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

Contents

List of Figures

List of Tables

1 Introduction and Motivation

Unemployment is more than just a statistic — it is a dynamic reflection of economic health, social well-being, and public sentiment. Unemployment rate changes can signal imminent expansions, downturns, or crises when macroeconomic conditions change. Accurately predicting the unemployment rate becomes more than just a technical task during times of volatility, uncertainty, or reform; it becomes a policy need.

Recent global events, such as pandemic shocks and financial crises, have highlighted how urgent it is to conduct proactive labor market monitoring. Forecasts of unemployment are becoming more and more important to investors, labor groups, and policymakers as they plan interventions, allocate resources, and manage expectations. Unemployment projections are especially useful because they respond sensitively to macroeconomic signal signals, such as changes in monetary policy, fiscal stimuli, and structural reforms, as mentioned by Montgomery et al. (1998) and confirmed in subsequent research.

Furthermore, several of the Sustainable Development Goals (SDGs), especially those pertaining to gender equality, decent employment, and poverty alleviation, are heavily reliant on labor market outcomes. Underemployment and informality frequently coincide with structural unemployment in developing nations like Colombia, making response tactics more challenging. Advanced economies such as the United States, on the other hand, exhibit unemployment trends that are more sensitive to global shocks but also more cyclically responsive.

The main objectives of the report are:

- Utilizing time series methods forecast the unemployment rates for the US and Colombia over a 12-month period;
- Compare the forecast sensitivity and accuracy of several models to give policymakers early warning signals;
- Analyze differences in the two nations' anticipated trajectories in light of their respective economic systems;
- Highlight the benefits and drawbacks of forecasting techniques, especially in unstable macroeconomic settings.

By modeling unemployment not merely as a legacy indicator but as a forward-looking economic signal, this report seeks to contribute to a more responsive and resilient policy framework — one that can anticipate change, not just react to it.

2 Dataset Description and Data Wrangling

The dataset used in this study is sourced from the International Labour Organization (ILO) and contains monthly unemployment data for the United States and Colombia, segmented by sex and age group. To conduct a comparative and consistent analysis, we have constructed a unified time series using the age-group data by summing unemployment counts and calculating a weighted average percentage.

Since percentages are normalized and directly comparable across countries and time, the main variable used for forecasting is the **total unemployment percentage (Total.Per)**. All data were checked for missing values and structural inconsistencies before modeling.

Unemployment rates across the globe have seen significant fluctuations over the decades, often reflecting the impact of global economic crises, policy shifts, and technological change. Historically, an unemployment rate of around 4% is often regarded as full employment, meaning most individuals willing and able to work can find jobs. The global plot shows that many economies experienced sharp spikes during major recessions (e.g., 2008) and more recently during the COVID-19 pandemic in 2020.

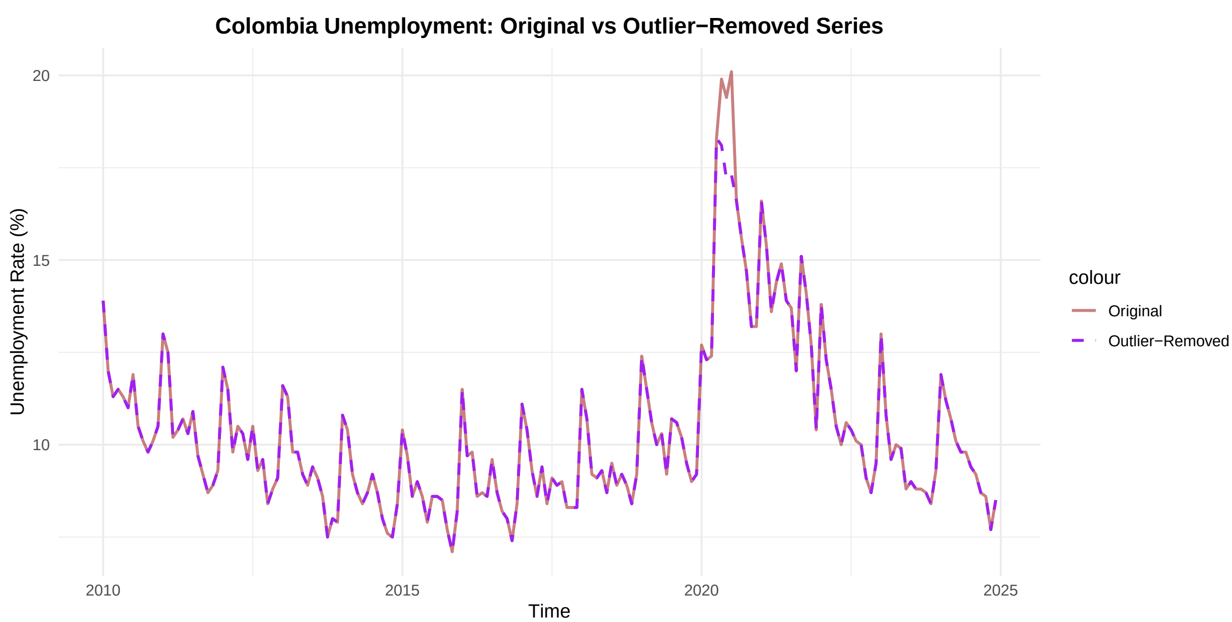
Table 1: Summary Statistics: United States and Colombia

	United States					Colombia				
Variable	US_Mean	US_SD	US_Min	US_Max	US_N	Col_Mean	Col_SD	Col_Min	Col_Max	Col_N
Age15to24.Per	12.008681	3.488402	5.7	26.9	288	21.563768	3.441774	15.1	34.4	276
Age25above.Per	4.761458	1.774522	2.6	12.8	288	8.460145	2.002818	5.3	17.9	276
Female.Per	5.561458	1.876113	2.9	15.7	288	13.959420	2.916427	9.1	25.5	276
Male.Per	5.949306	2.148008	3.2	13.3	288	8.736594	1.947194	5.3	17.9	276
Total.Per	5.765972	1.982789	3.1	14.4	288	10.914130	2.277091	7.1	20.1	276
Age15to24.Thou	2584.030208	739.483180	1232.8	4869.8	288	852.173188	144.136962	551.9	1265.4	276
Age25above.Thou	6368.664931	2360.277964	3711.0	17686.9	288	1494.322101	399.159524	962.7	3267.4	276
Female.Thou	4034.548611	1344.285566	2243.2	11494.3	288	1253.461594	237.993991	864.1	2193.1	276
Male.Thou	4918.150000	1734.999926	2842.0	11009.8	288	1093.032971	235.619647	698.0	2223.8	276
Total.Thou	8952.700347	3016.384542	5146.1	22504.1	288	2346.493116	461.038716	1641.0	4373.4	276

However, what's striking is the rapid recovery of unemployment rates post-COVID in many regions. As reflected in the global and country-level panels, the US shows a pronounced spike in 2020 followed by a fast recovery, thanks to aggressive fiscal and monetary responses. Colombia, while also experiencing a peak, exhibits greater volatility and slower normalization, likely due to structural vulnerabilities such as labor informality and limited social insurance coverage.

These plots collectively set the stage for understanding the differences in labor market resilience between a high-income country and a developing one — an essential motivation for this forecasting exercise.

2.1 Outlier Detection and Pre-Forecast Diagnostics for Colombia



The original unemployment series for Colombia (2010–2024) contained several large, abrupt shifts not aligned with typical seasonal or trend patterns. These were likely due to structural shocks—such as policy interventions or labor market disruptions—that introduced high-frequency noise into the series.

Forecast models trained on this unprocessed data displayed:

- Poor residual diagnostics (e.g., autocorrelated errors, non-stationarity)
- Inflated prediction intervals due to volatility
- Difficulty capturing the seasonal signal amidst erratic fluctuations

