Business Economic and Financial Data

Unemployment Rate Time Series Analysis

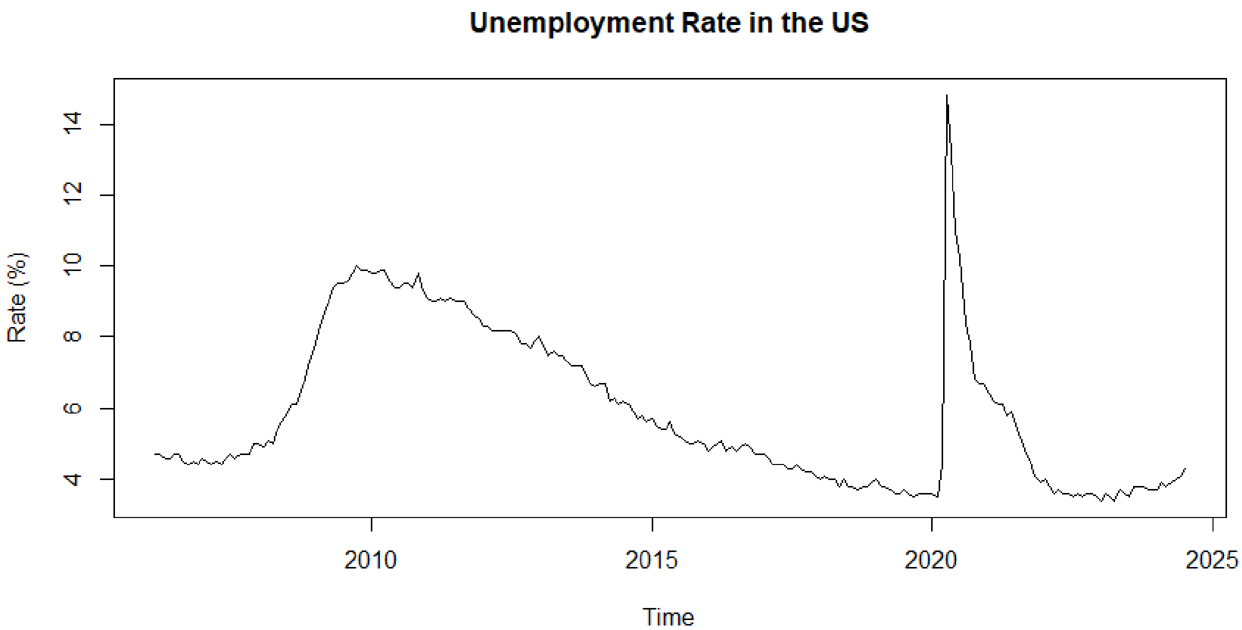
Ivan Tepliashin (2112455), Nikita Paltov (2107498)

Project Description

In this project we perform the analysis of an unemployment rate in the United States time series and apply different methods and models trying to fit the data and predict future values. We consider 6 different models and model families: ARIMA, Exponential Smoothing, Prophet, TSLM, GAM and gradient boosting. To validate predictions of our models we leave a part of data out as test data and fit the models on the rest. Next, predictions are made on this previously unseen data and MSE metric is calculated. Models from one family are generally compared with each other using AIC. In some models, where we do not use external regressors, we perform the training on the full data and predict the future values, which are unknown to us.

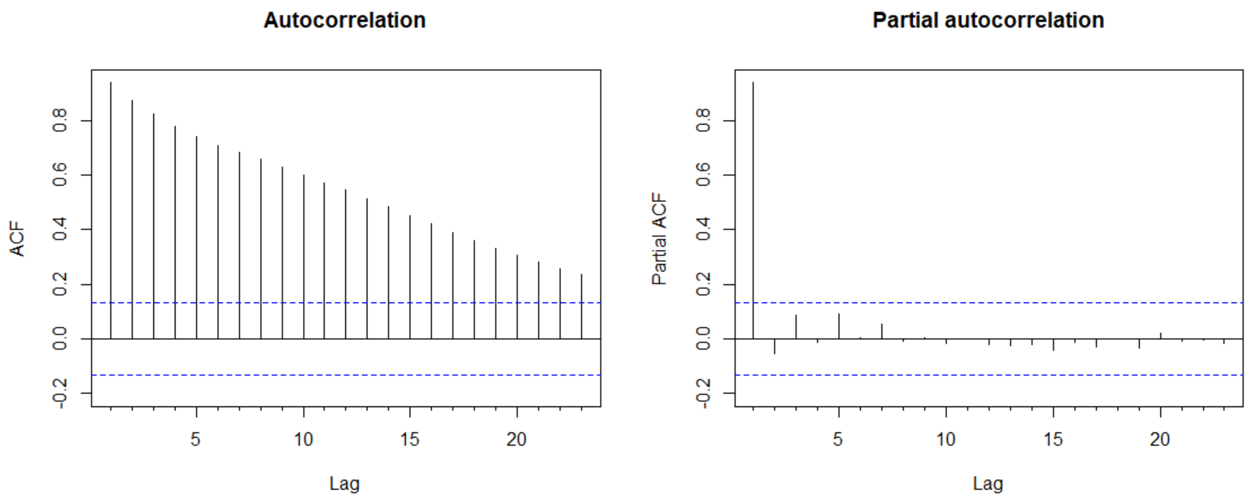
Data Presentation

We took the Unemployment Rate data and most of the additional data from the FRED website. Unemployment rate time series displays monthly unemployment rate in the U.S. since 1948. This rate is measured in percents. However, in our project we focus on the period from March 2006 up to July 2024, to ensure compatibility with our other data.



