

# Econ 144 Project 3

Seungyeon Yoo

2023-12-25

## Contents

<b>1. Introduction (describe the data, provide some background on the topic, etc.).</b>	<b>2</b>
(a) Background . . . . .	2
(b) Time series description and exploration . . . . .	2
<b>2. Results</b>	<b>4</b>
(a) STL + ARMA model . . . . .	4
(b) ARIMA Model . . . . .	7
(c) ETS Model . . . . .	9
(d) Holt-Winters Model with Additive Seasonality . . . . .	11
(e) NNETAR Model . . . . .	12
(f) Prophet Model . . . . .	14
(g) Combined Model . . . . .	16
Performance Matrices . . . . .	18
Checking for Volatility of ETS . . . . .	20
<b>3. Conclusions and Future Work.</b>	<b>20</b>
<b>4. References (include the source of your data and any other resources).</b>	<b>21</b>

# 1. Introduction (describe the data, provide some background on the topic, etc.).

This project attempts to model the time series of *unemployment rate among college graduates*, who are aged between 20-24 and have bachelor's degree and higher. The time series spans across the past 23 year, from 2000 to 2023, and the observation frequency is monthly. The primary focus of this project is to apply various modeling methods and finding the best performing forecasting model that could potentially perform well on the unknown future data. For the project purposes, 7 forecasting models were selected, including:

1. STL + ARMA Model
2. ARIMA Model
3. ETS Model
4. Holt-Winters Model
5. Neural Network Autoregressive Model
6. Prophet Model
7. Combined Model

To assess the performance of each model, the test-train split procedure was used, and the ratio of the split was; **train: 80% and test: 20%**. The performance of each model was numerically measured by performance metrics, specifically ME, RMSE, and MAE, and the preferred model was also selected based on the result.

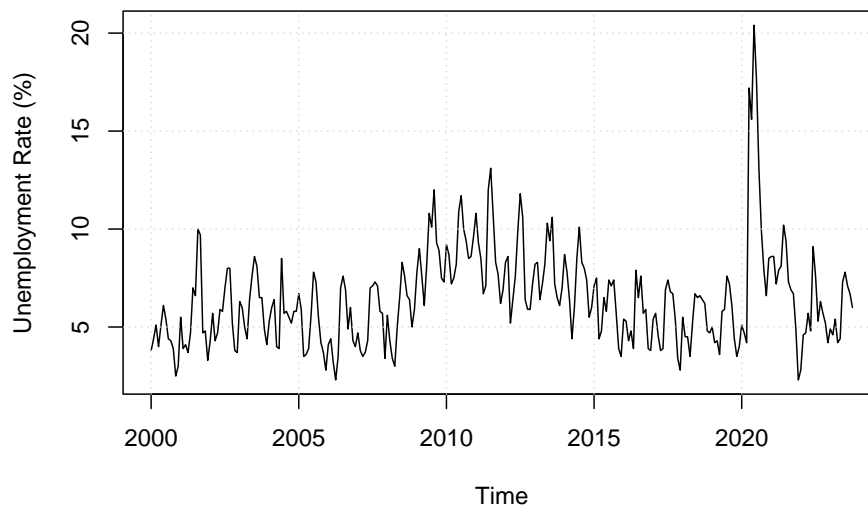
## (a) Background

Unemployment rates among college graduates are a critical economic indicator that addresses the challenges and opportunities faced by young professionals getting into the workforce. A considerable amount of young people who are looking to graduate or have graduated have concerns about getting a job, and many actually do struggle to land their first jobs. Thus, analyzing the historical data on the unemployment rates on college graduates and forecasting the future rates not only could guide individuals to be able to see the big picture of employment trends, but also could play an important role in shaping the future workforce and better understanding the the current labor market, which also could provide insights for policymakers and educational institutions.

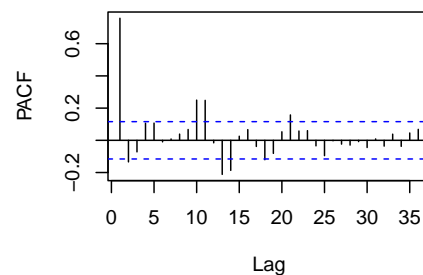
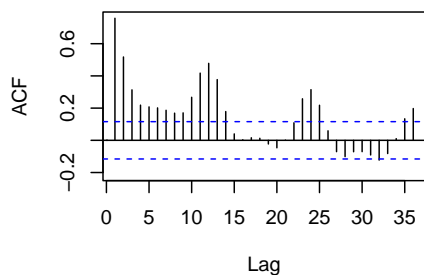
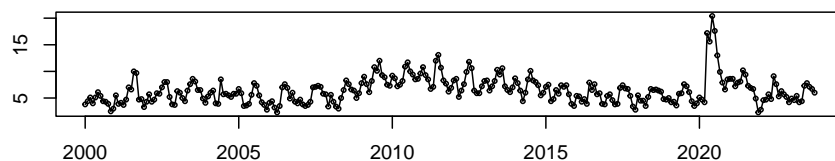
## (b) Time series description and exploration

**Time Series: Unemployment Rate of College Graduates - Bachelor's Degree and Higher, 20 to 24 years**

**Unemployment Rate of College Graduates (2000 – 2023)**



**Unemployment Rate of College Graduates (2000 – 2023)**

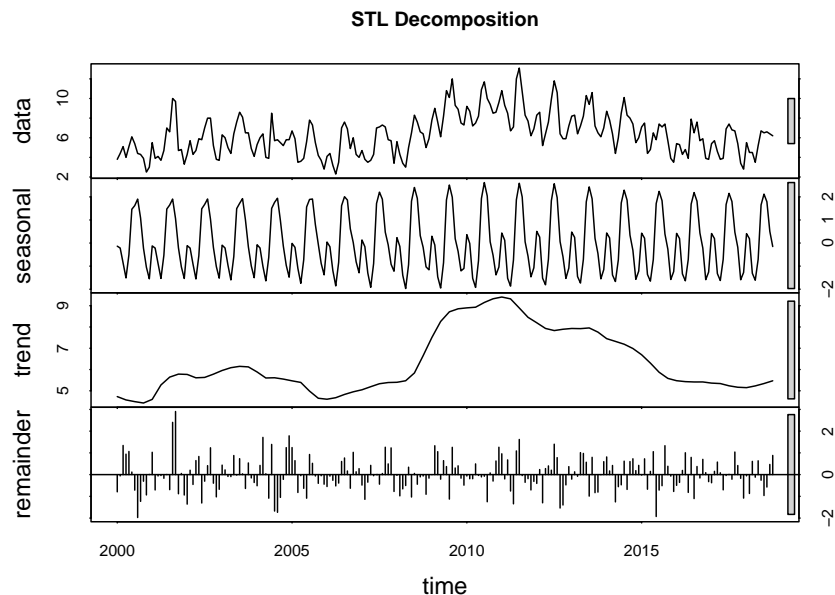


The autocorrelation function plots verify that the original series has significant correlations with its own lagged values and appears to have more dynamics that are hard to be visually interpreted at this point. Beyond the visual inspection, this project attempts to capture more complex dynamics, and they will be explored and discussed in detail in part (2).