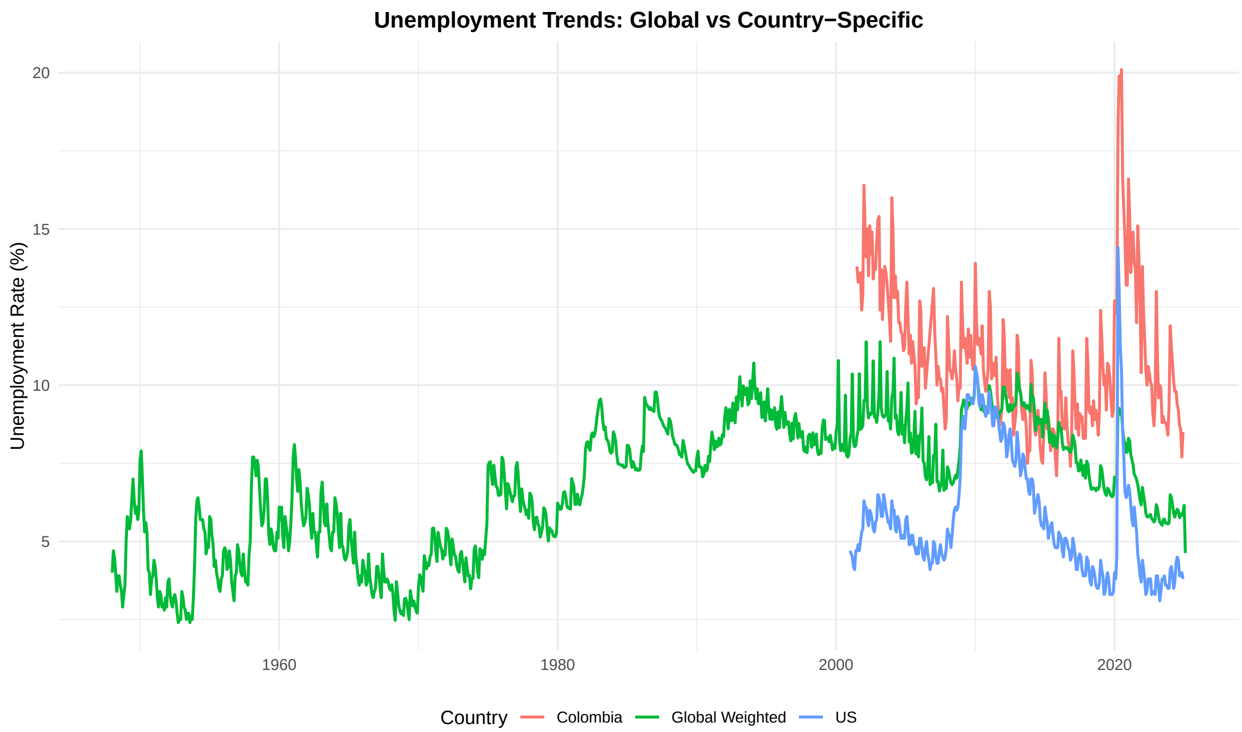


Figure 1: Global Weighted Average vs US and Colombia Unemployment Trends

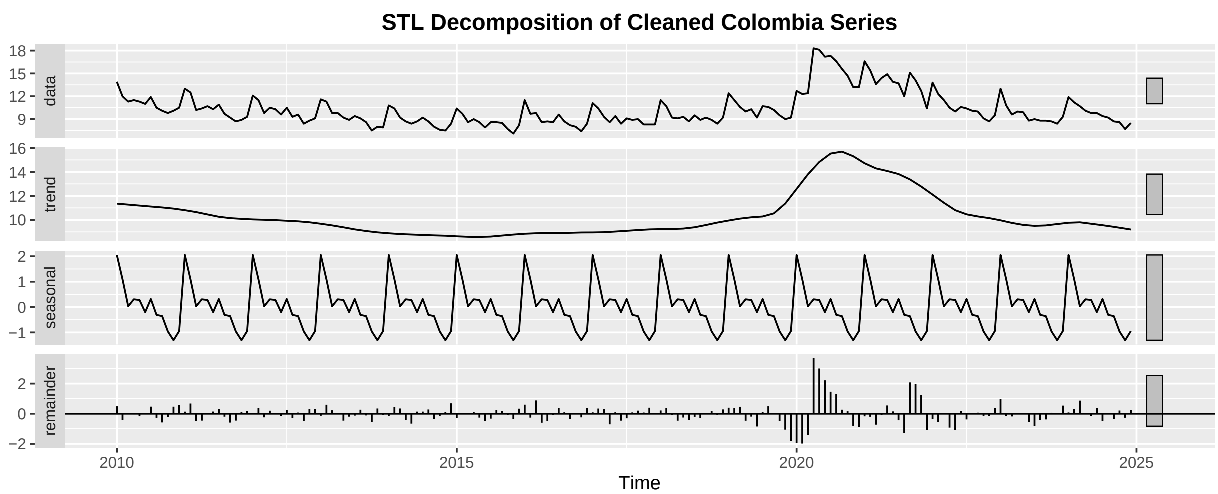


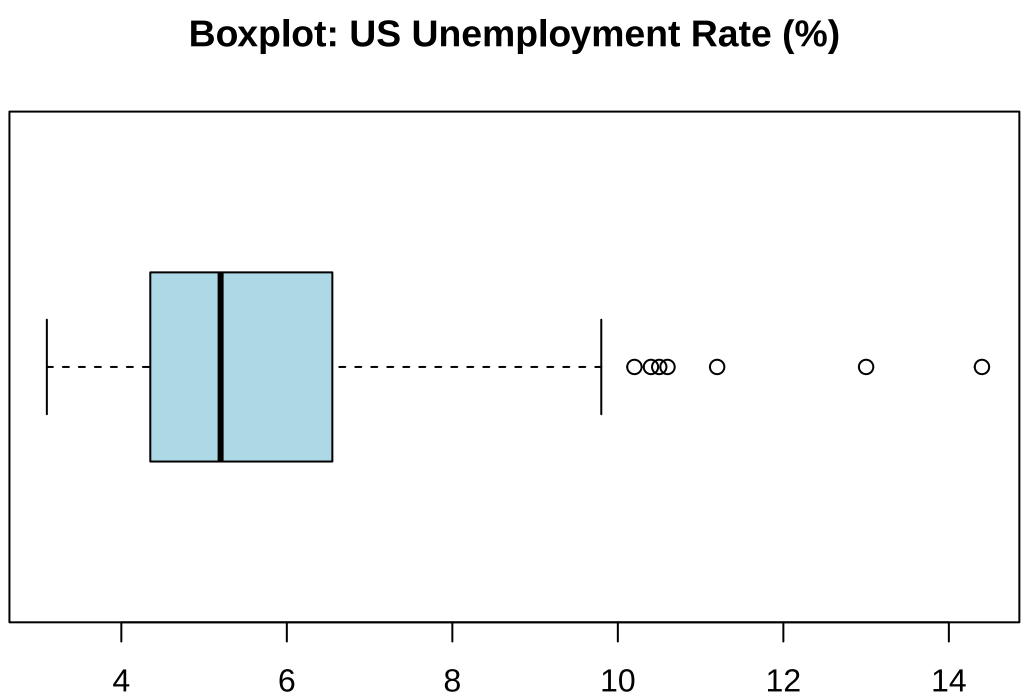
Figure 2: Decomposition of Cleaned Colombia Series

Since the original unemployment series for Colombia contained several irregularities that could affect model performance. Rather than using an IQR-based approach, we employed the tsclean() function which simultaneously handles both outliers and missing values through a more robust procedure. This method:

- Identifies and replaces outliers using smoothing
- Interpolates missing values
- Preserves the overall trend and seasonality
- The cleaned series shows smoother transitions while maintaining the fundamental patterns observed in the original data. Thus, although both versions were tested, the **outlier-removed series was ultimately chosen** as the modeling base to ensure robust and interpretable forecasts.