On the plot of the data we can see two clear peaks, the first one caused by the consequences of the 2007-2008 financial crisis and the second one caused by the COVID-19 pandemic in 2020. After the crisis in 2008 the unemployment rate fell gradually until 2020. Moreover, we can see that it started to rise a bit in 2024, probably due to economic instability. The second peak is much higher and sharper and would be probably harder to predict by the models.

We can also take a look at the autocorrelation plots:



The autocorrelation function shows that our data is nonstationary. Some models like ARIMA can take that into account.

Additional Data Presentation

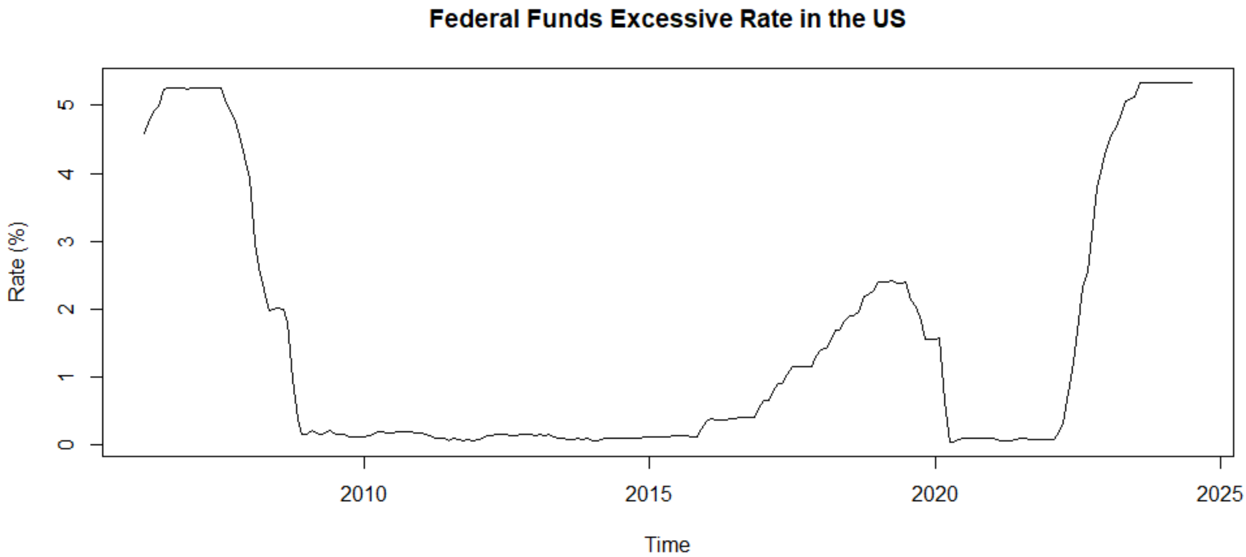
To assess the impact of macroeconomic and other indicators, we used additional data as external regressors:

1. Federal Funds Effective Rate – the interest rate at which depository institutions trade federal funds (balances held at Federal Reserve Banks) with each other overnight. The effective federal funds rate is essentially determined by the market but is influenced by the Federal Reserve through open market operations to reach the federal funds rate target. The federal funds rate is the central interest rate in the U.S. financial market. It is measured in percents.
2. Consumer Price Index for All Urban Consumers (CPI) – a price index of a basket of goods and services paid by urban consumers. This index includes roughly 88 percent of the total population, accounting for wage earners, clerical workers, technical workers, self-employed, short-term workers, unemployed, retirees, and those not in the labor force. The index measures price changes (as a percent change) from a predetermined reference date. The index at the time of 1982-1984 equals 100.
3. Gross Domestic Product (GDP) – the market value of the goods and services produced by labor and property located in the United States. GDP is usually calculated quarterly, so we augmented it for missing months with simple linear interpolation. It is measured in billions of dollars.
4. Real Gross Domestic Product – the inflation adjusted value of the goods and services produced by labor and property located in the US. Real GDP is also calculated quarterly, so we augmented it for missing months in the same way we did with GDP, using simple linear interpolation. It is measured in billions of dollars.
5. Average Weekly Hours of all Employees – the total weekly hours divided by the employees paid for those hours. It is measured in hours.
6. Average Hourly Earnings of All Employees – a measure of the average hourly earnings of all private employees on a “gross” basis, including premium pay for overtime and late-shift work. This measure excludes benefits, irregular bonuses, retroactive pay, and payroll taxes paid by the employer. It is measured in dollars per hour.