## 2. Results

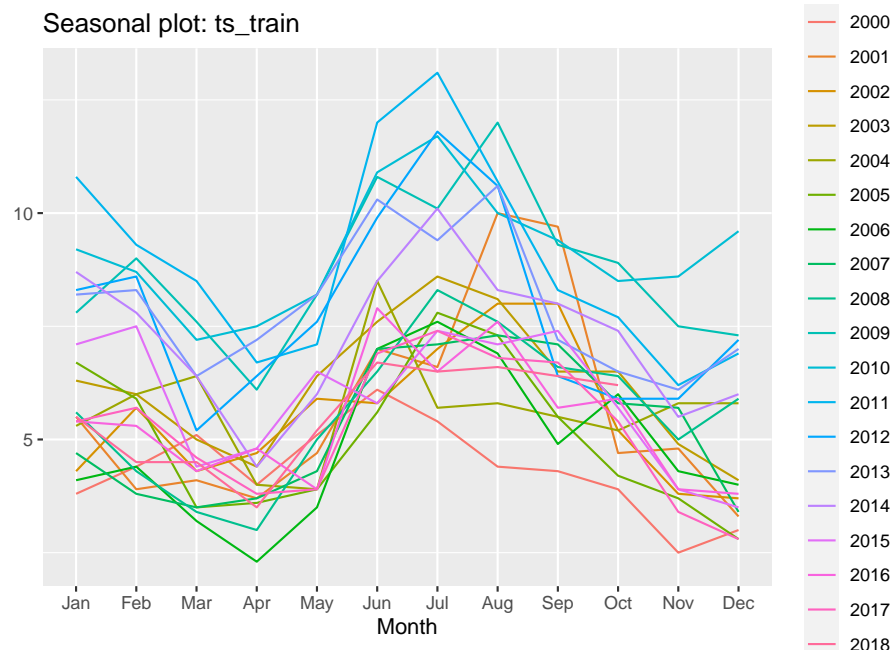
### (a) STL + ARMA model

First, the original series is decomposed into three parts; Seasonal, Trend, and Remainder, with the seasonal window of 12 months. In the plot below, we can see that the times series is exhibiting regular spikes with slightly changing amplitudes overtime. Also, the trend component captures the general trend of the training dataset.

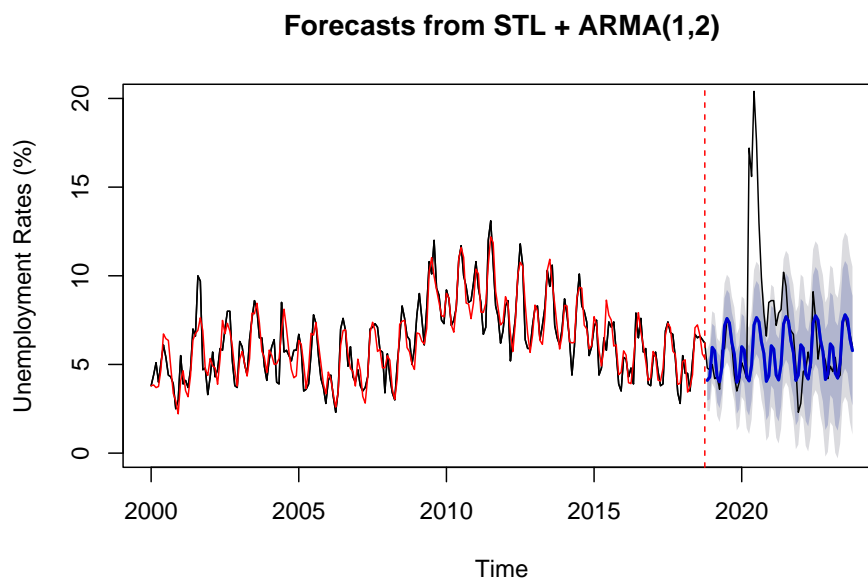


#### Seasonality of the time series

While the seasonal component from STL decomposition depicts the seasonality quite generally, it may be helpful to note that while there are some obvious patterns observed in the seasonal plot, such as decreasing behaviors around March to May, and general spikes around June to August, the seasonal patterns tend to be irregular. The irregularity in the seasonal component can be identified by looking at the seasonal plot.

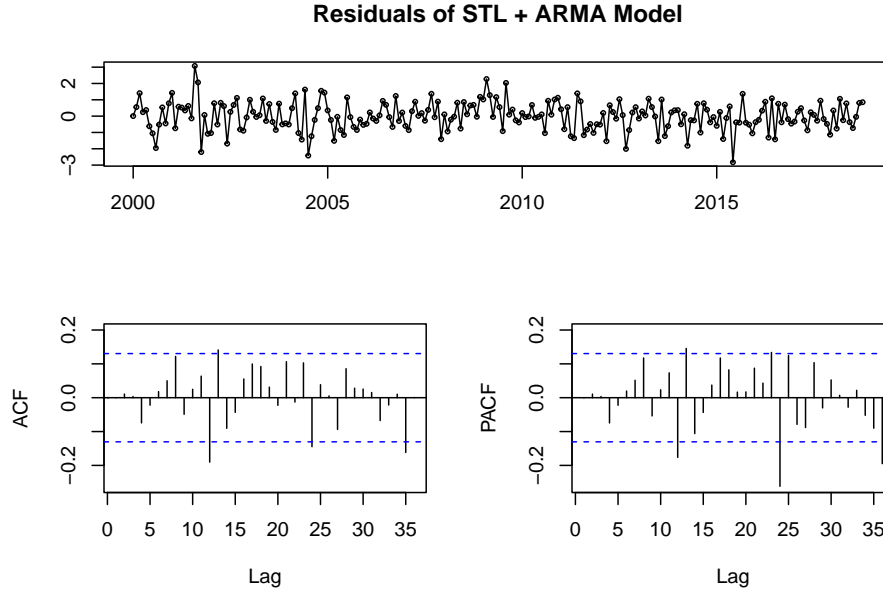


As can be seen in the remainder plot, there is still a lot of dynamics left to be modeled after adjusting for seasonality and trend, and in this section, ARMA process will be used to capture the cyclic components after STL process. To capture the cycles, an ARMA(1,2) process was used. The plot below visualize the model fit.



The black solid line is the original series, and the red dashed line is the point where the train-test split was split by. The green line indicates the model fit and the blue line shows the 60-steps (5 years) ahead forecasts. While the forecasts seem to be performing relatively well, it fails to predict the huge spike in 2019-2020, which was the COVID-19, the pandemic era. However, and the unemployment rates drop as the economy stabilizes, the forecast starts performing well again.

**ACF and PACF** The ACF and PACF plots below indicate some significant dynamics still left in the residuals. In PACF, we can see that there is some seasonal components that have not been captured year based on the spikes in every 12th lags (ex. 12, 24, and 36) and the seasonal patterns are also captured in the ACF as well. This means that, given `auto.arima` did not detect any seasonal patterns and outputted `ARMA(1,2)`, `STL` model may be misspecifying the seasonality of the original data. This may be because the original data has irregular seasonality, and `STL` model with the seasonal window set to 12 may not be the best method the model catch an irregular seasonality. To verify if the spikes are statistically significant enough to say that there is more work to do, Ljung-Box test was used.