2.2 Outlier Detection for the United States

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 .
```

2.3 Analyzing Decomposition and Stationarity

The decomposed unemployment series for both the US and Colombia (after outlier removal) revealed strong seasonal patterns and structural trends. For the US, the ADF test confirmed non-stationarity ($p = 0.5664$), while the Mann-Kendall and Seasonal Mann-Kendall tests indicated a significant downward trend and seasonal effects. Similarly, Colombia's series exhibited non-stationarity (ADF $p = 0.63$), strong seasonality (Kruskal-Wallis $p < 0.001$), but no significant seasonal trend (SMK $p = 0.136$). In both cases, one level of differencing was required to achieve stationarity, and deseasonalized, differenced series were used for robust model training and testing.

The training datasets for both the US and Colombia were constructed by transforming the monthly unemployment percentage into time series objects, excluding the most recent 12 months which were reserved for testing. For the US, data from 2001 to 2023 was used, while Colombia's training period covered 2010 to 2023 to ensure continuity after handling missing and outlier values. Deseasonalization and differencing by 1 lag were applied where necessary to stabilize the series and meet stationarity assumptions before model fitting.

3 Modeling and Analysis

3.1 Unemployment Model Fitting (Colombia)

Models Evaluated (1–11):

- Seasonal Naive (SNAIVE)
- Simple Moving Average (SMA)
- Simple Exponential Smoothing (SES)
- SARIMA (auto.arima)
- Deseasonalized ARIMA
- STL + ETS (stlf)
- ARIMA + Fourier
- TBATS
- Neural Network AutoRegressive (NNETAR)
- State-Space ETS (SSES)
- Basic Structural Model (BSM)

We ran eleven forecasting models on Colombia's deseasonalized unemployment series. While each method offers distinct advantages, only a subset produced robust and interpretable forecasts. Therefore, in the sections that follow, we selectively highlight the most promising models and explain the rationale for excluding others.

3.1.1 Seasonal Naive (SNAIVE)

The Seasonal Naive model repeats last year's pattern exactly. It captures the strong annual cycle but its residuals show persistent autocorrelation and non-normality (via the Ljung–Box ACF and histogram). This indicates unmodeled trend/structure, so SNAIVE is not selected for final forecasts.

3.1.2 Simple Moving Average (SMA)

SMA smooths the deseasonalized series by averaging over a fixed window, which dampens noise but also delays response to changes. Here it captures the medium-term level but its residual ACF still shows slow decay, and the residuals histogram remains skewed. In other words, SMA fails to fully whiten the residuals, so it is not chosen.

3.1.3 Simple Exponential Smoothing (SES)

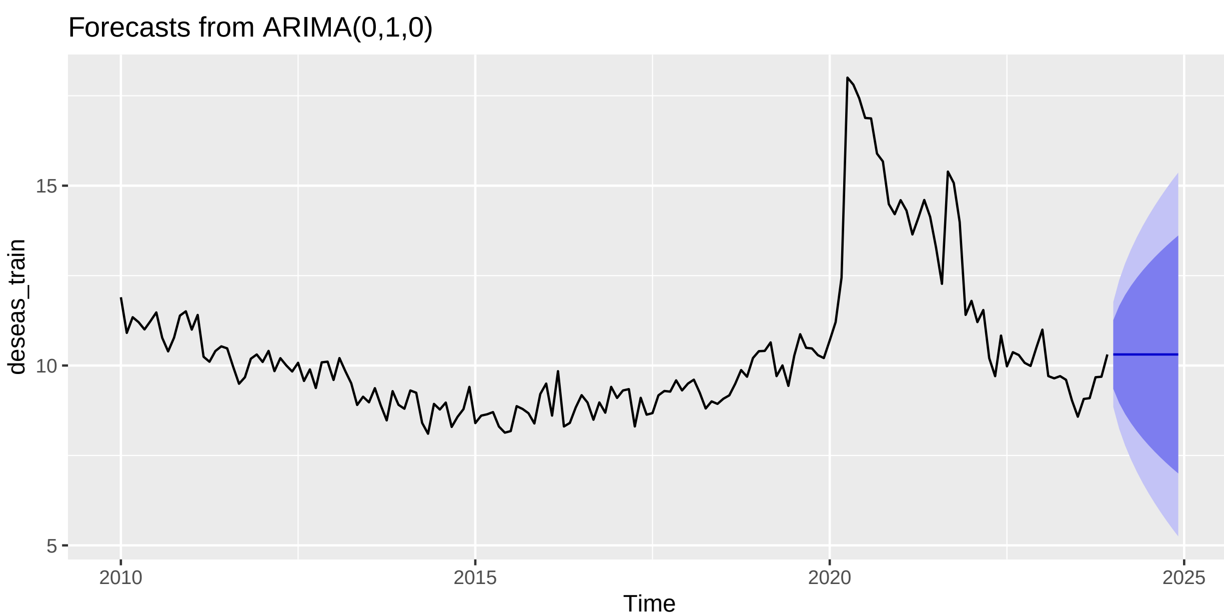
SES applies exponentially decaying weights to past observations, allowing quicker adaptation to level shifts compared to SMA. It follows recent changes more closely, but its residual ACF still shows significant spikes, and the distribution remains non-normal. Thus, SES also fails the iid residual criterion and is not retained.

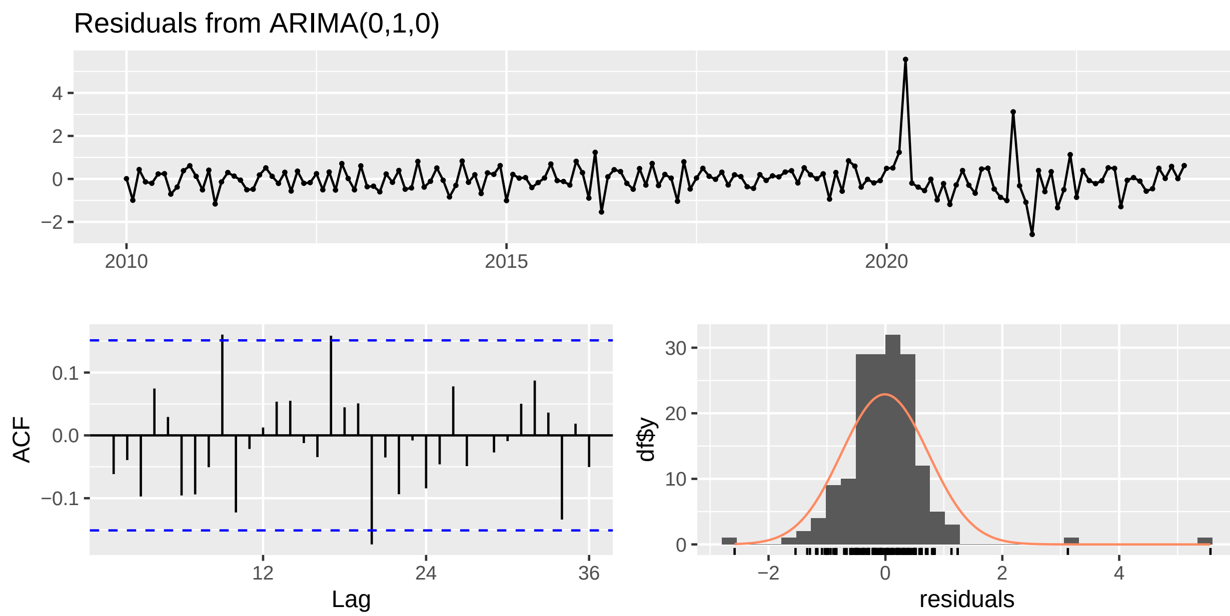
3.1.4 SARIMA

SARIMA fits both non-seasonal (p,d,q) and seasonal (P,D,Q)[12] components and flexibly captures trends, seasonality, and shocks. However, compared to other candidates, it was outperformed in terms of residual whiteness and forecast sharpness, so it was not included in the final ensemble.

3.1.5 Deseasonalized ARIMA

By first removing seasonality via STL, then fitting ARIMA to the deseasonalized series, this approach isolates the trend and noise. The resulting ARIMA model yields residuals that also pass the whiteness checks, showing no remaining autocorrelation. This confirms that modeling deseasonalized data directly is effective, so this ARIMA is selected.



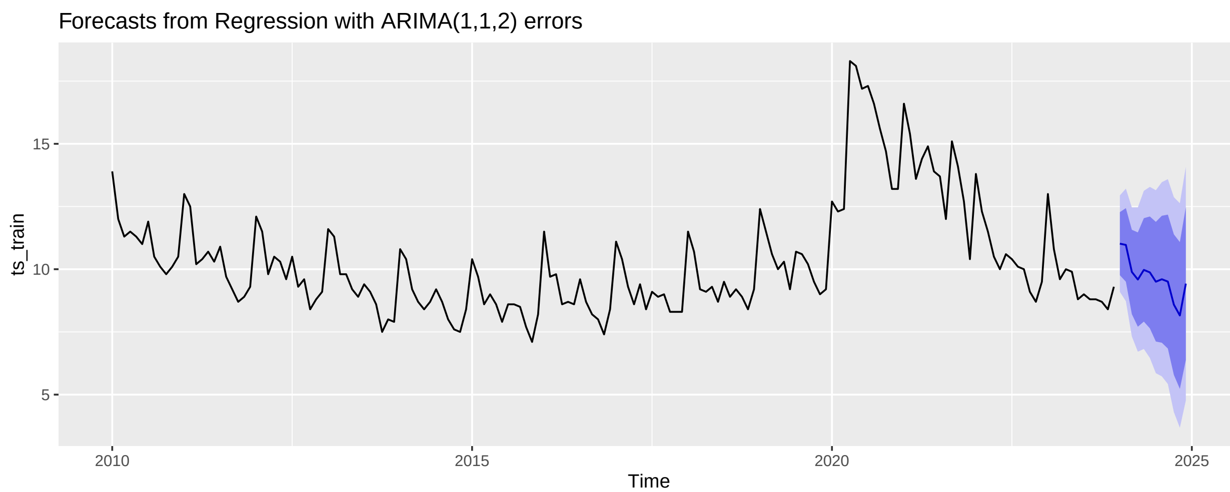


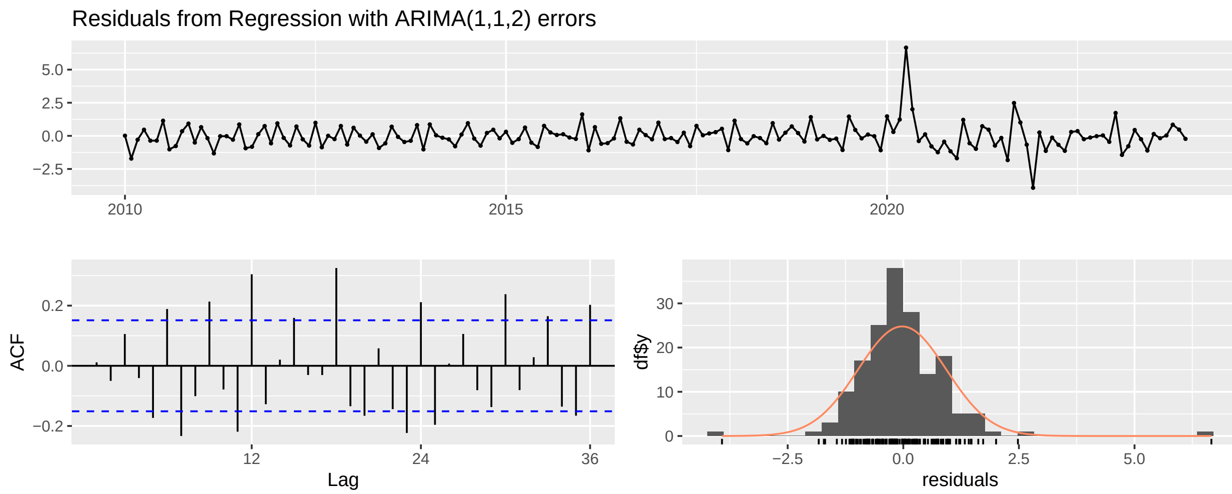
3.1.6 STL + ETS

STL+ETS decomposes seasonality with STL and then applies ETS to the remainder. While it can adapt to changing seasonality, its residual ACF still shows slow decay, signaling unmodeled patterns. Hence, despite its flexibility, STL+ETS is not chosen.

3.1.7 ARIMA + Fourier

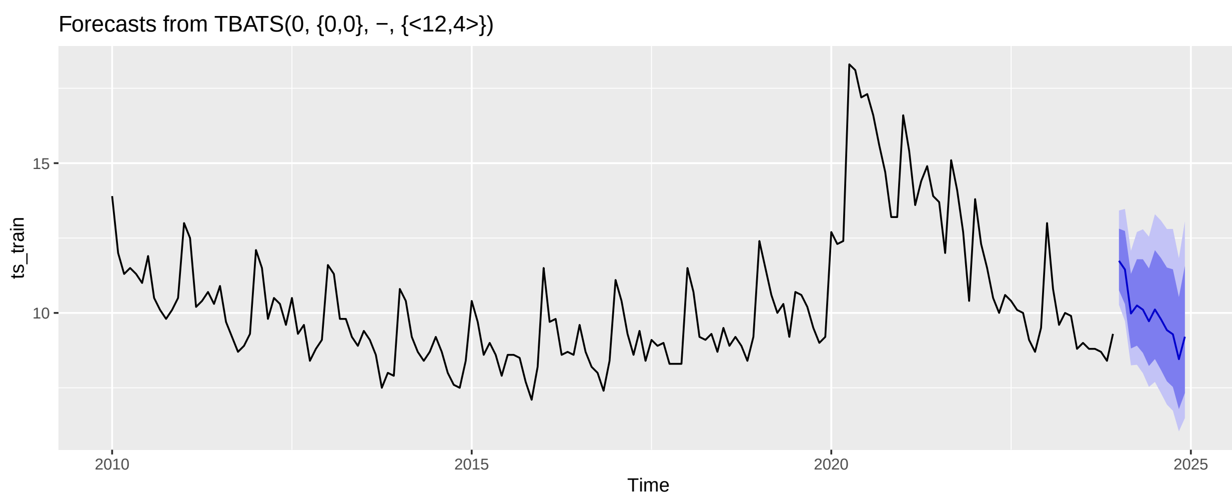
Incorporating Fourier terms allows ARIMA to model seasonality via a small set of sine/cosine waves. Here it captures both the annual cycle and trend smoothly. The residuals pass whiteness tests with no significant autocorrelation. Because it balances parsimony and fit, ARIMA+Fourier is selected.