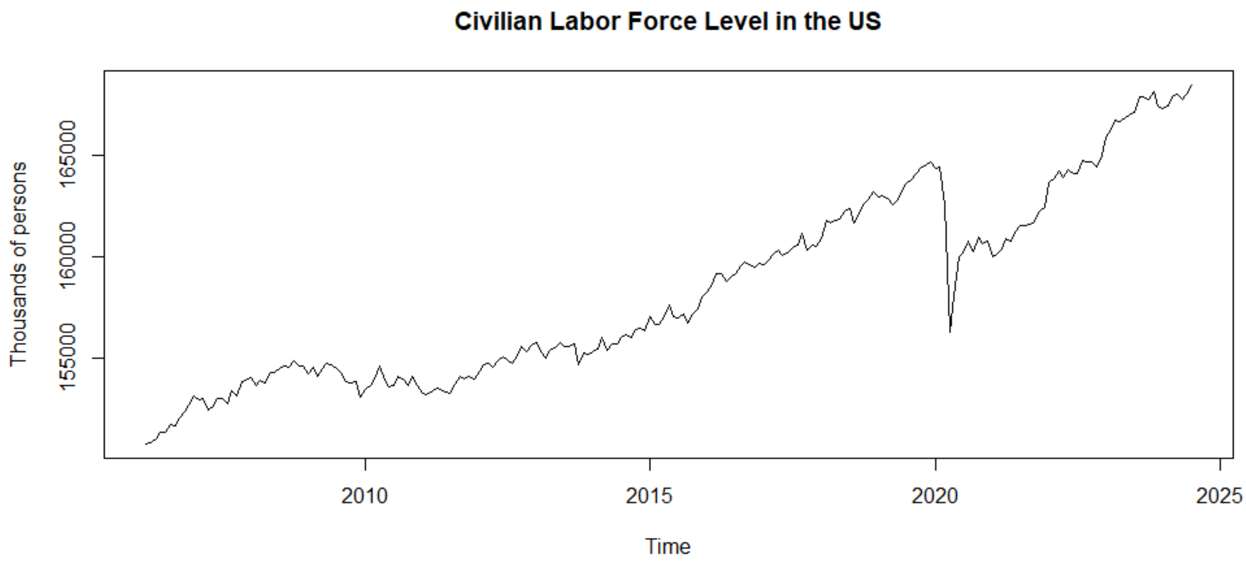
7. Civilian Labor Force Level – the amount of the working-age U.S. civilian population that is either employed or available for employment. It is measured in thousands of persons.
8. BBK Monthly GDP Growth – quarterly estimates of real GDP growth in the U.S. Brave-Butters-Kelley (BBK) index is the byproduct of research originally conducted by the Federal Reserve Bank of Chicago. It is represented by annualized percent change from preceding period.
9. Covid Data – cumulative number of COVID-19 cases. This data was originally collected by NY Times in real time as they were identified after testing. It covers time period from 2020 to 2023. The data is recorded daily, but most of our data is monthly data, so we took only values on the 1st of each month. Moreover, as it is cumulative, in our work we added zero values for each month from 2006 up to 2020 and the last value from 2023 for each month in 2024 and used first difference of this time series. It is measured in total number of cases.
10. SPX – historical data on S&P 500 (^SPX ticker). It is a stock market index tracking the stock performance of 500 of the largest companies listed on stock exchanges in the United States. The index can provide a broad view of the economic health of the U.S. because it covers so many companies in so many different sectors.

All data were taken for the period from March 2006 to July 2024.

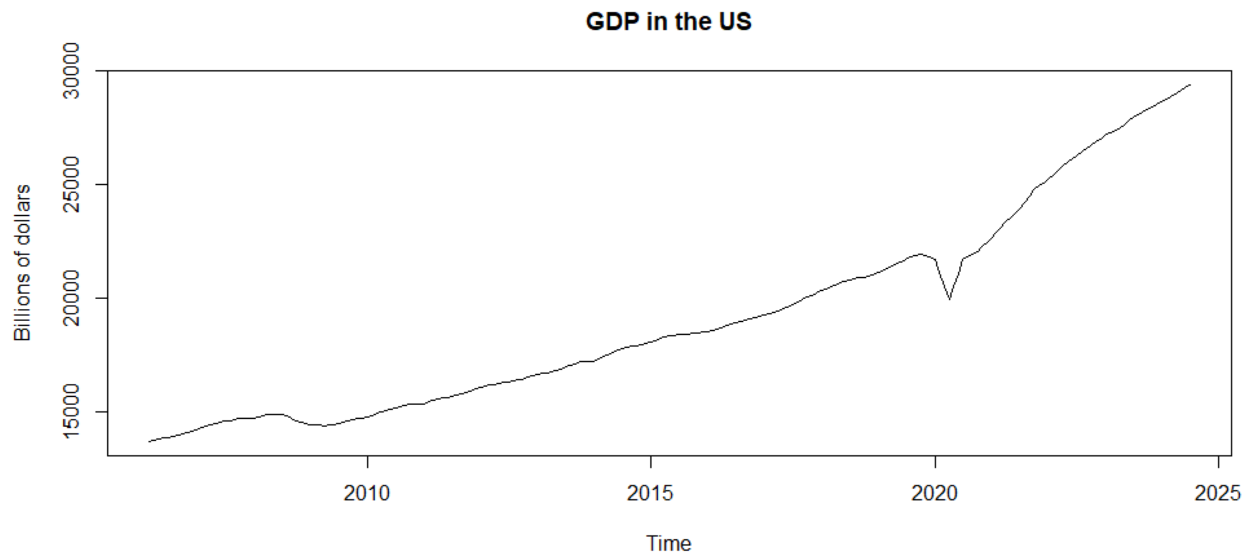
Let us also take a look at plots of some of these data:



The Federal Funds Excessive Rate over time reflects the Federal Reserve's monetary policy response to economic conditions. From this plot we can see that the Federal Reserve lowered the rate dramatically during the financial crisis in 2008 to stimulate economic activity. A lower federal funds rate makes borrowing cheaper, encouraging businesses and consumers to spend and invest. Next, from 2015 the rates started to rise gradually until the sharp drop in 2020 due to the economic shock caused by the COVID-19 pandemic. Starting in 2022, the rates were raised aggressively likely to combat rising inflation.