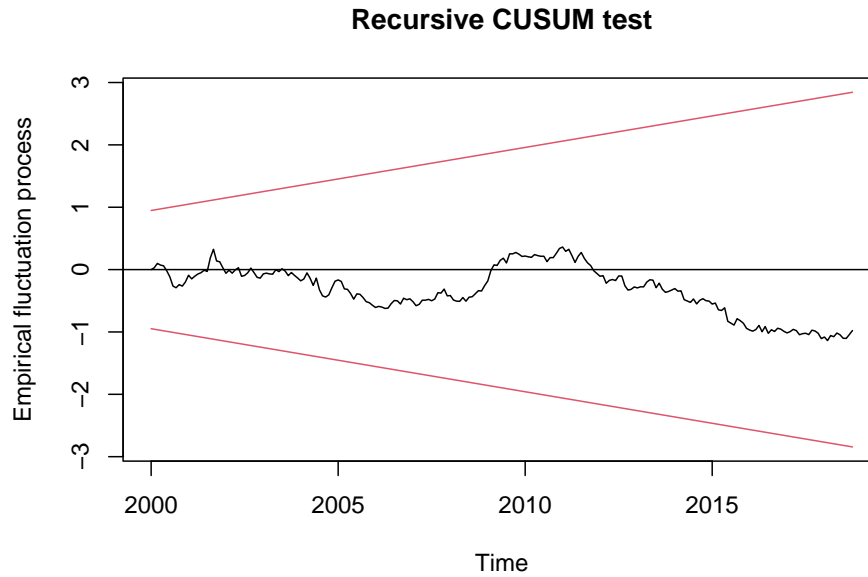


```
##
## Box-Ljung test
##
## data: stl_arima_model$residuals
## X-squared = 3.7659e-05, df = 1, p-value = 0.9951
```

Based on the the p-value of the test, we fail to reject the null hypothesis that there is no autocorrelation in the residuals. In other words, while there are some significant spikes, the Ljung-box concludes that there is no sufficient evidence for autocorrelation in the residuals.

### Cumulative Sum Test

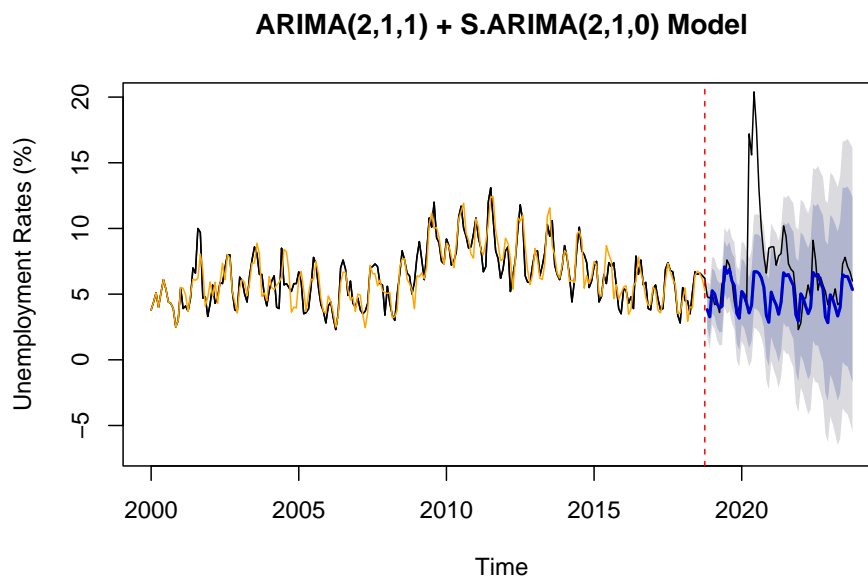
By visualizing the cumulative sum of residuals, we can see that the `CUSUM` line stays within the control lines. Thus, it is confirmed that there is no structural break.



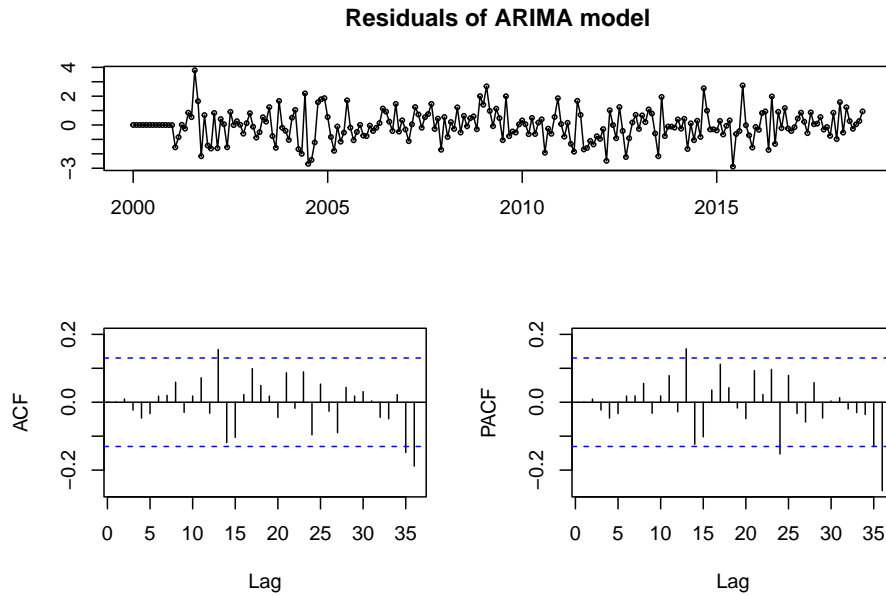
## (b) ARIMA Model

For this model, ARIMA model alone captures all the components previously explored; trend, seasonality, and cycles. Using `auto.arima` function, the optimal order was obtained;  $\text{ARIMA}(2,1,2) + \text{S.Arima}(2,1,0)$ . As observed in the decomposition, there is a trend component and it is captured by  $I(1)$  and Seasonal first order differencing. The seasonal component is modeled by  $\text{S.Arima}(2,0)$ , and  $\text{ARMA}(2,2)$  handles the cycle components.

Also, it is worth noting that the prediction intervals are wider than the  $\text{STL} + \text{ARMA}$  model, which may be an indicator of more uncertainty in the forecasting performance.



## ACF and PACF

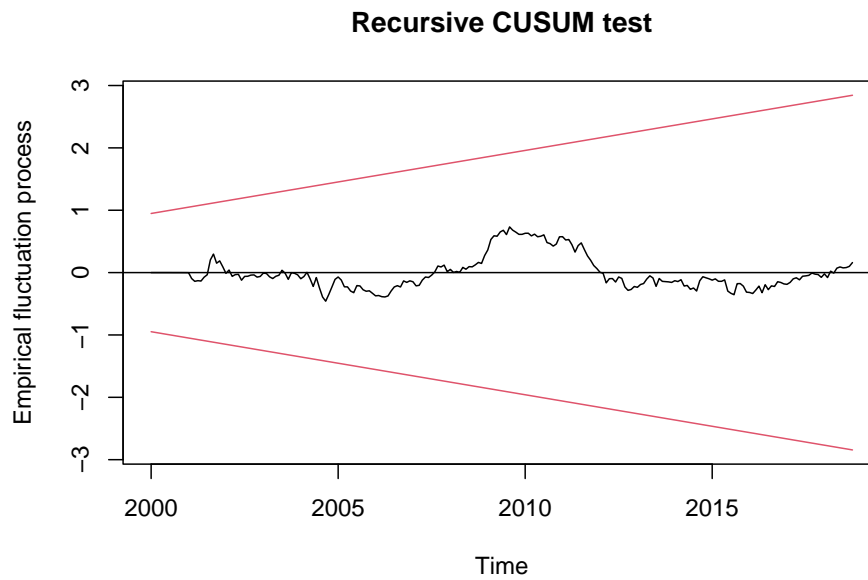


```
##  
## Box-Ljung test  
##  
## data:  arima_model$residuals  
## X-squared = 0.00017781, df = 1, p-value = 0.9894
```

Based on the ACF and PACF, the ARIMA model is capturing the dynamics of the data well. Although some spikes appear to be “significant”, the lag values are far away from the present, and the magnitude of the spikes are not high. To confirm that the residuals are behaving like white noise, Ljung-Box test was used. Since the p-value is high (0.9894), we reject the null hypothesis that there is statistically significant correlation between lags. Therefore, the residuals are like white noise.