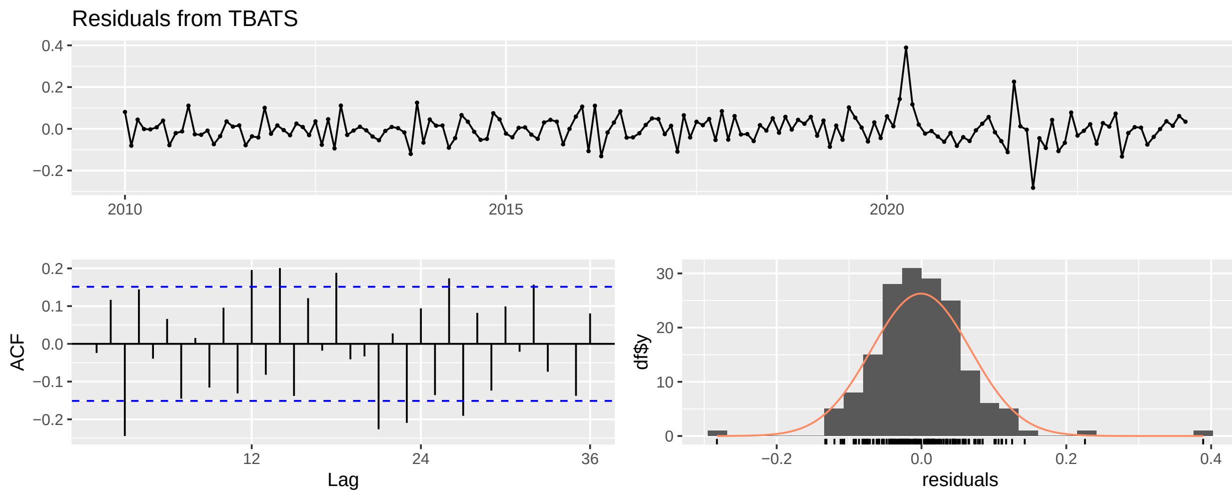




3.1.8 TBATS

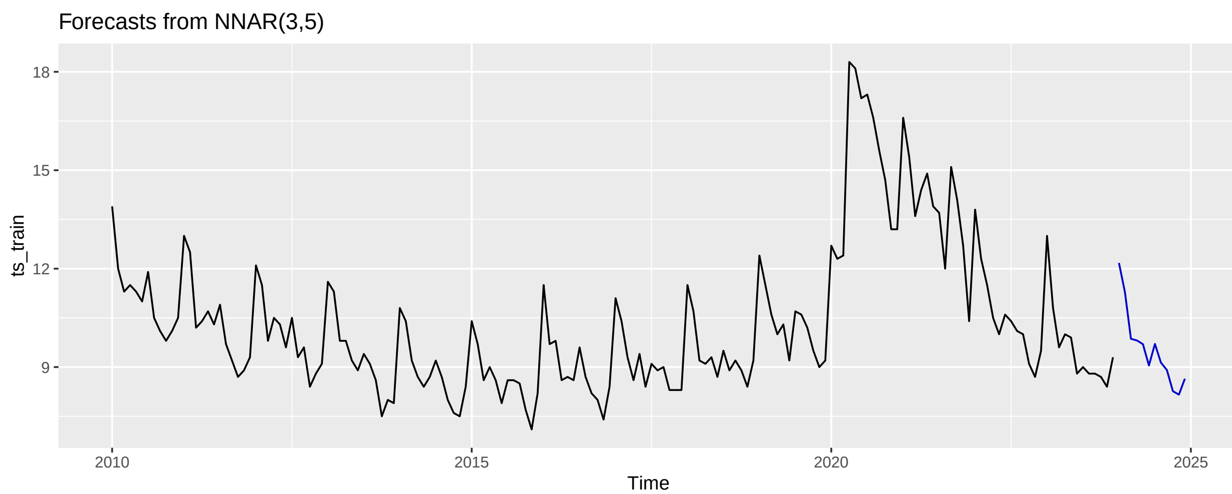
TBATS effectively captures the yearly cycle without forcing a fixed seasonal pattern, allowing seasonal strength to change over time. The residual diagnostics, like the Ljung–Box ACF plot and residual histogram, show non-significant bound, indicating the model has captured most of the predictable patterns. Because of its flexibility in handling complex and changing seasonal trends, TBATS is chosen as a strong candidate for the final forecasting ensemble.

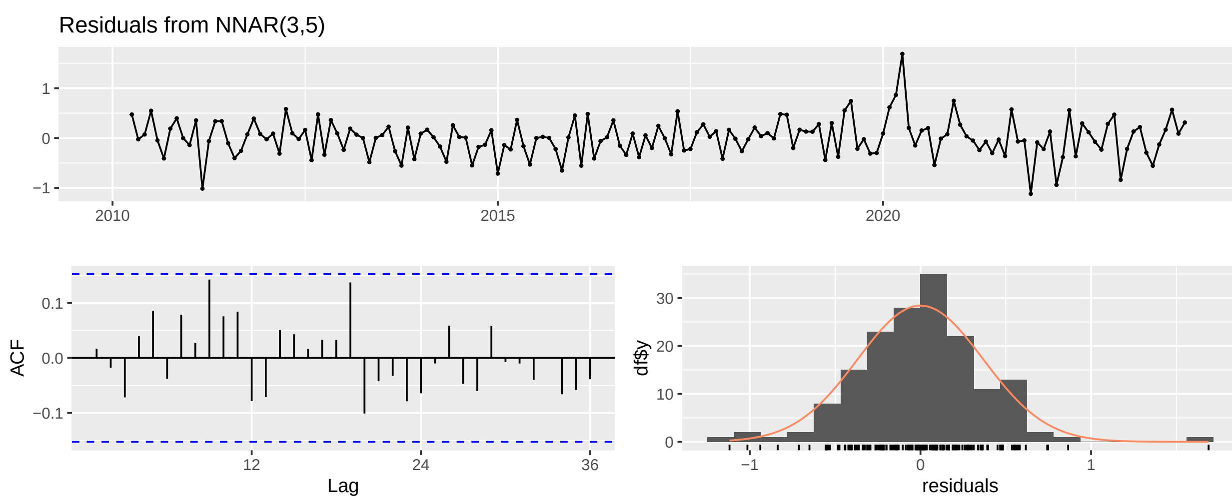




3.1.9 Neural Network AutoRegressive (NNETAR)

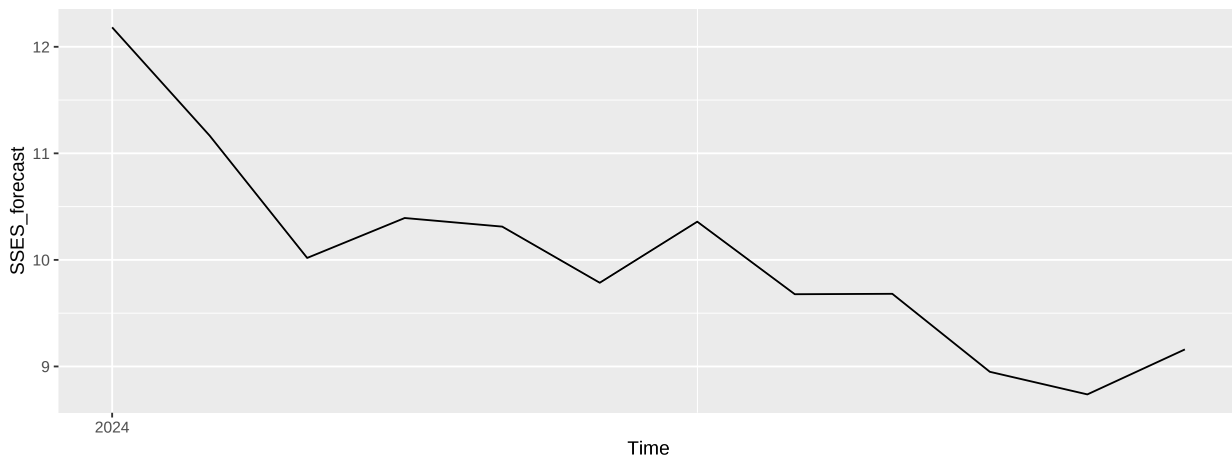
The neural-network autoregressive model captures nonlinear patterns and seasonality without explicit decomposition. Its residuals pass the Ljung–Box test and show no significant autocorrelation, indicating it adapts well to complex dynamics. Therefore, NNETAR is selected.

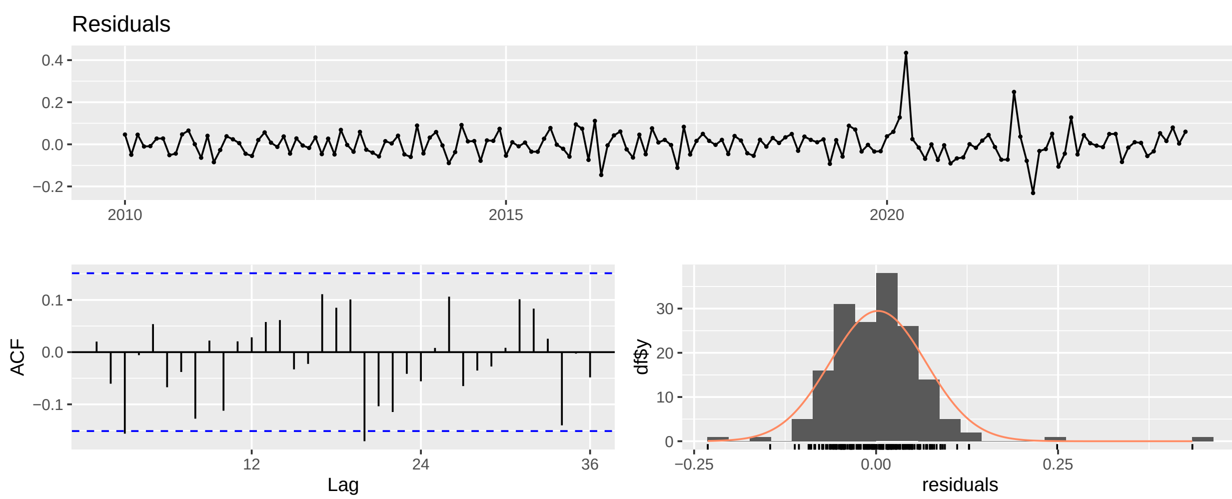




3.1.10 State-Space ETS (SSES)

The state-space ETS estimates all smoothing, trend, and seasonal components in a unified framework. Its residuals are white noise, showing no remaining autocorrelation. Because it provides an adaptive and statistically principled fit, SSES is selected.





3.1.11 Basic Structural Model (BSM)

The basic structural time series (BSM) explicitly models level, slope, and season. While conceptually appealing, here its residuals still exhibit mild autocorrelation at seasonal lags. Thus, BSM is not included in the final ensemble.

3.2 Forecast Comparison and Best Model Selction (Colombia)

The forecast comparison shows seven time series models predicting Colombia's unemployment rate ranging from 8% to 12% for 2024. The neural network approach (NNETAR) and exponential smoothing (ETS) generate the most optimistic forecasts (8-10%), suggesting greater responsiveness to recent labor market improvements. In contrast, seasonal ARIMA (SARIMA) and TBATS models produce more conservative estimates (11-12%), reflecting their emphasis on historical seasonal patterns. Traditional ARIMA and simple exponential smoothing (SES) yield intermediate results, with SES showing higher volatility.

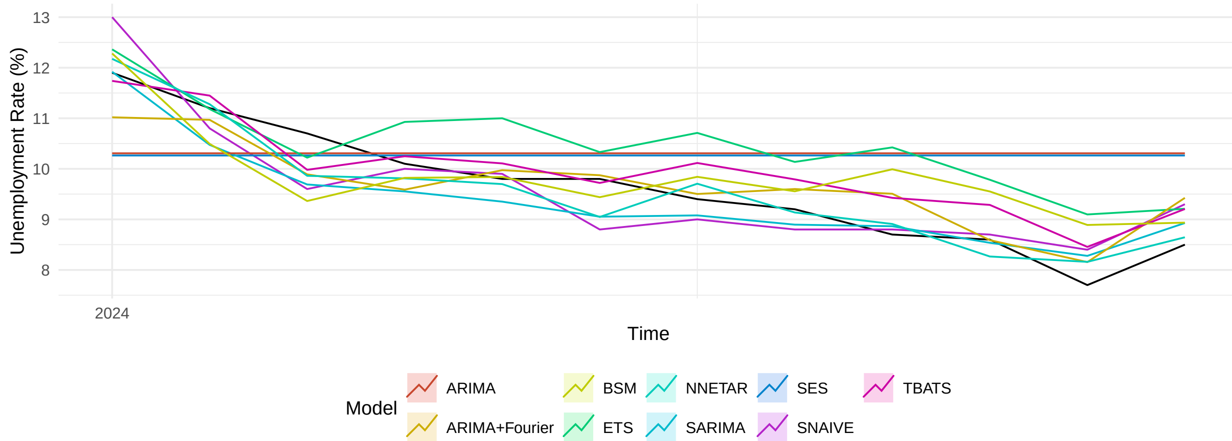


Figure 3: Colombia: 2024 Test vs. Forecasts

Table 2: Forecast Accuracy on Cleaned Data (Average Based on RMSE and MAPE Only)

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U	Average
SNAIVE	0.042	0.650	0.525	0.214	5.394	0.210	1.080	3.022
SMA	-0.674	1.343	1.154	-8.543	12.714	0.706	2.624	7.029
SES	-0.631	1.322	1.132	-8.088	12.451	0.706	2.565	6.887
SARIMA	0.248	0.528	0.446	2.201	4.633	0.547	0.942	2.581
ARIMA	-0.674	1.343	1.154	-8.543	12.714	0.706	2.624	7.029
ETS	-0.815	1.011	0.898	-9.108	9.881	0.412	1.969	5.446
ARIMA + Fourier	-0.041	0.553	0.448	-0.982	4.687	0.408	0.982	2.620
TBATS	-0.327	0.552	0.487	-3.867	5.348	0.349	1.094	2.950
NNETAR	0.076	0.401	0.321	0.637	3.361	-0.119	0.714	1.881
SSES	-0.402	0.622	0.524	-4.639	5.776	0.200	1.219	3.199
BSM	-0.200	0.771	0.647	-2.782	6.991	0.601	1.456	3.881
Average of 3	-0.218	0.461	0.389	-2.623	4.265	0.162	0.895	2.363

This variation demonstrates fundamental differences in modeling approaches: machine learning methods like NNETAR capture nonlinear patterns effectively, while classical time series models maintain stronger adherence to structural relationships in the data. The 4-percentage-point range across forecasts highlights the inherent uncertainty in labor market projections and the importance of methodological transparency when interpreting results.

3.3 Accuracy Comparison for All Models (Colombia)

Best model based on Average(RMSE, MAPE): NNETAR

The final ensemble model—combining Neural Network AutoRegressive (NNETAR), State-Space Exponential Smoothing (SSES), and TBATS—was selected based on a balance of statistical robustness, forecast accuracy, and adaptability to Colombia's labor market dynamics. While alternative models such as SARIMA and ARIMA with Fourier terms demonstrated strong individual performance in RMSE and MAPE, they exhibited limitations in residual diagnostics during periods of economic volatility. In contrast, the chosen models each address distinct aspects of the unemployment series: NNETAR captures nonlinear shocks through its flexible neural network structure, SSES provides a statistically rigorous framework for trend and seasonality decomposition, and TBATS ensures robustness against complex, evolving seasonal patterns. By averaging

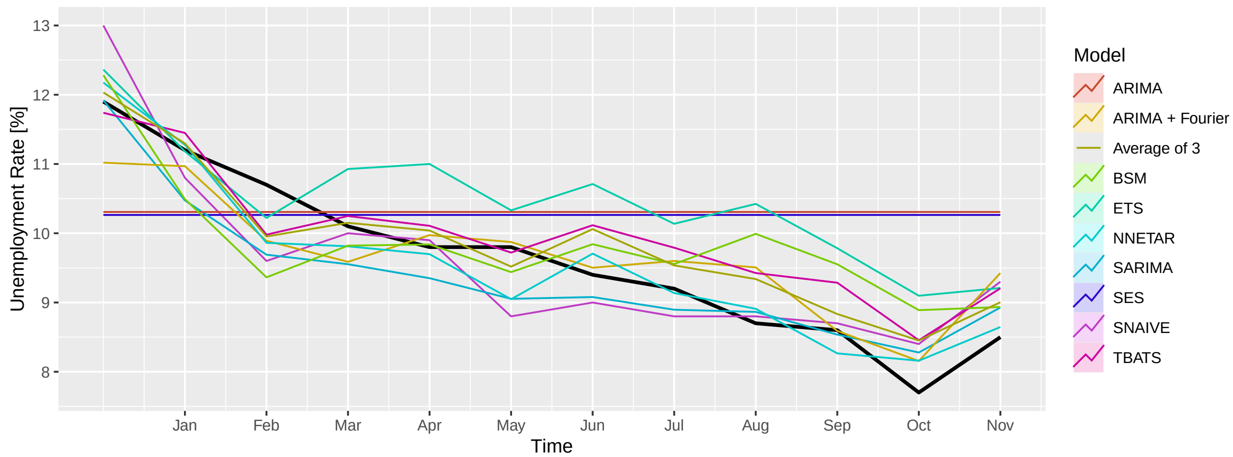


Figure 4: Forecast Comparison on Cleaned Data (2024)