Over time, the labor force generally grows due to population increases and more individuals entering the workforce. During periods of economic growth, more people participate in the labor force as job opportunities increase. From the plot we can see that during and after the 2008 crisis, labor force participation stagnated and declined in some periods due to job losses, discouragement among workers, and slower recovery in employment opportunities. Also, there is a sharp drop in 2020, which reflects a decline caused by the pandemic, when many people became temporarily or permanently unemployed.



GDP constantly grows over time due to numerous factors, such as population (and therefore labor force) growth, technological advancements, trade and globalization and human capital development. Also, this is a plot of the nominal GDP, meaning that inflation also plays its role here. We can notice a slight fall in 2008 (due to the crisis) and a more significant drop in 2020, caused by the pandemic.

We can notice that all these data depict in some way the influence of two major crises (2008 and 2020) on the economy.

ARIMA models

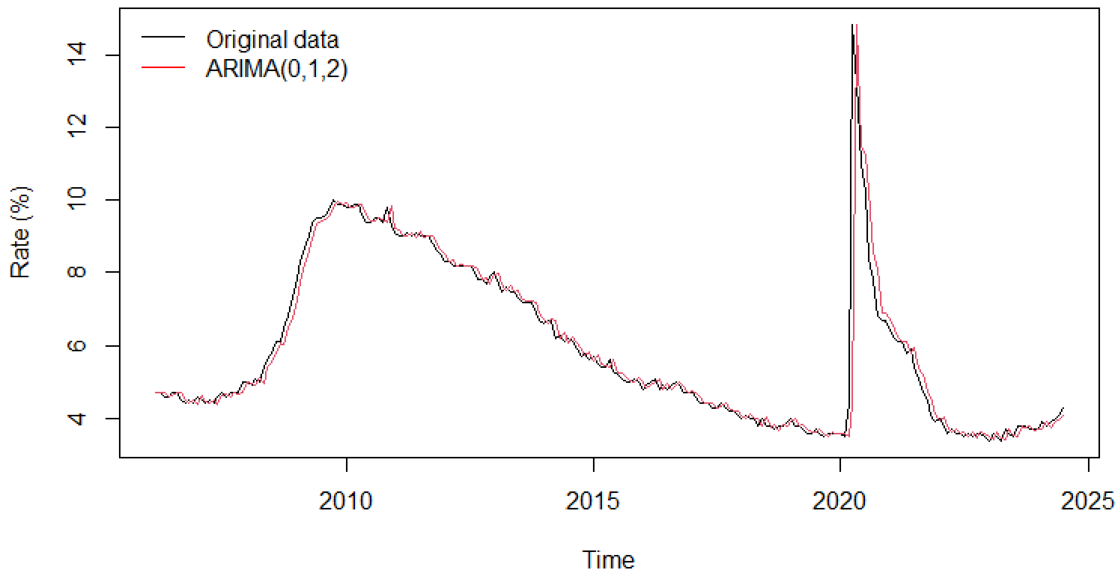
First, we tried to predict the unemployment rate with an ARIMA model. We selected the optimal ARIMA model based on the Akaike Information Criterion (AIC). Next, we performed an exhaustive search for the optimal ARIMA model by testing different combinations of the autoregressive order p, the differencing order d, and the moving average order q, within the range of 0 to 3 for each parameter. The AIC values were calculated for each combination, and the model that produced the lowest AIC value was selected as the best model.