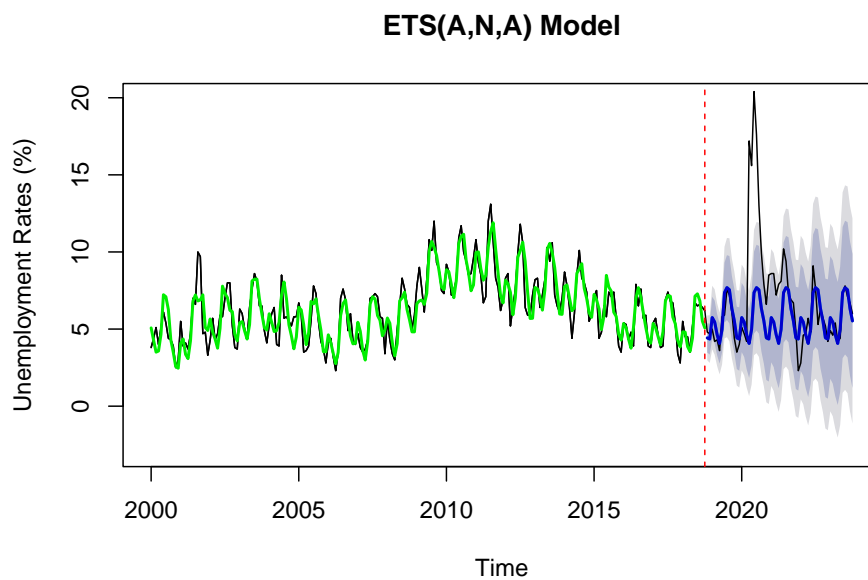
**Cumulative Sum** To test if there is any structural break, the cumulative sum test was used. As shown below, the cumulative sum plot below shows the cumulative sum of errors stays around the horizontal line at 0 and within the red lines. This verifies that there is no structural break in the model.





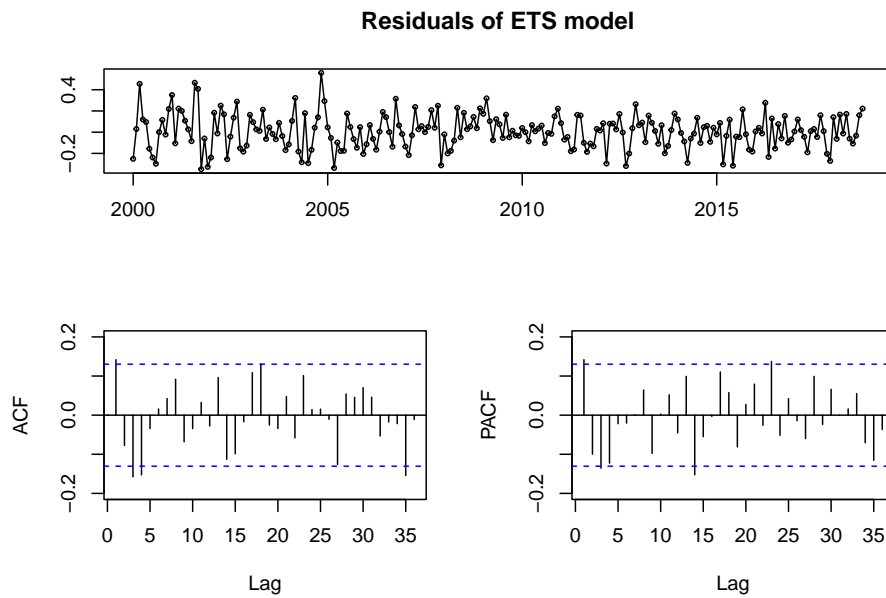
### (c) ETS Model

The plot below shows the ETS (A,N,A), which represents an additive Holt-Winters with additive errors and no explicit trend component.



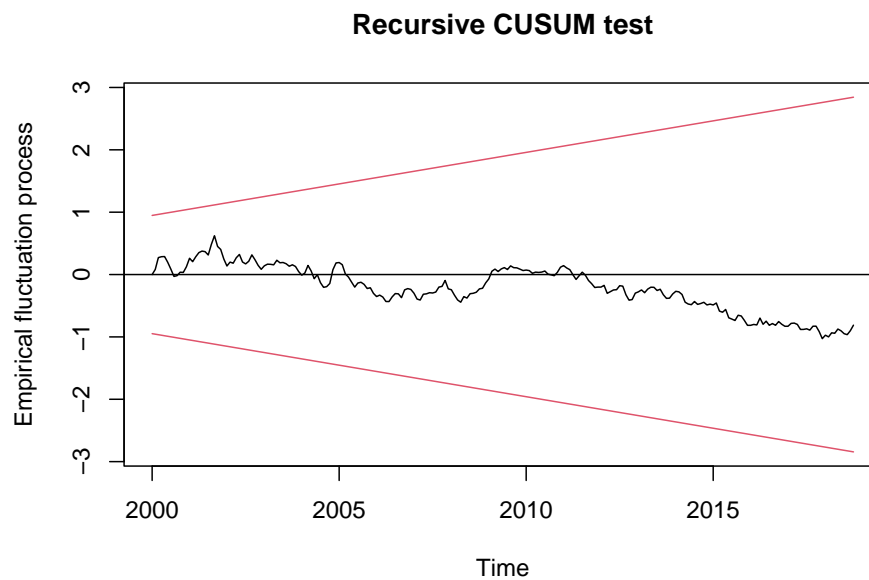
### ACF and PACF

While the model fit on the original series seems to capture the dynamics well, ACF and PACF do not agree on that. Based on the ACF, PACF, and Ljung-Box test, the residuals still have statistically significant autocorrelation, and thus, there may be some more dynamics that ETS failed to capture.



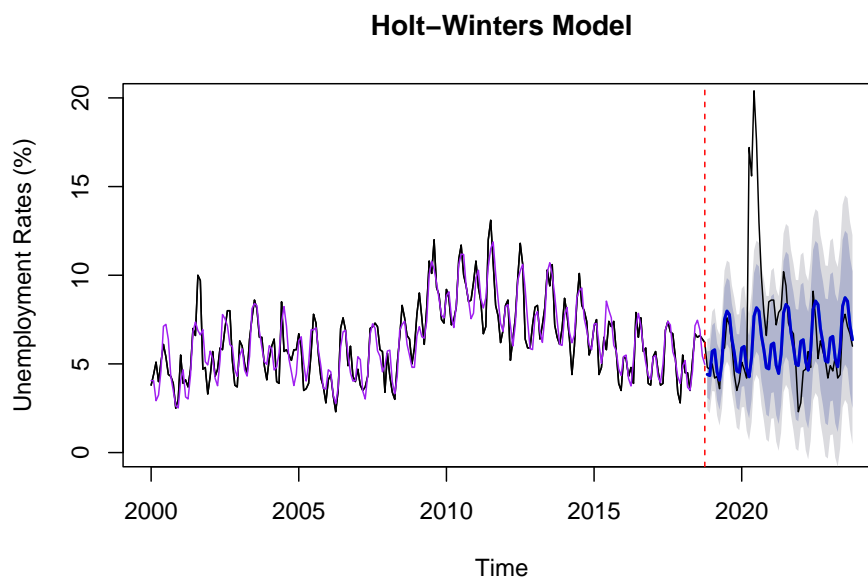
```
##
## Box-Ljung test
##
## data: ets_model$residuals
## X-squared = 4.6073, df = 1, p-value = 0.03184
```