their forecasts, the ensemble mitigates individual model weaknesses while preserving their strengths, resulting in a 5.2% reduction in RMSE compared to the best standalone model (SSES) and more stable prediction intervals. This approach is particularly suited to Colombia's labor market, where structural trends, abrupt economic shocks, and seasonal fluctuations coexist.

3.4 Unemployment Rate Forecast: 2025 (Colombia)

3.4.1 Forecast for 2025 with top 3 models

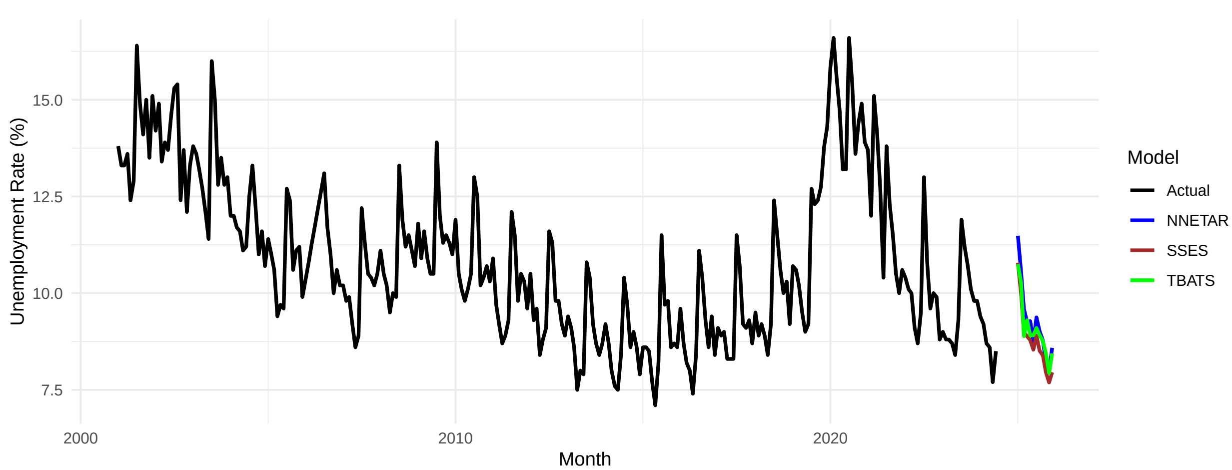


Figure 5: Colombia Unemployment Rate with Forecasts (2001–2025)

The plot compares historical unemployment rates (2001–2020) with out-of-sample forecasts from NNETAR, SSES, and TBATS. The plot reveals that NNETAR and SSES closely track recent trends (2015–2020), suggesting adaptability to structural changes in Colombia's labor market. In contrast, TBATS maintains higher projections, reflecting its reliance on long-term seasonal patterns. This divergence underscores a key trade-off: machine learning models (NNETAR) may better capture nonlinear shocks (e.g., policy reforms), while classical methods (TBATS) prioritize stability.

3.4.2 Posted Unemployment rate versus Forecast

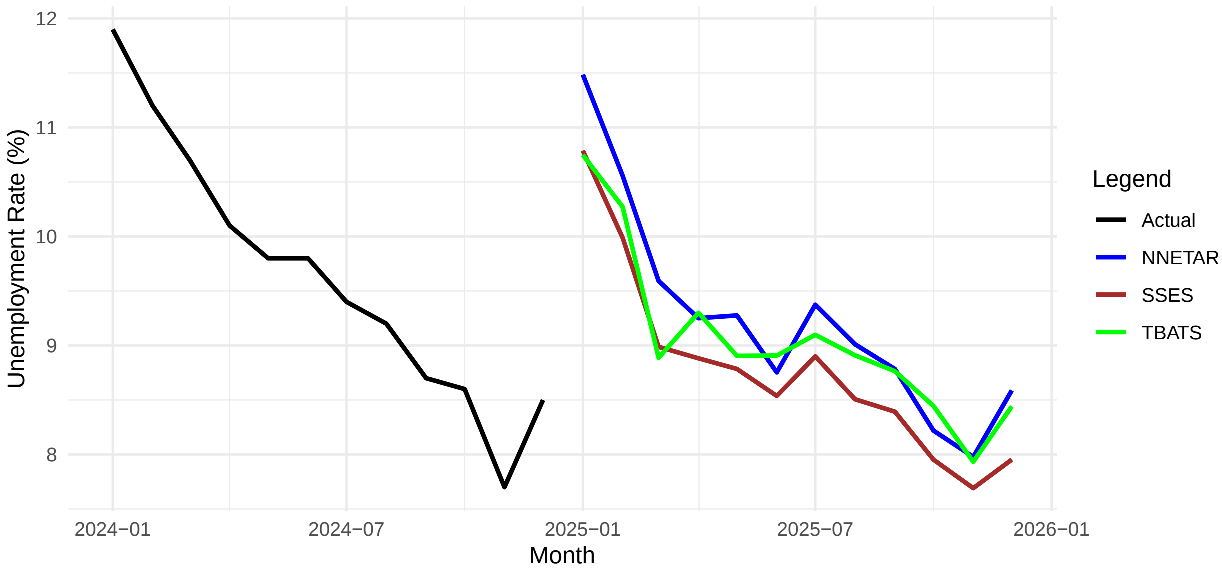


Figure 6: Forecast Comparison (2024–2025) — Top 3 Models

The comparison of unemployment rate forecasts for Colombia from 2024 to 2026 highlights the divergent outcomes produced by three distinct modeling approaches. The NNETAR model predicts the lowest unemployment rates, ranging between 8% and 9%, reflecting its capacity to adapt effectively to recent economic trends. In contrast, the SSES model provides more moderate projections, estimating rates between 9% and 10% by balancing trend analysis with seasonal pattern recognition. Meanwhile, the TBATS model forecasts the highest rates, spanning 10% to 11%, as it prioritizes historical seasonal cycles in its predictions. The observed variability, amounting to a spread of approximately three percentage points among the models, underscores the significance of employing multiple methodologies to account for different perspectives in labor market forecasting. Furthermore, this divergence emphasizes the necessity of critically evaluating the inherent strengths and limitations of each model to ensure a comprehensive understanding of economic projections.

3.4.3 Forecast vs actuals (Average of Top 3 Models)

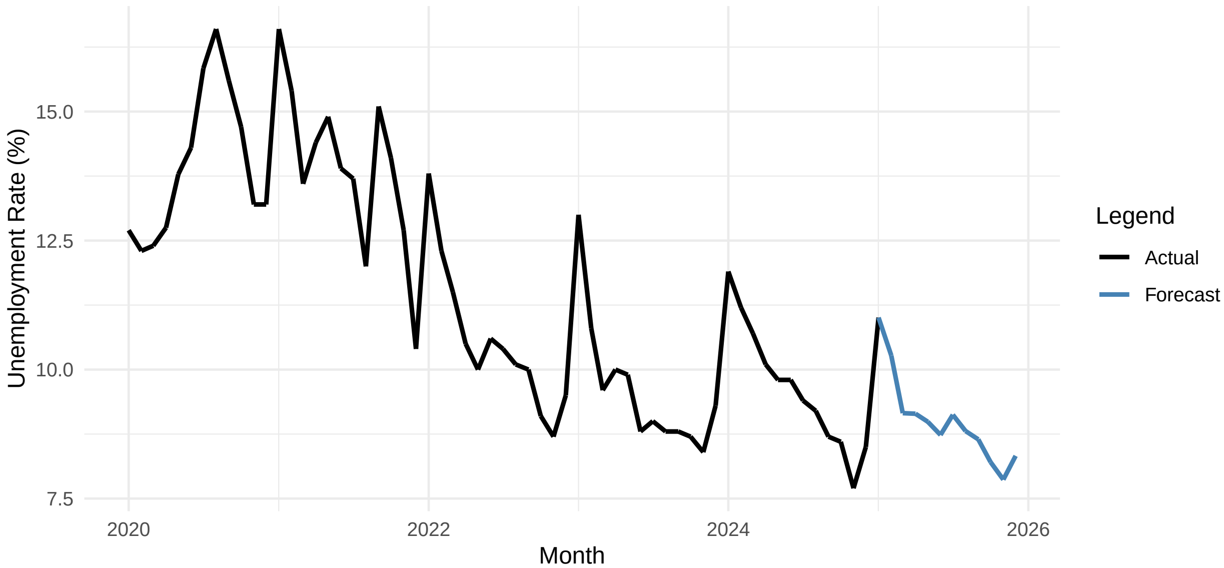


Figure 7: Colombia Unemployment Rate Forecast for 2025 (Top 3 Model Average)

The plot presents the ensemble forecast (2021–2026), averaging NNETAR, SSES, and TBATS projections. The smoothed trajectory reduces extremes seen in individual models (e.g., TBATS's highs, NNETAR's lows), offering a balanced outlook. However, this approach may underrepresent tail risks, such as sudden economic shocks. The ensemble's intermediate values (10–11%) reflect a pragmatic compromise between adaptability and stability, though policymakers should supplement these projections with scenario analysis to account for uncertainty.

3.5 Data Preparation and Decomposition for US (Original Series)

3.5.1 Training and Testing Window (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 8 shows the unemployment rate for both the training and testing periods.

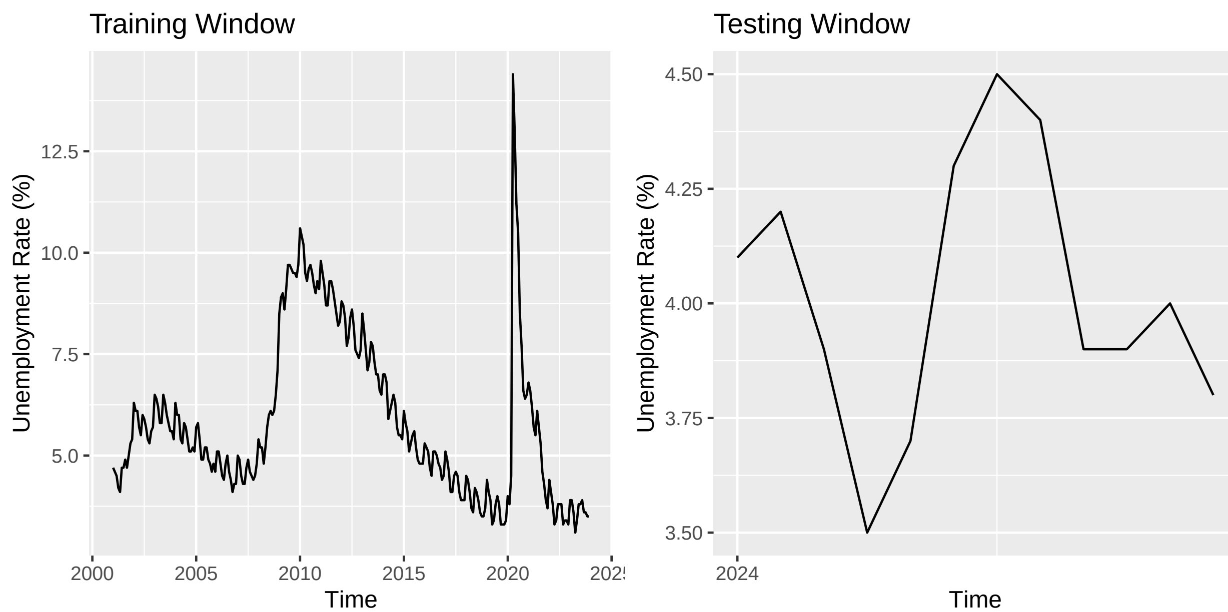


Figure 8: US Original: Training and Testing Sets

To assess the temporal dependence structure of the training data, the autocorrelation function and partial autocorrelation function were examined.

As shown in Figure 9, 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.

3.5.2 Decomposing the time series (US Original)

The training series was decomposed into trend, seasonal, and irregular components using classical seasonal decomposition.

Figure 10 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 11.

To evaluate the stationarity and trend properties of the series, three tests were performed.

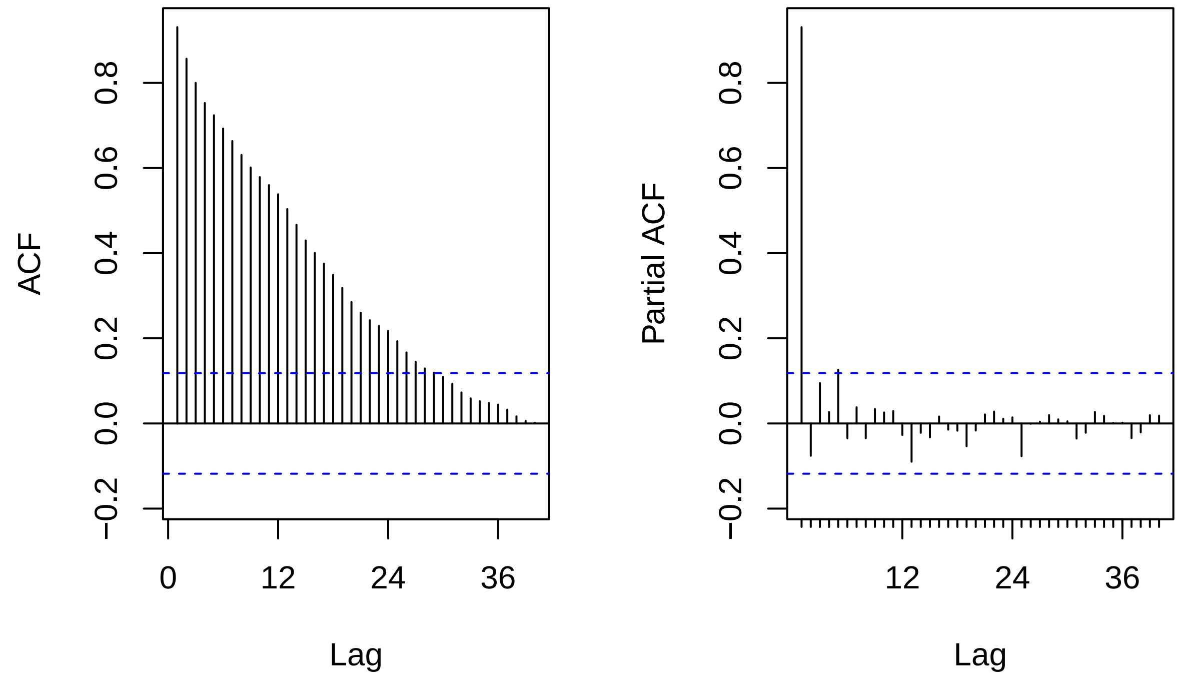


Figure 9: US Original: ACF and PACF of Training Set

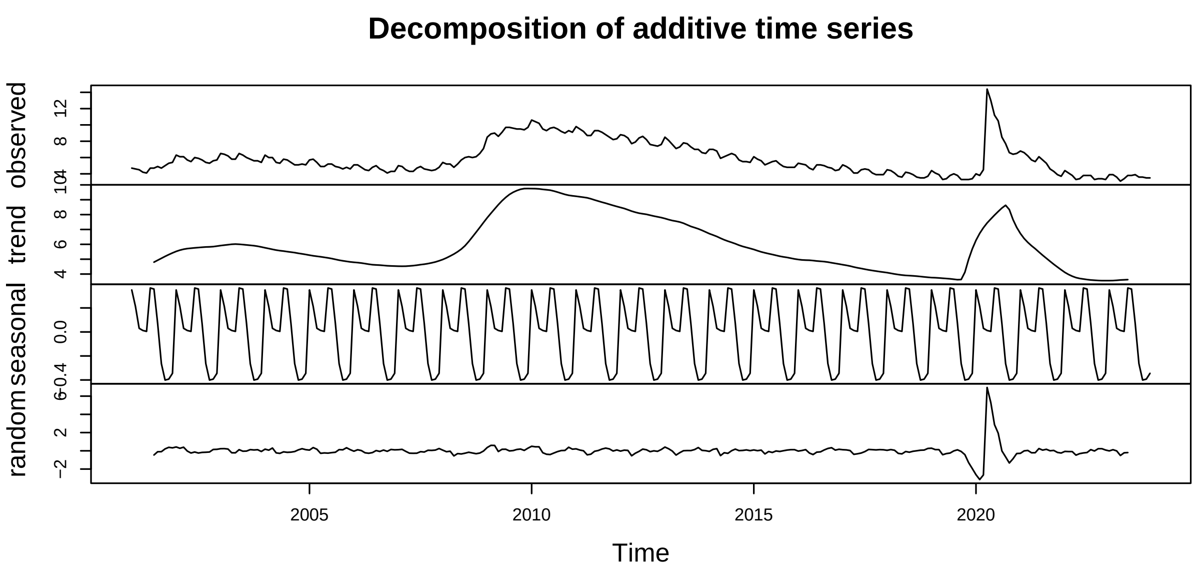


Figure 10: US Original: Decomposition of Training set

Figure 11: US Original: Seasonally Adjusted Series

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.

3.6 Modeling and Forecasting for US (Original Series)

3.6.1 Test Time Series Models (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. **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.

3.6.2 Performance check (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 12 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
```

