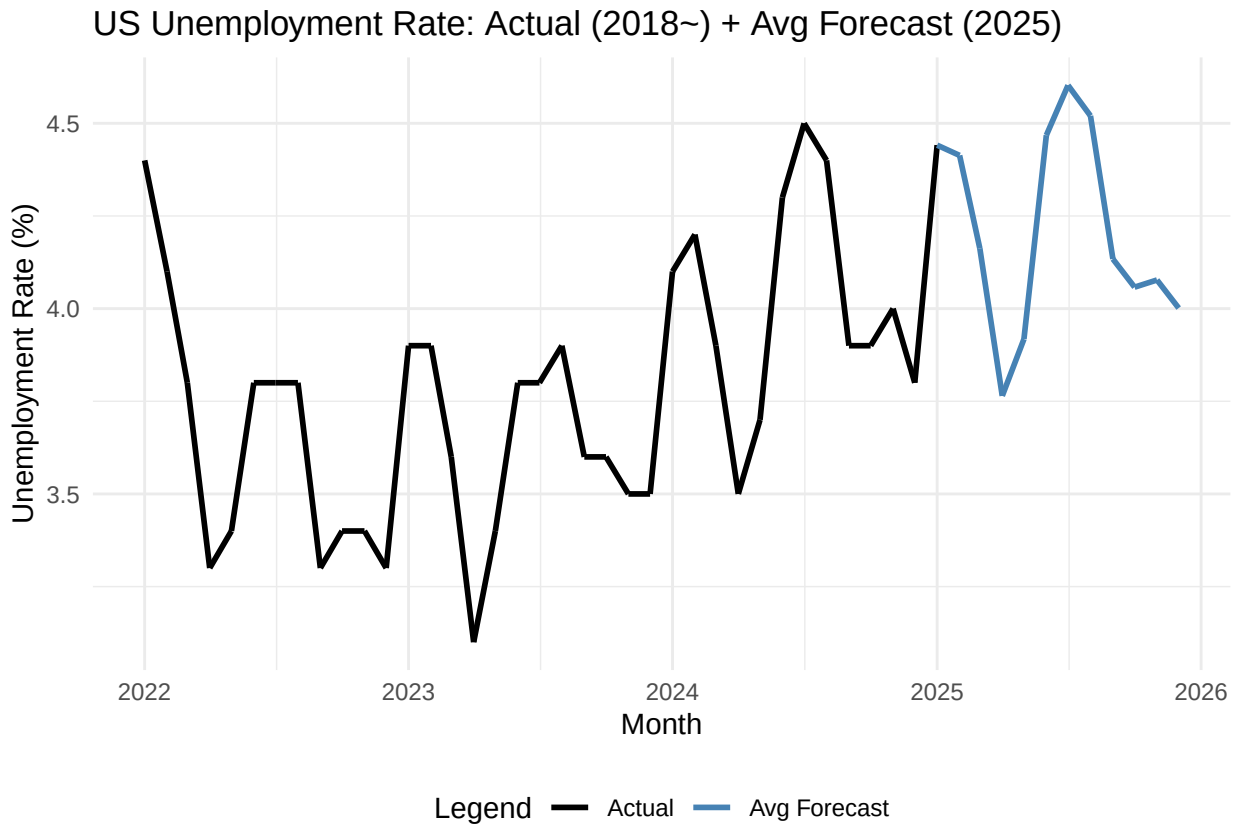

Figure 15: Average Forecast of U.S. Unemployment Rate for 2025