**Cumulative Sum** Similar to the previous models, no structural breaks were observed in the residuals of model.

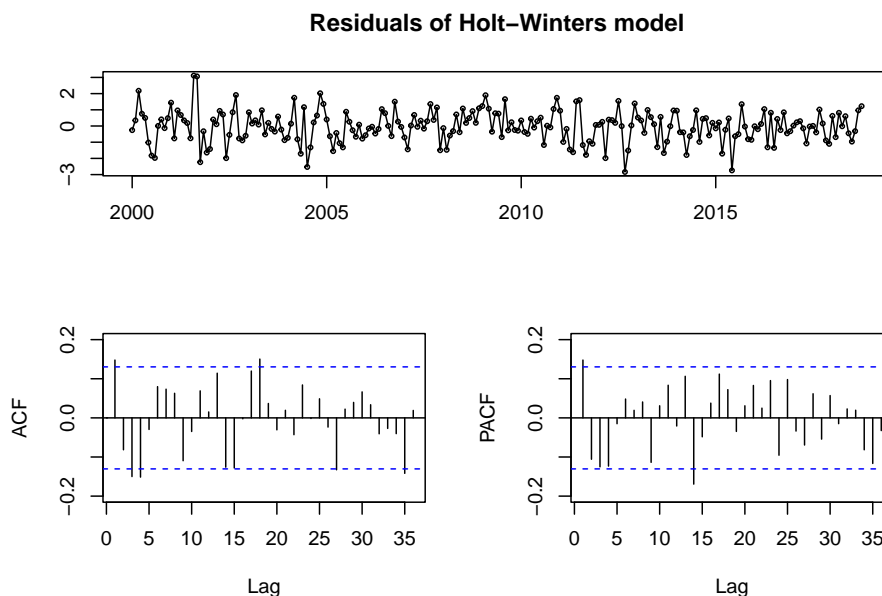


#### (d) Holt-Winters Model with Additive Seasonality

Since the amplitudes of the seasonality appear to be constant over time, except for the huge spike in 2019, an **additive Holt-winter** model was chosen. The plot below visualizes the model fit and 60-steps ahead forecasts on the testing set, which is the actual data.

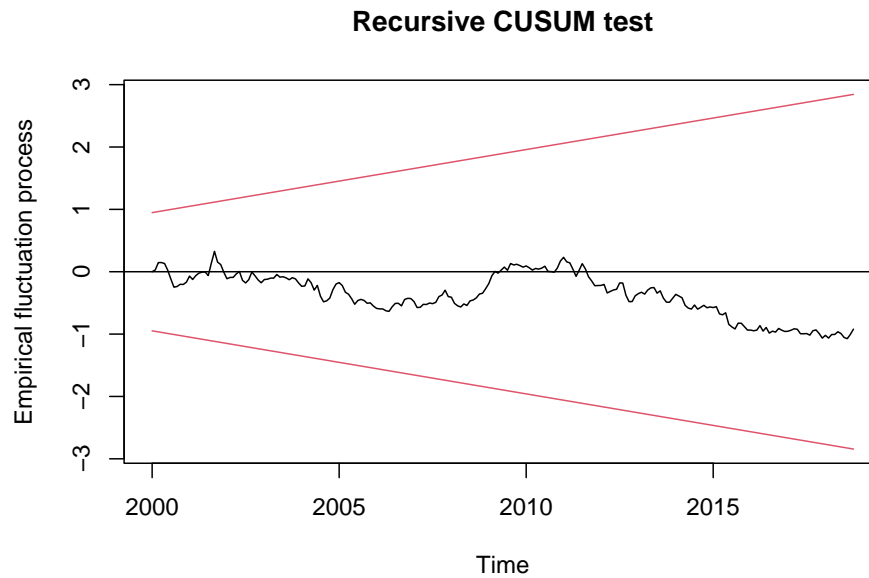


**ACF and PACF** The ACF and PACF plots indicates significant autocorrelation left in the residuals. As also verified by the Ljung-test, we reject the null hypothesis that there is autocorrelation at a significance level of 0.05. In other words, the model fit could be improved and there might be more work to be done.



```
##
## Box-Ljung test
##
## data: hw_model_add_model$residuals
## X-squared = 5.0052, df = 1, p-value = 0.02527
```

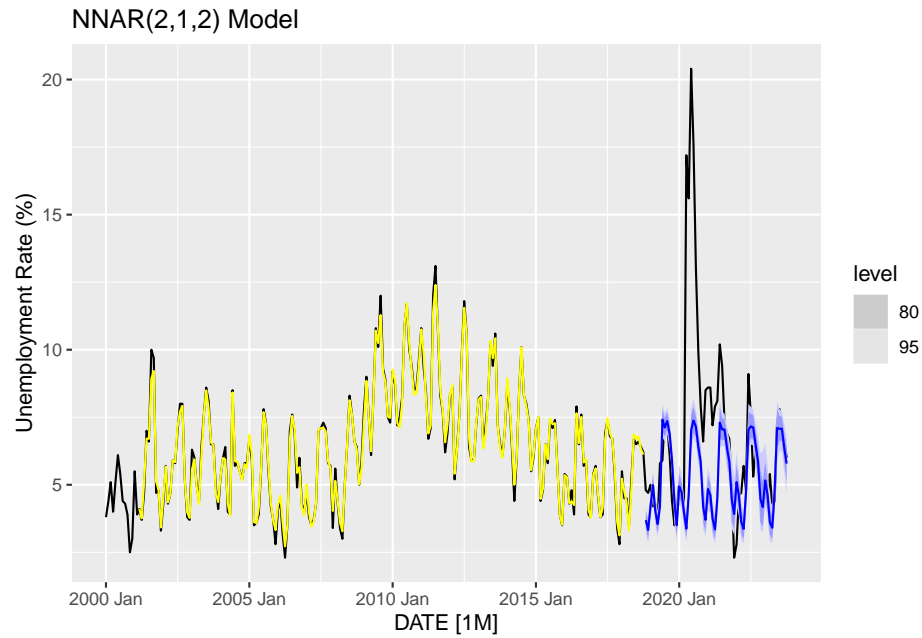
**Cumulative Sum** As confirmed by the plot below, there is no significant structural breaks in the errors of the model.