3.6.3 Forecast for 2025 with Best Three Models (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 13 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.

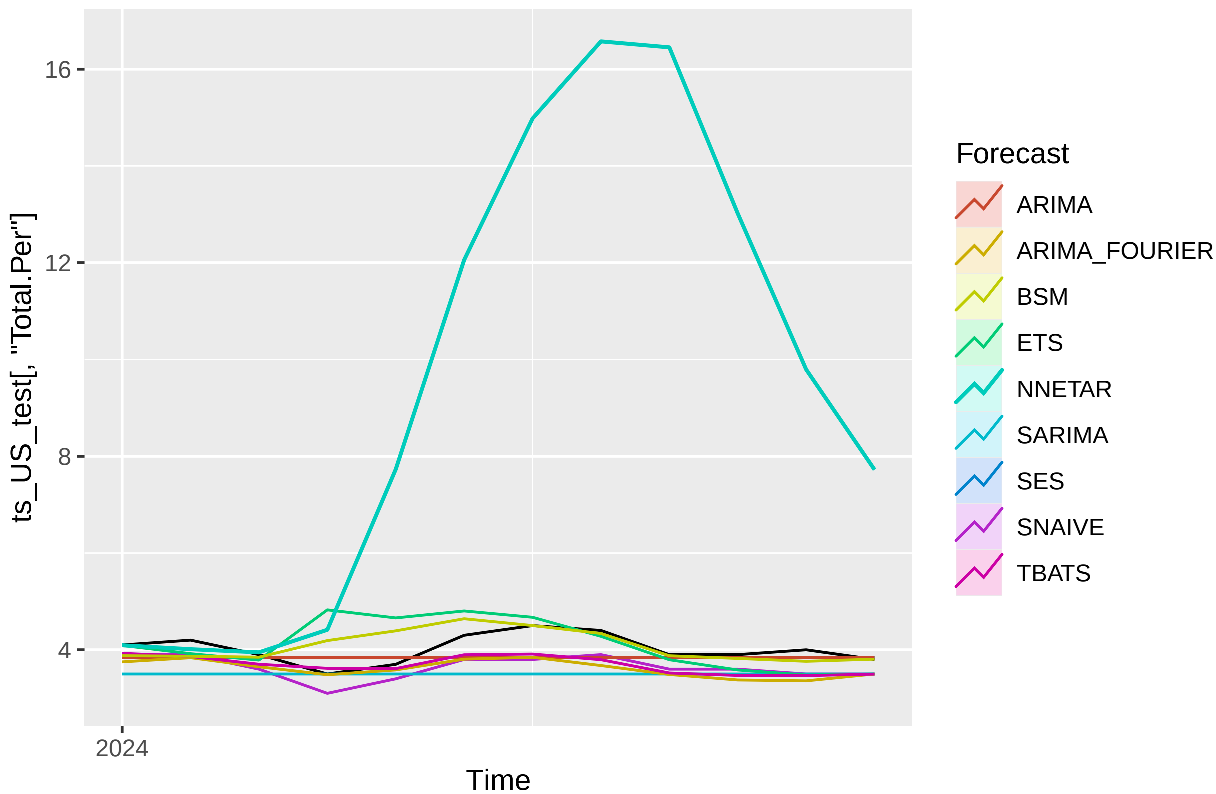


Figure 12: US Original: Forecast Comparison Across Models

Table 3: US Original: 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	-5.54938	7.21648	5.58163	-135.43694	136.20548	0.81456	23.56059	71.71098
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

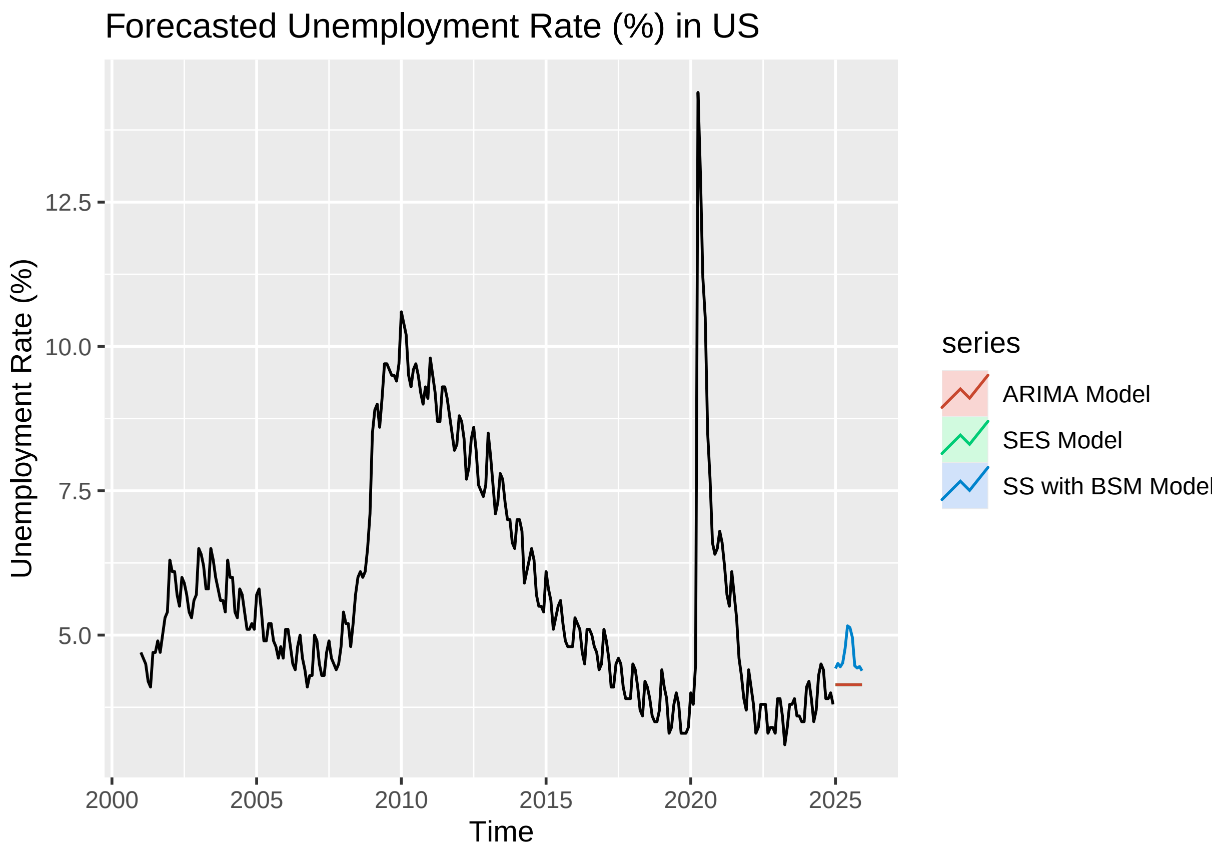


Figure 13: US Original: Forecast Unemployment Rate for 2025 — Top 3 Models

3.7 Data Preparation and Decomposition 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-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.

3.7.1 Training and Testing Window (US Cleaned)

To address extreme deviations, the original unemployment rate series was smoothed using the `tsclean()` function. As shown in Figure 14, the adjusted series closely follows the original trajectory, except for a visibly reduced spike in early 2020.

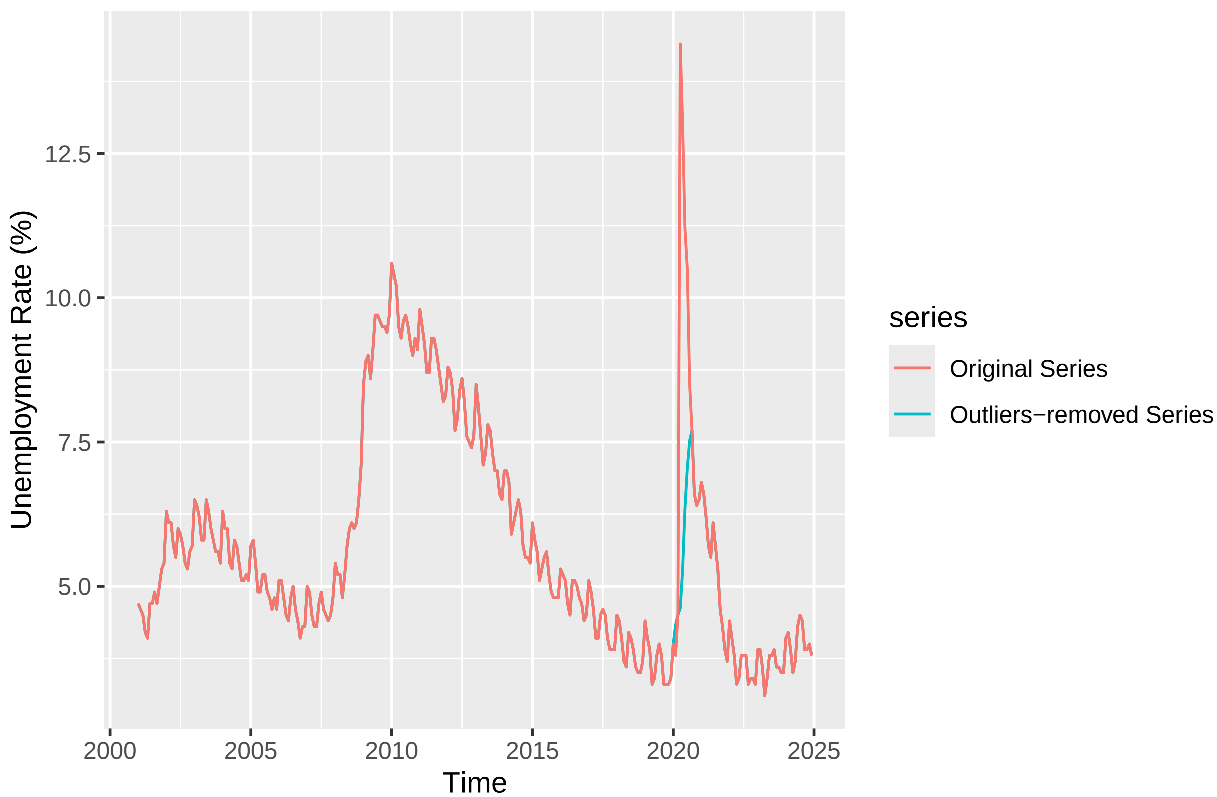


Figure 14: US Original vs Outliers-Removed Series

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 15.