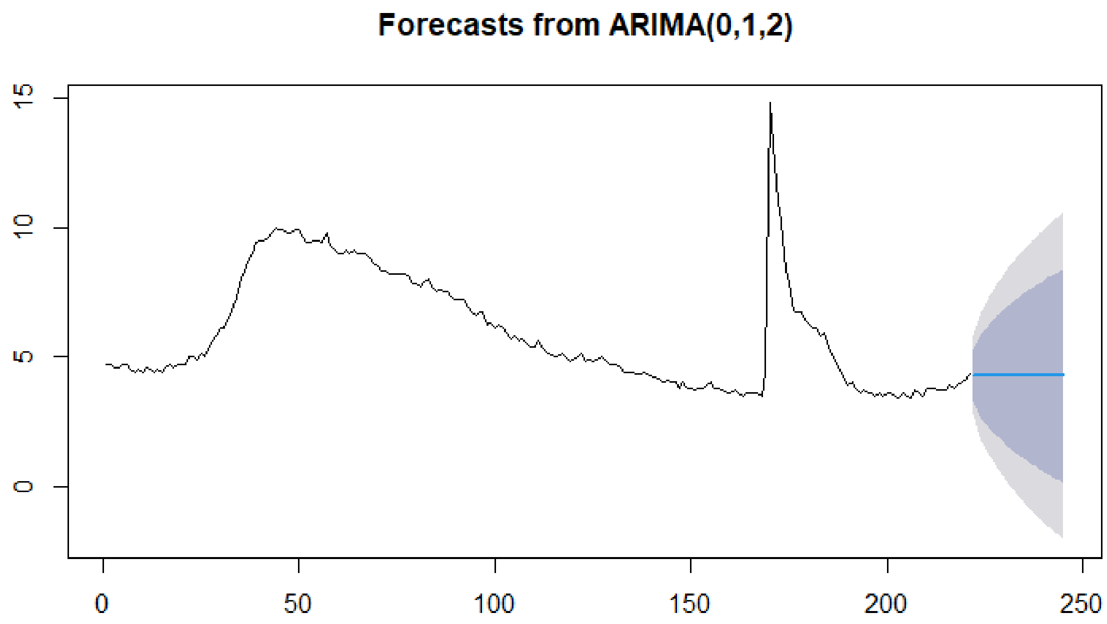
After this process, the best-fitting ARIMA model was found to be ARIMA(0,1,2), with an AIC of 506.8715. This model suggests that first differencing of the data ($d = 1$) is needed to achieve stationarity, while the moving average component requires two lags ($q = 2$) to adequately capture the autocorrelation structure in the residuals. The autoregressive term was not significant in this case, as ($p = 0$). Originally, our data was nonstationary, but we do not have to calculate first difference by ourselves, because the model does it for us in this configuration.

By looking at the Partial Autocorrelation Function of our data (presented earlier), it appears that the autoregressive order should be 1, as the first lag showed a significant correlation. However, despite this indication, the selected ARIMA model did not include an autoregressive term, suggesting that a simple differencing and moving average model provided a more optimal fit for the data.

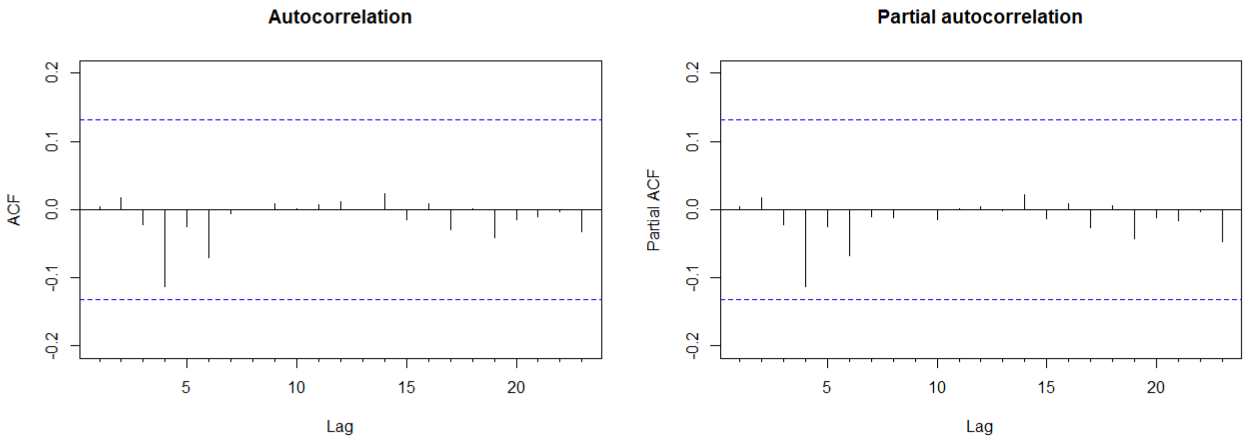
The model is used to forecast the unemployment rate for the next 24 months, or two years. The forecast horizon was chosen to provide a reasonable outlook for short-term trends in the unemployment rate. Importantly, no significant seasonality was detected in the data, which was confirmed by both the model selection process and the seasonal component tests.

Fit by ARIMA(0,1,2):