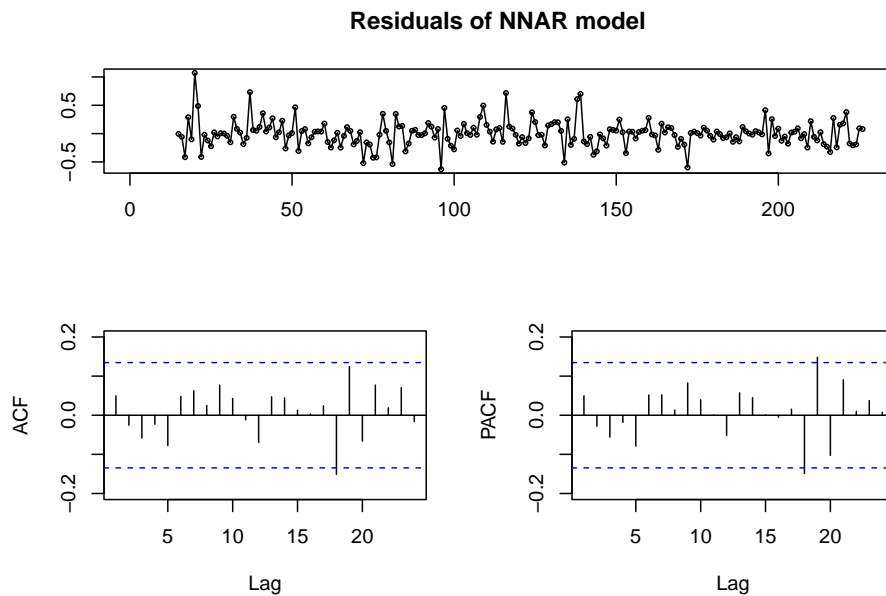
### (e) NNETAR Model

As shown below, NNAR(2,1,2) was chosen as the optimal order of Neural Network Autoregressive model with inputs of:

- $y_{t-1}, y_{t-2}, y_{t-3}$
- $y_{t-12}$
- two neurons in the hidden layer

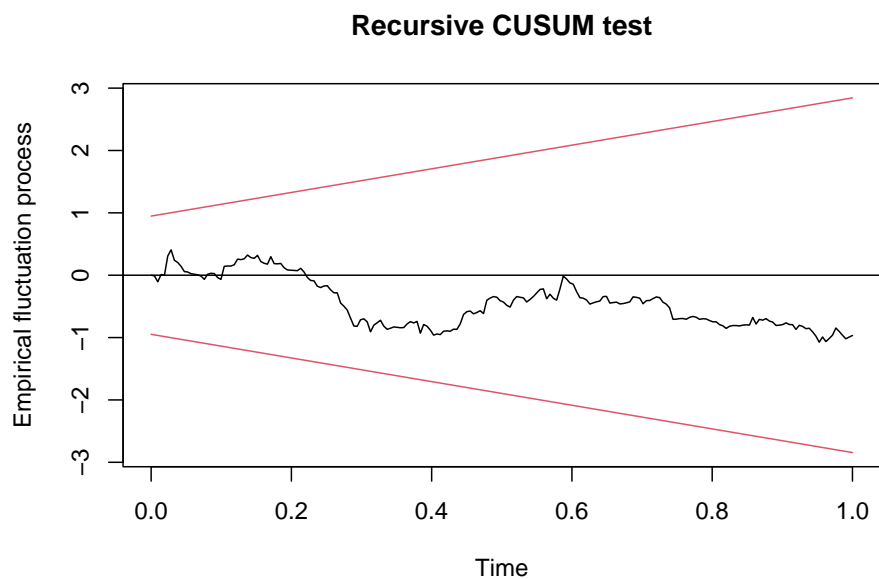


**ACF and PACF** For NNAR model, the ACF and PACF are almost completely wiped out, which is a very good sign that the NNAR model is capturing the dynamics of the original time series well. As expected from the plots, Ljung-box test confirm, with p-value of 0.9315, that there is no evidence of autocorrelation.



```
##
## Box-Ljung test
##
## data: resid(NNAR_model)$.resid
## X-squared = 0.52668, df = 1, p-value = 0.468
```

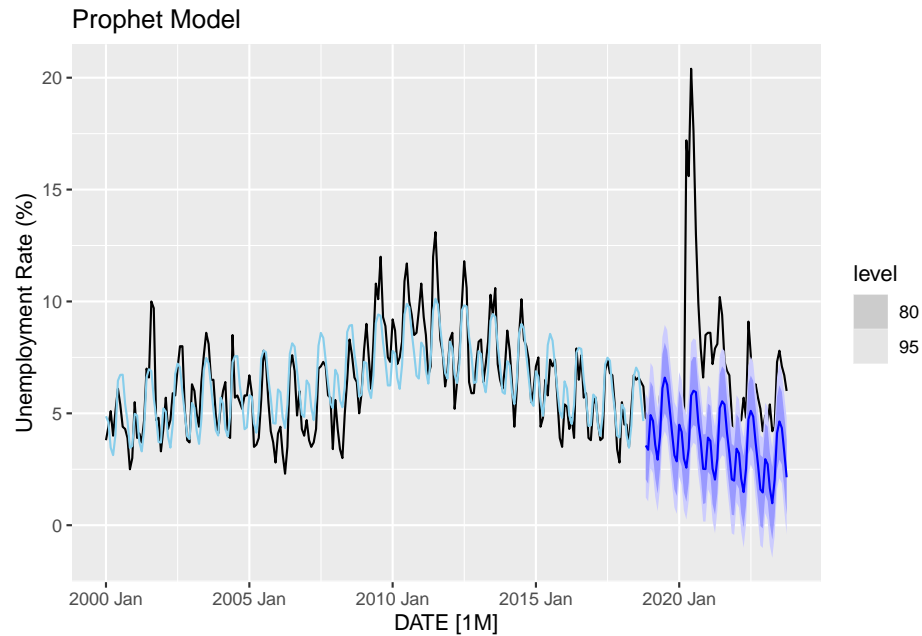
**Cumulative Sum** NNAR model also does not have any structural breaks detected, which suggests that the model is stable over time.



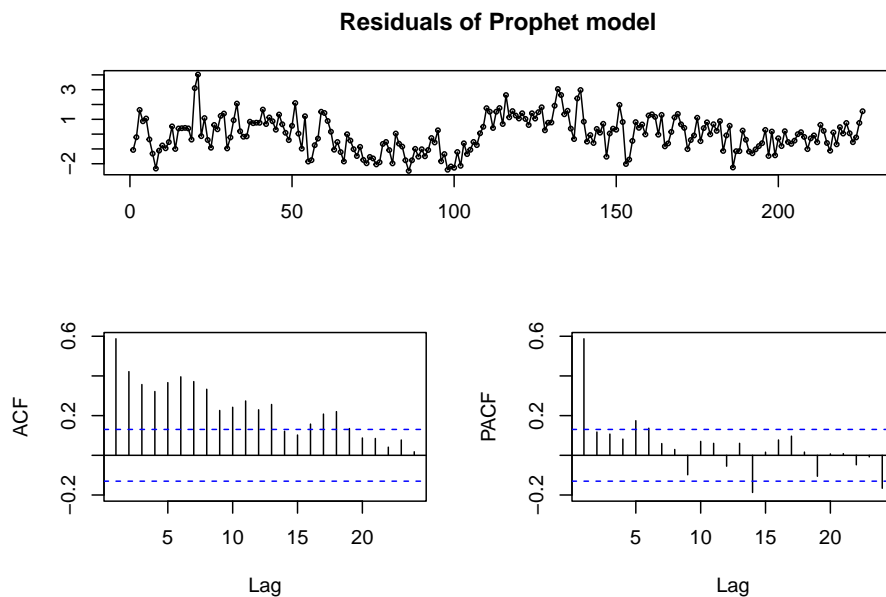
### (e) Prophet Model

The Prophet model is designed to capture more complex seasonal patterns typically found in higher frequency data, such as minute or hourly data. Since the prophet function expects at least daily frequency, the monthly frequency was explicitly specified in the model. Additionally, the order argument is set to 6, indicating the inclusion of 6 Fourier terms to model the monthly seasonality, allowing the model to adapt to the lower frequency of the data.