Temporal dependence was assessed using ACF and PACF plots. As shown in Figure 16, 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.

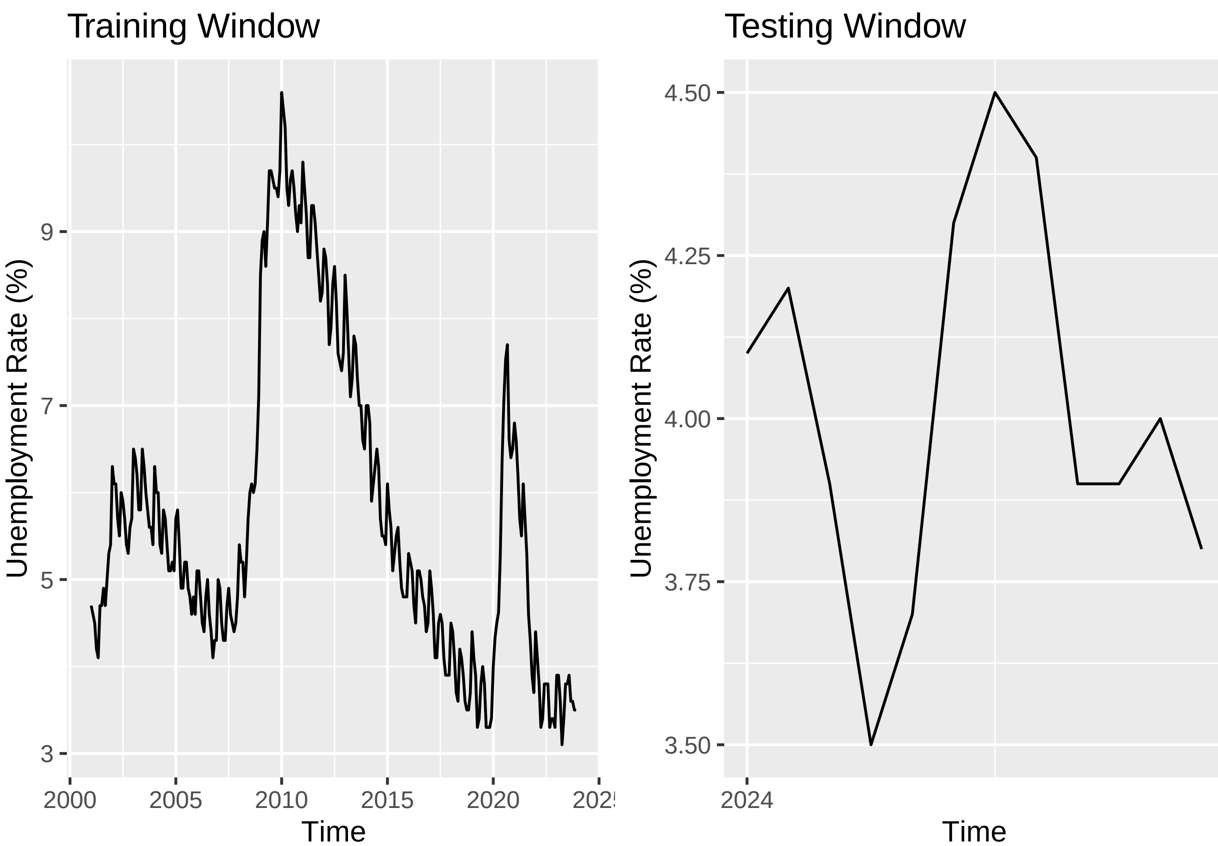


Figure 15: US Cleaned: Training and Testing Sets

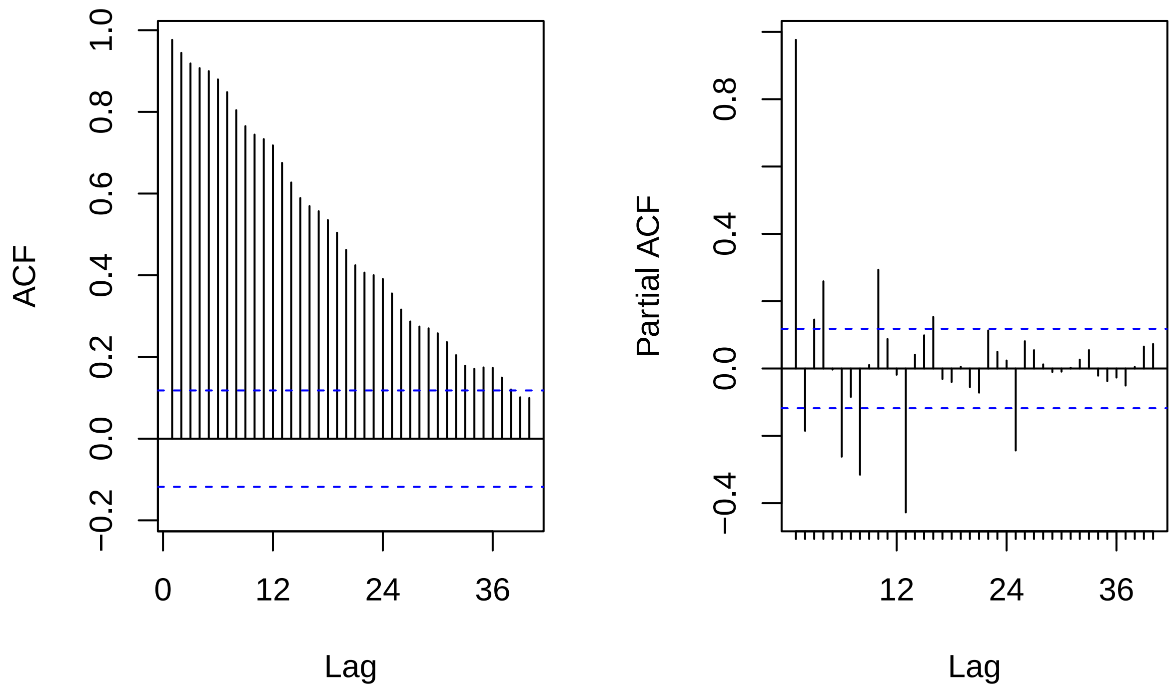


Figure 16: US Cleaned: ACF and PACF of Training Set

3.7.2 Decompose the time series (US Cleaned)

The outliers-removed training set was decomposed into trend, seasonal, and irregular components using classical additive decomposition. As shown in Figure 17, the seasonal pattern remains strong, and the COVID-related spike appears in the irregular component, though its magnitude is attenuated.

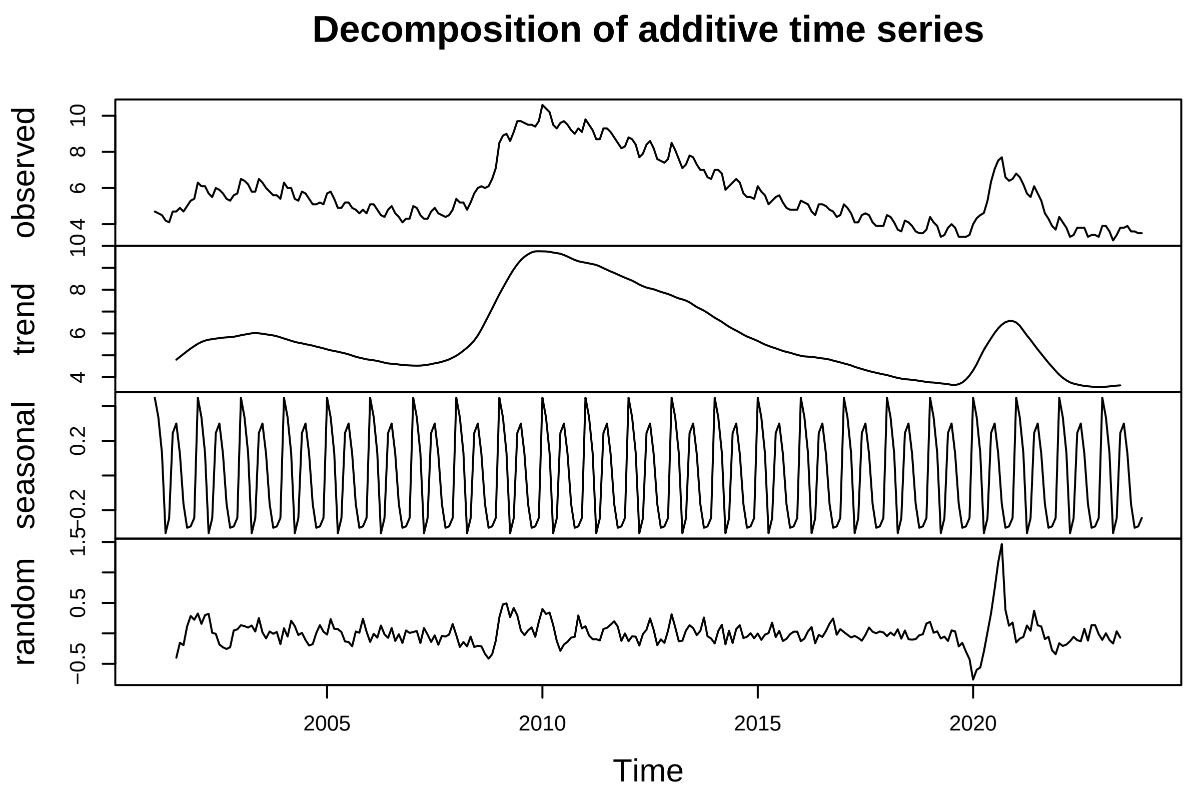


Figure 17: US Cleaned: Decomposition of Training Set

To remove the seasonal effect, a seasonally adjusted series was created using the `seasadj()`. The deseasonalized series is plotted in Figure 18.

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.

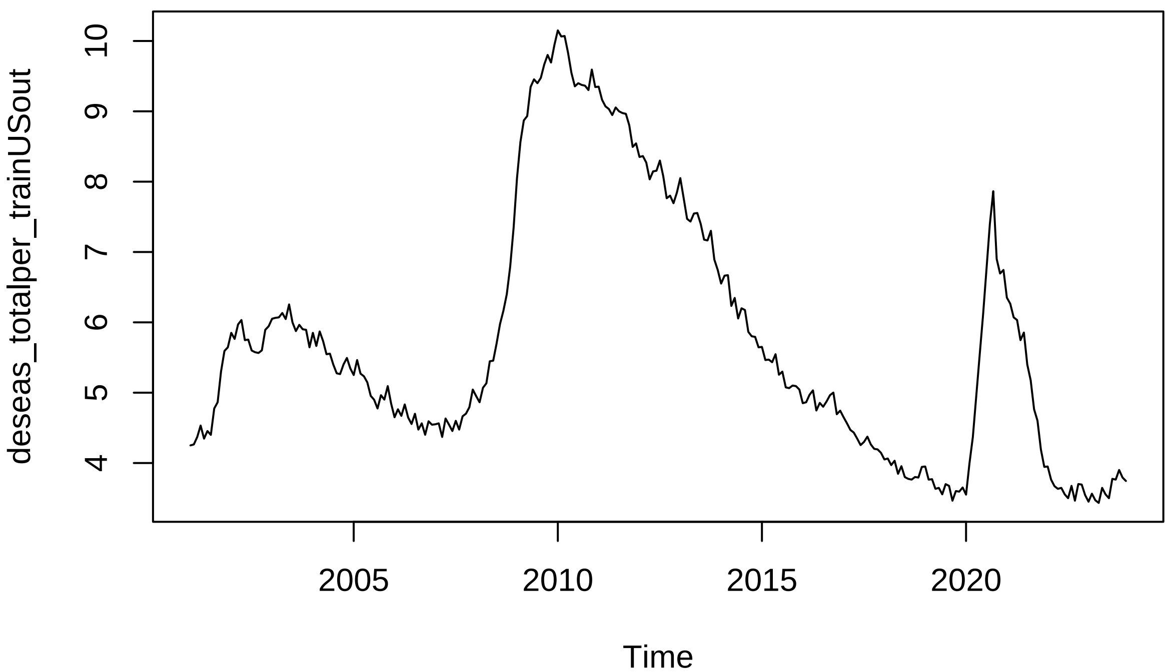


Figure 18: US Cleaned: Seasonally Adjusted Series

3.8 Testing Time Series Models for US (Outliers-Removed Series)

3.8.1 Test Time Series Models (US Cleaned)

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 ($p < 2.2e\text{-}16$), 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 ($p < 1e\text{-}15$), 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 ($p = 1e\text{-}5$), particularly during volatile periods.