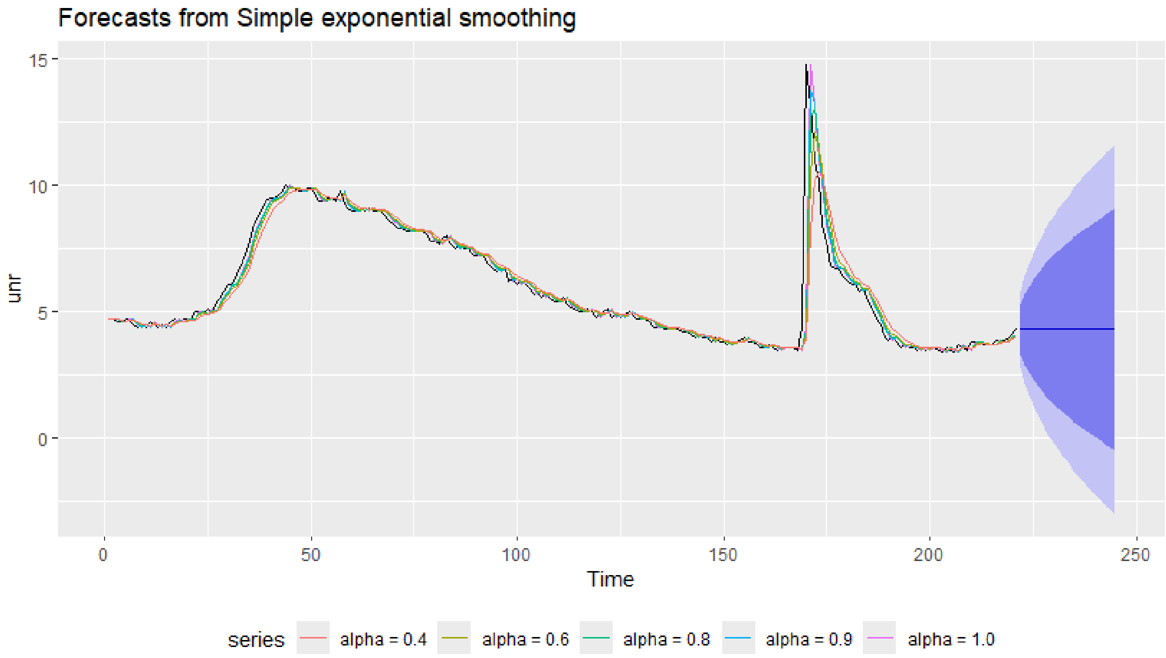
Autocorrelation and Partial Autocorrelation functions of the residuals from ARIMA(0,1,2):



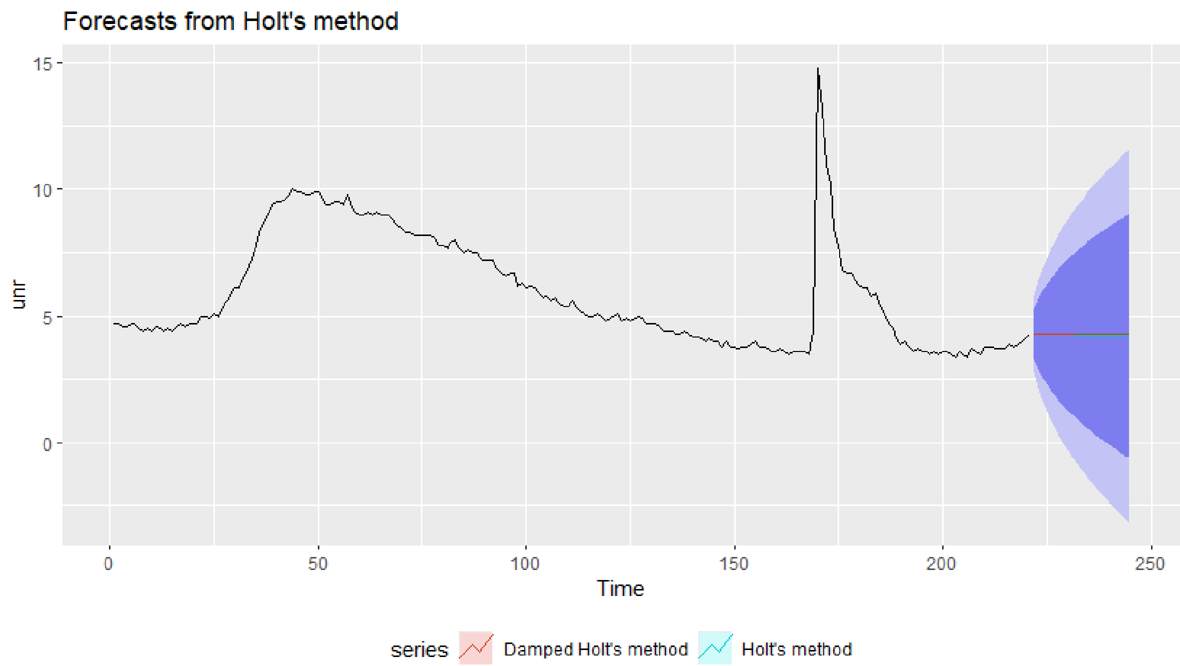
These plots show us that residuals are white noise, and no further improvement of the model is required.