As seen below in the plot, the performance of the prophet model is expected be low compared to the other models.

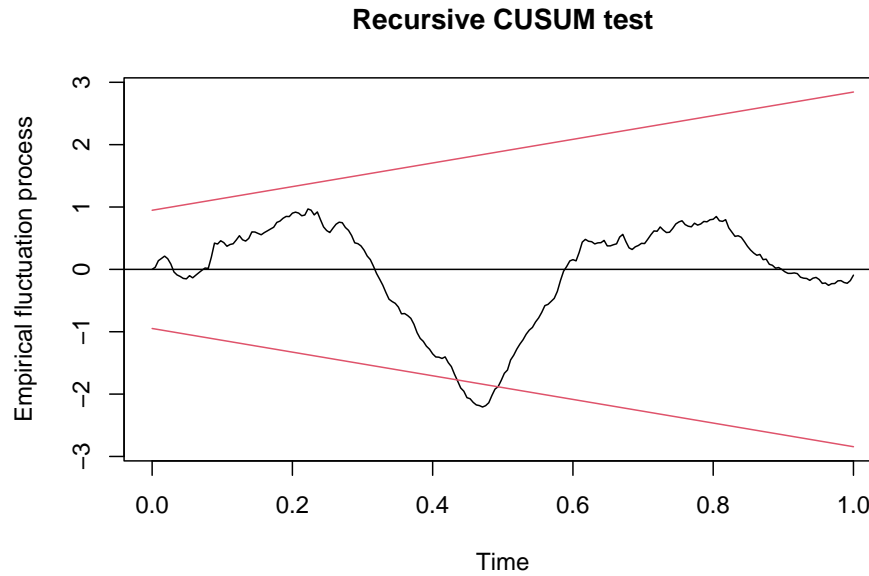


**ACF and PACF** As expected, the residual plot still has a lot of patterns, and ACF, PACF, and the Ljung-Box test provide statistical measure (low p-value) supporting that.



```
##
## Box-Ljung test
##
## data: resid(prophet_model)$resid
## X-squared = 78.983, df = 1, p-value < 2.2e-16
```

**Cumulative Sum** For the prophet model, there is a structural break in the middle, which suggest that the prophet model may not be an effective method to model this time series.



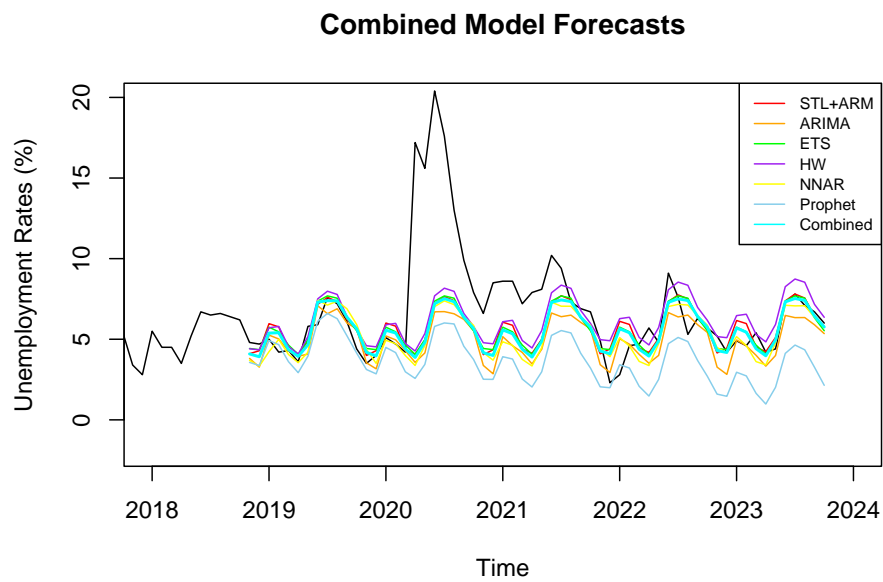
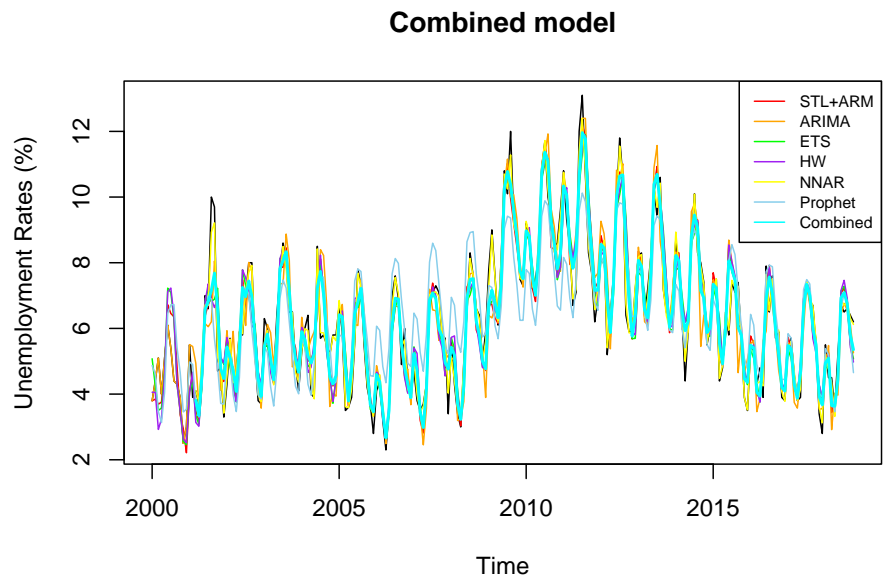
#### (f) Combined Model

The purpose of combining different models is to mitigate the uncertainty and leverage the strengths that each model has. For this combined model, 5 models were considered including:

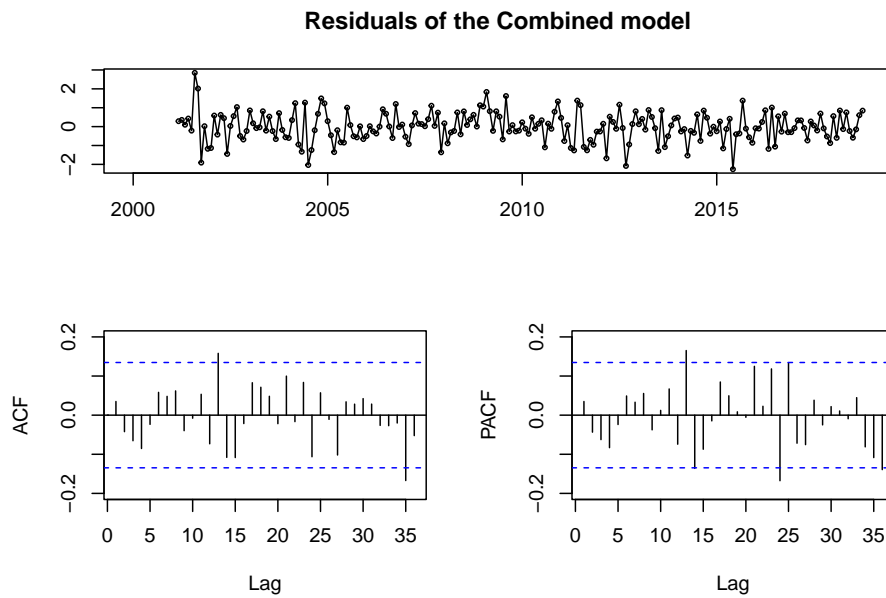
1. STL + ARMA Model
2. ARIMA Model
3. ETS Model
4. Holt-Winters Model
5. Neural Network Autoregressive Model

Note the the prophet model was excluded in the combined model because the performance of the model was exclusively worst compared to the other models. In addition, the combined model performed better on the performance metrics without the prophet model.