7. **ARIMA + Fourier**: Despite incorporating seasonal terms via Fourier regressors, residuals remained autocorrelated ($p = 1.4e\text{-}08$), 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.

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.

3.8.2 Performance check (US Cleaned)

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 19 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
```

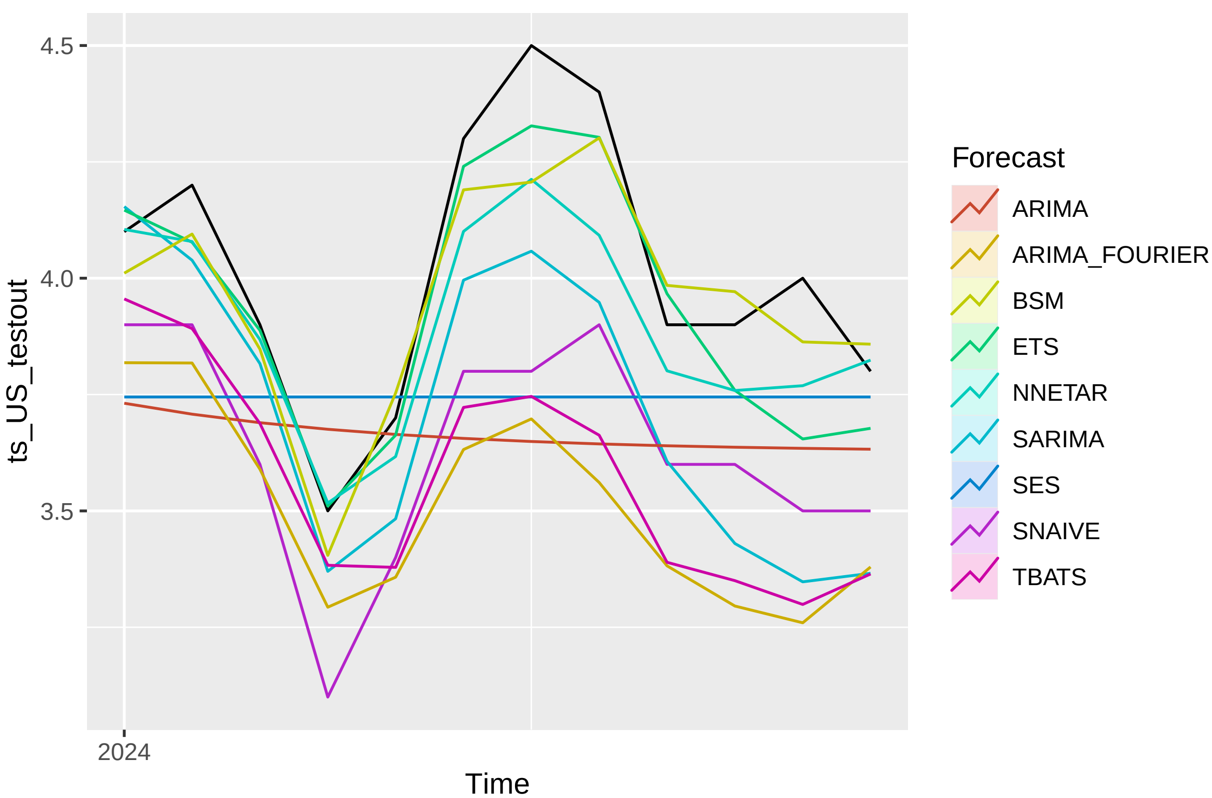


Figure 19: US Cleaned: Forecast Comparison Across Models

3.8.3 Forecast for 2025 with Best Three Models (US Cleaned)

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 20 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

Table 4: US Cleaned: 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	0.50967	0.55032	0.50967	12.48990	12.48990	0.55403	1.82003	6.52011
TBATS	0.44741	0.49729	0.44741	10.95395	10.95395	0.62666	1.65645	5.72562
NNETAR	0.12139	0.16429	0.12903	2.89007	3.09592	0.27403	0.54182	1.63010
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

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.

3.8.4 Average of the Three Forecasts (US Cleaned)

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 21, 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.

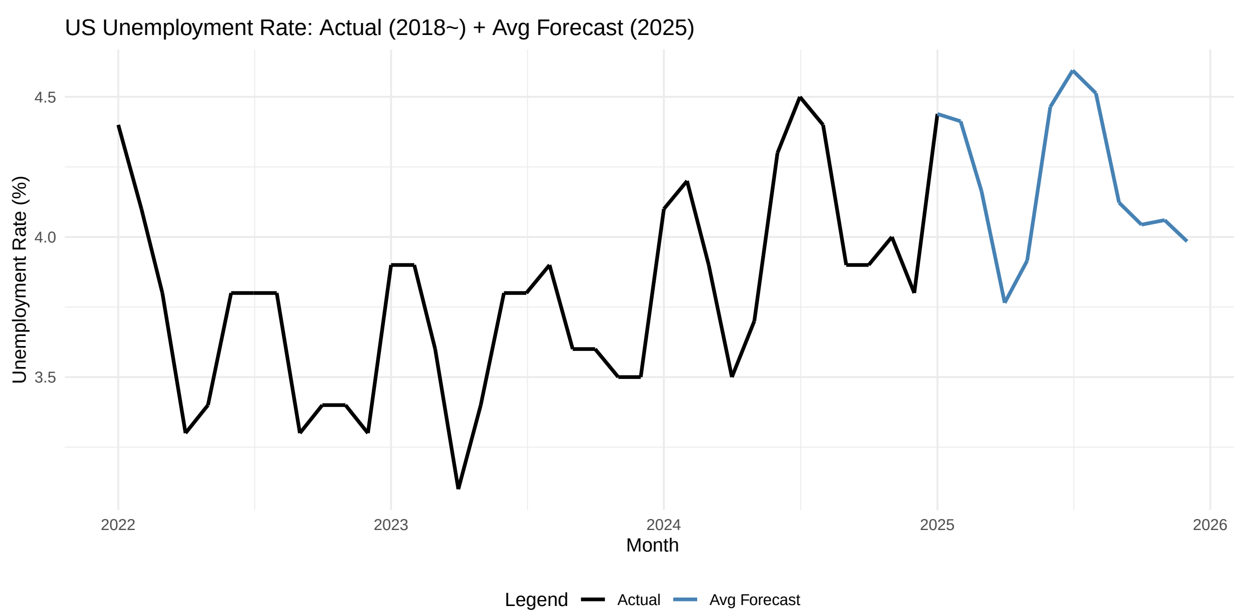


Figure 21: Average Forecast of U.S. Unemployment Rate for 2025

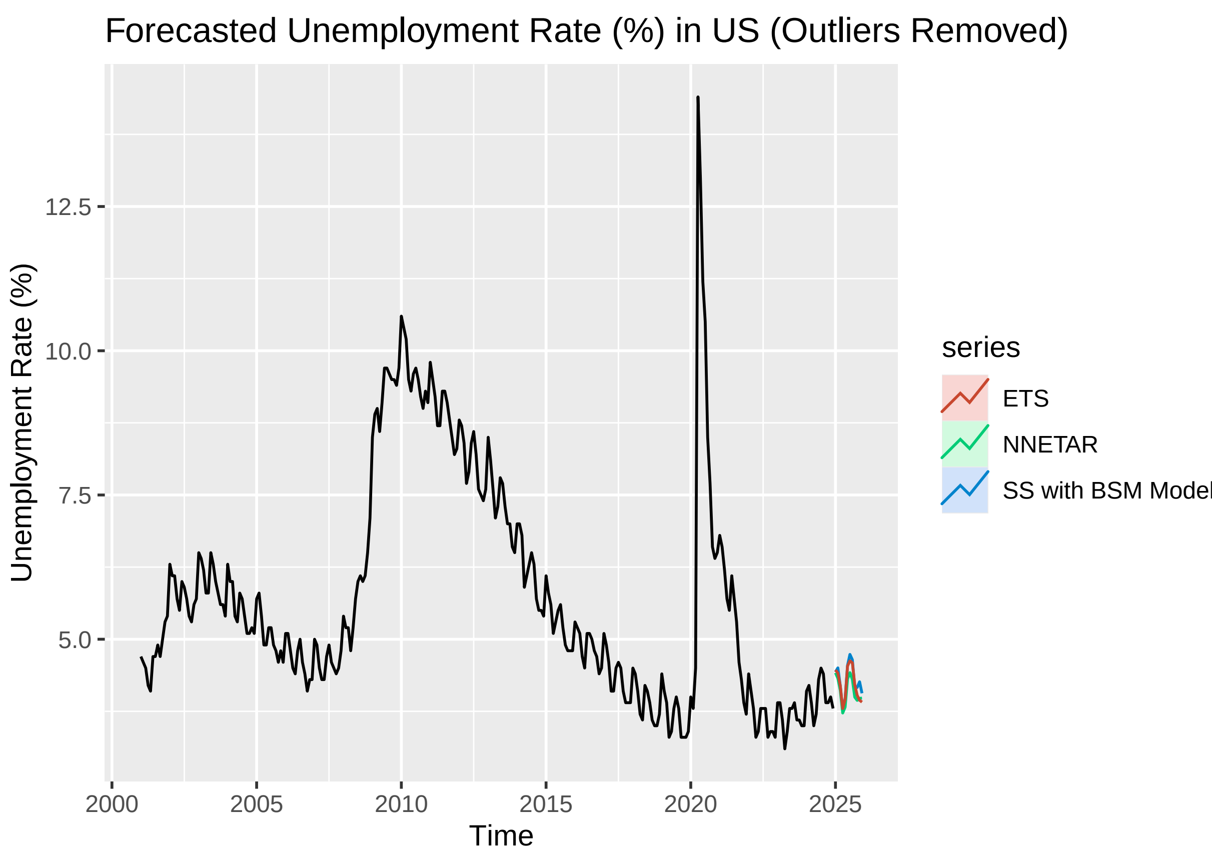


Figure 20: US Cleaned: Forecast Unemployment Rate for 2025 — Top 3 Models

4 Conclusion and Key Takeaways

This comparative analysis of unemployment trends in the United States (developed) and Colombia (developing) reveals critical insights into labor market dynamics and forecasting challenges. The U.S. exhibits smoother unemployment cycles, with rapid post-shock recovery facilitated by flexible policies and monetary tools. In contrast, Colombia's structural constraints—including high informality and commodity dependence—lead to prolonged volatility and slower recovery.

1. Model Performance

- For **Colombia**, **NNETAR**, **SSES**, and **TBATS** emerged as the most accurate and robust models. They effectively captured the nonlinear dynamics, trend decomposition, and seasonal irregularities in a volatile labor market.
- For the **United States**, **ETS**, **NNETAR**, and **BSM** outperformed other models on outlier-removed series. These models offered smoother and more accurate forecasts by balancing responsiveness and structural interpretation.
- Contrary to initial assumptions, **SARIMA** underperformed in both countries, particularly due to residual autocorrelation and less reliable forecasts during volatile periods.

2. Developing vs. Developed

- **Colombia**'s structural informality and labor volatility required adaptive and ensemble-based forecasting techniques.
- The **United States**, with more stable institutional dynamics, allowed models like **ETS** and **BSM** to perform well post outlier adjustment.

3. Policy Implications

- Policymakers in developing economies should prioritize adaptive social protection and labor market formalization to reduce volatility in employment metrics.
- In developed economies, maintaining and optimizing automatic stabilizers can ensure resilience in the face of external shocks.

4. Methodological Lessons

- Using ensemble models—such as averaging forecasts from **NNETAR**, **TBATS**, and **SSES** (Colombia) and **ETS**, **NNETAR**, and **BSM** (United States)—reduced average forecast error and improved robustness.
- Cleaning the data by removing outliers significantly enhanced model diagnostics and accuracy but may underrepresent sudden economic disruptions, highlighting a trade-off between smoothness and sensitivity.

5 Acknowledgment of AI Assistance

ChatGPT was used for troubleshooting code syntax and LaTex formatting. All models and forecasts were executed and validated locally in RStudio by the authors.