Exponential Smoothing Models

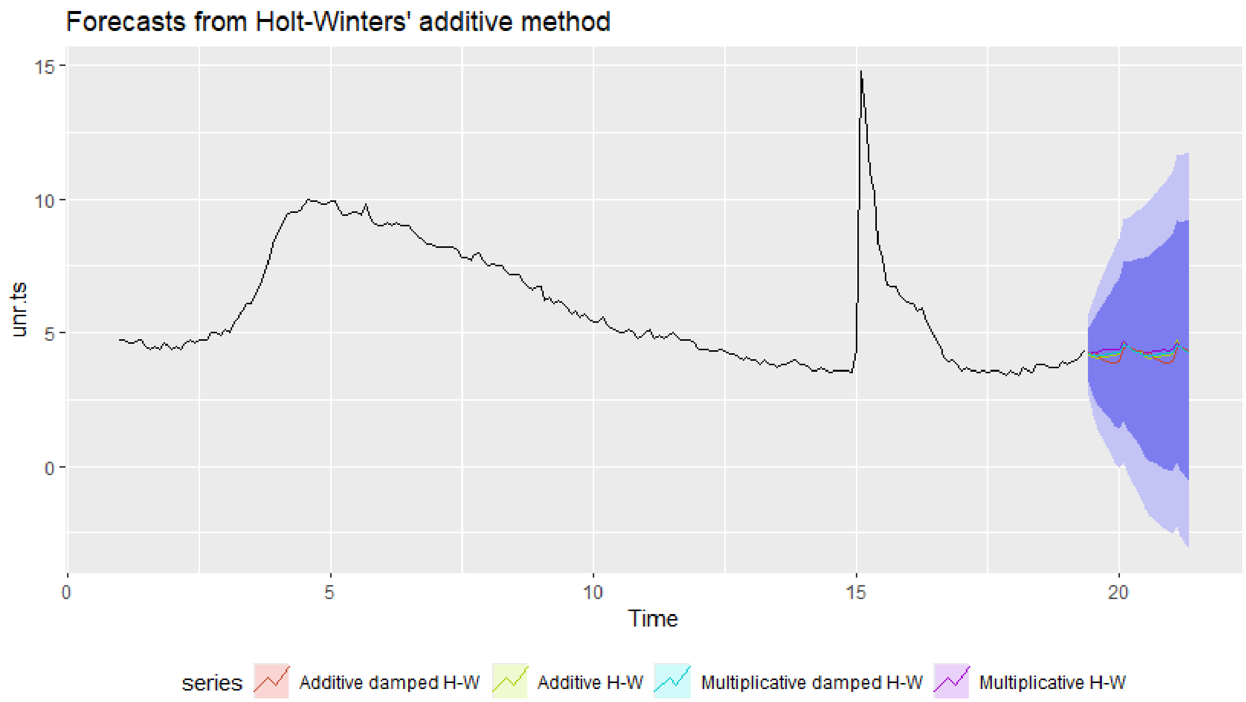
In this part, we present Exponential Smoothing models. First, we tried the Simple Exponential Smoothing (SES) model. The SES model was tested with different smoothing parameters (α) set to 0.4, 0.6, 0.8, 0.9, and 1.0. All these models fit the data well, but they all produce flat, horizontal lines as expected. The best fit for SES is achieved with $\alpha = 1.0$, yielding RMSE of 0.7612. We used 24 month forecast horizon here and for both Holt's and Holt-Winters' models as well.



Next, we tested both the basic Holt model and the damped version. The basic Holt model performed slightly better, with an AIC of 1082.427 and an RMSE of 0.7613. Smoothing parameters automatically selected by the function: $\alpha = 0.9999,\ \beta = 0.0001$. These values were chosen by the model, indicating a near-zero trend and a very small smoothing factor for the trend component.



For the Holt-Winters' models, we explored four variants: additive, additive damped, multiplicative, and multiplicative damped. Among these, the additive model provided the best fit, with smoothing parameters $\alpha = 0.9999, \beta = 0.0001, \gamma = 0.0001$, along with an RMSE of 0.7447. The AIC for this model was 1096.679. As with Holt's model, the smoothing parameters were automatically chosen by the function, reflecting the lack of trend and seasonality in the data.



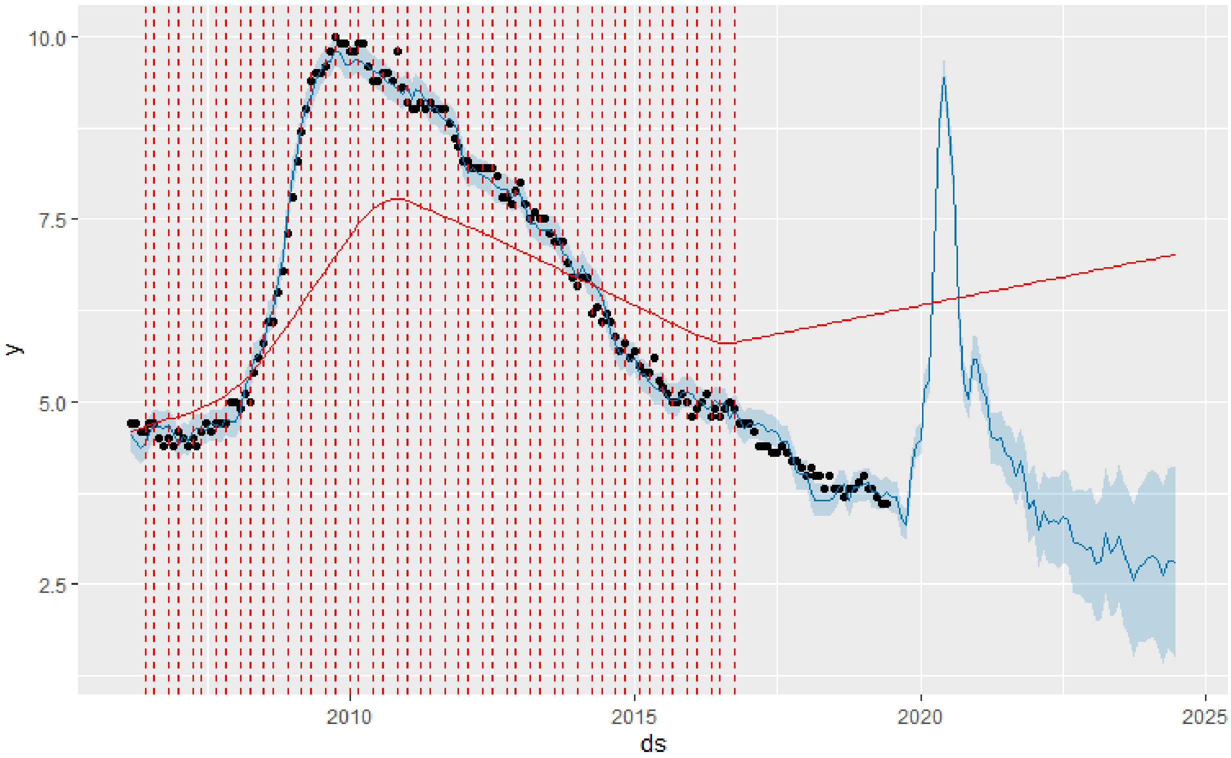
In summary, the SES, Holt, and Holt-Winters models all return consistent results with near-zero trend and no seasonality. Such results confirm that the data does not exhibit significant trends or seasonal patterns. This is the reason why models such as SES, Holt, and Holt-Winters are not effective in capturing meaningful patterns in this data.