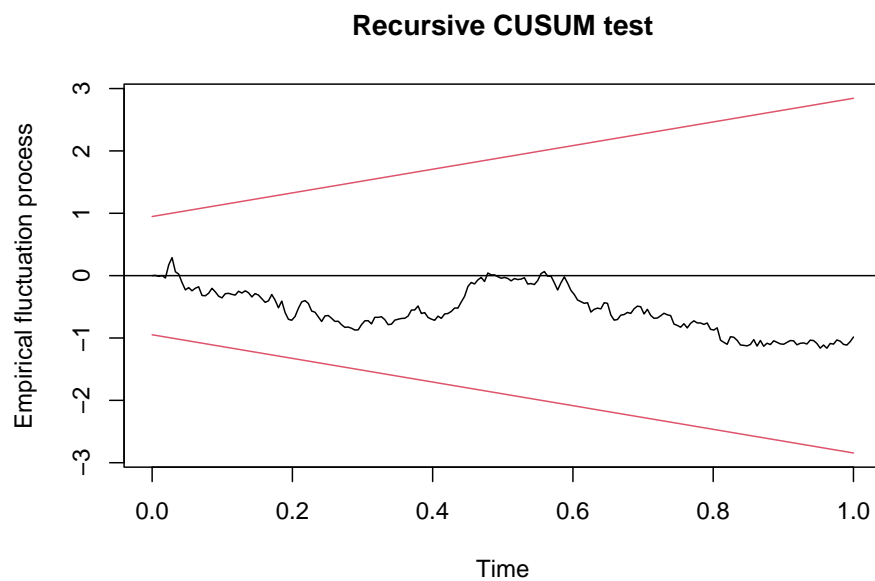


**ACF and PACF** The residuals of the combined model is behaving like a white noise. Except for a couple of random spikes in the further lags, there is no autocorrelation left to be modeled, which is also supported by the Ljung-Box test.



```
##
## Box-Ljung test
##
## data: ts_train - combined_fit
## X-squared = 0.26237, df = 1, p-value = 0.6085
```

**Cumulative Sum** The combined model also does not exhibit any structural break.



**Performance Matrices**

	ME	RMSE	MAE
STL+ARMA	0.0034661	0.8767425	0.6877418
ARIMA	-0.0236640	1.0521351	0.7875806
ETS	0.0071651	0.1619837	0.1267875
Holt-Winters	-0.0258617	0.9810644	0.7584231
NNAR	-0.0008457	0.2278277	0.1589421
Combined	-0.0056362	0.7806043	0.6007044

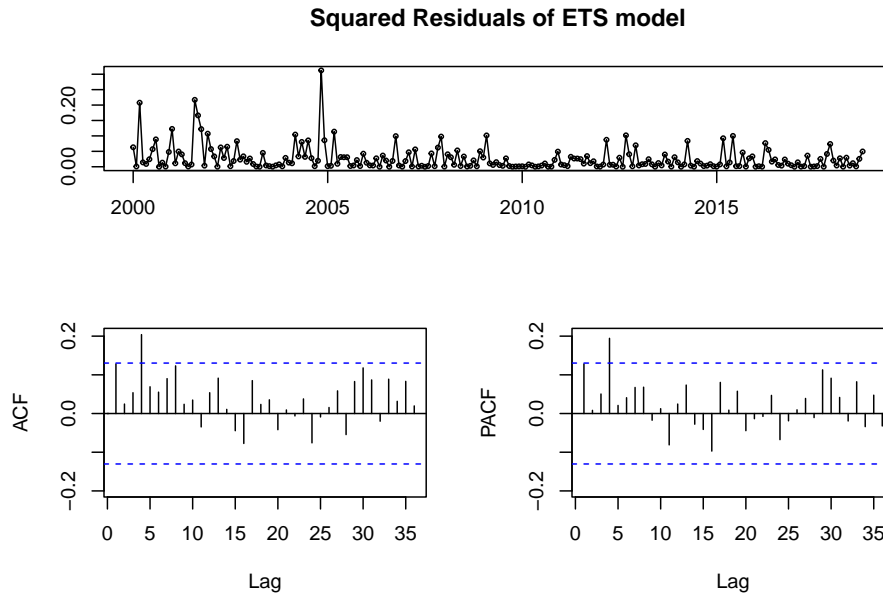
	ME	RMSE	MAE
STL+ARMA	1.2125086	3.501253	1.957406
ARIMA	1.9451674	3.802638	2.230059
ETS	1.2280216	3.470641	1.920897
Holt-Winters	0.7502916	3.384005	1.995046
NNAR	1.6454187	3.667321	2.057873
Combined	1.3562816	3.534797	1.939295

To select the preferred model, ME, RMSE, and MAE are used. Based on the model fit and ACF and PACF plots, I initially thought that the best performing model would either be ARIMA, NNAR, or the combined model. Those are the models with no leftover dynamics in the residuals plot. However, according to the performance metrics above, we have come to a little different conclusion than initially expected.

First, the positive ME's show that the models are generally *underestimating* the actual values. This may be partially because of the significant spike in the unemployment rate during the pandemic. The models in this project were not able to capture the sudden shock caused by a random event, COVID-19.

In terms of the model performance, either ETS or NNAR could be the best performing model on the training dataset as they scored the lowest on RMSE and MAE. On the test dataset, Holt-Winter's and ETS model are the first and the second place. Therefore, considering the performance scores on both training and testing set, the preferred model would be ETS model.

## Checking for Volatility of ETS



```
##  
## Box-Ljung test  
##  
## data: ets_model$residuals^2  
## X-squared = 3.753, df = 1, p-value = 0.05271
```

At a significance level of 0.05, we conclude that there is no volatility to be captured in our preferred model, ETS.

## 3. Conclusions and Future Work.

In conclusion, this project attempted to model the unemployment rate among college graduates with a bachelor's or higher degree, aged between 20 and 24, using 7 different models including STL+ARMA, ARIMA, ETS, Holt-Winters, NNAR, prophet, and a model that combines of the above models. To assess the model fit, the behavior of the residuals, and the performances of the models, the ACF and PACF correlograms, Ljung-Box tests, and the cumulative sum tests were used. Specifically, to determine the preferred model, each model's performance on both training and testing datasets was evaluated. As a result, we conclude that the ETS model performs the best in capturing the dynamics of the original time series, and potentially in forecasting the future rates.

For future work, although they were not included in this particular project, a model that could account for other economic factors correlated with the unemployment rate, such as the VAR model or a method that splits the series into parts and models the varying trends and seasonal components may be able to perform better in terms of the model fit and forecasting.

#### **4. References (include the source of your data and any other resources).**

U.S. Bureau of Labor Statistics. “Unemployment Rate - College Graduates - Bachelor’s Degree, 20 to 24 Years.” FRED, Federal Reserve Bank of St. Louis, 1 Jan. 2000, <https://fred.stlouisfed.org/series/CGBD2024>.