Prophet Model

Next, we applied the Prophet model to predict the unemployment rate, experimenting with both the additive and multiplicative seasonality modes. The additive mode provided better results. We also tested various numbers of changepoints to evaluate their effect on the model's performance. In general, increasing the number of changepoints improved the model's ability to capture future trends. We tested different values, including 0, 15, 25, 50, and 100 changepoints, and found that 50 changepoints provided a satisfactory balance between model complexity and accuracy. The trend parameter in the model was chosen to be linear. The data is not sales data and therefore probably does not have any saturation point, so logistic trend is unapplicable here. A flat trend also would not make a good fit, as the unemployment rate usually fluctuates due to various factors.

Moreover, we incorporated several external regressors in the Prophet model to improve the analysis by capturing possible external influence. These regressors included all our additional data: the federal funds effective rate, the consumer price index, the civilian labor force level, GDP, real GDP, average weekly hours, average hourly earnings, GDP growth, the number of COVID-19 cases and the S&P 500 index (SPX).

For this analysis, we trained the model on the first 160 data points (up to June 1, 2019). The displayed plot depicts this model with additive seasonality and 50 changepoints:



While adding more changepoints helped capture future trends more accurately, it was noted that the height of the COVID-19 peak was predicted to be much lower with multiplicative seasonality compared to the additive seasonality. Increasing the number of changepoints led to narrower confidence intervals, though it did not significantly improve the model's mean predictions, so we decided to stop on 50.

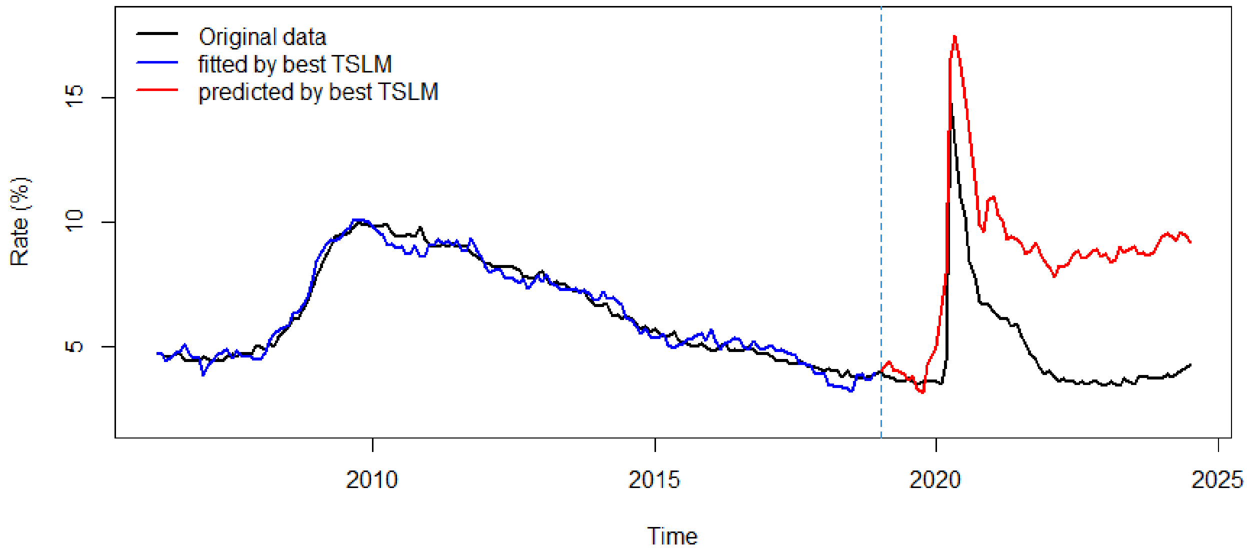
Overall, the Prophet model provided reliable results and showed strong predictive performance on unseen data. The predicted unemployment rates closely mirrored the actual values, indicating that the model is well-suited for this time series and capable of capturing key trends effectively.

TSLM

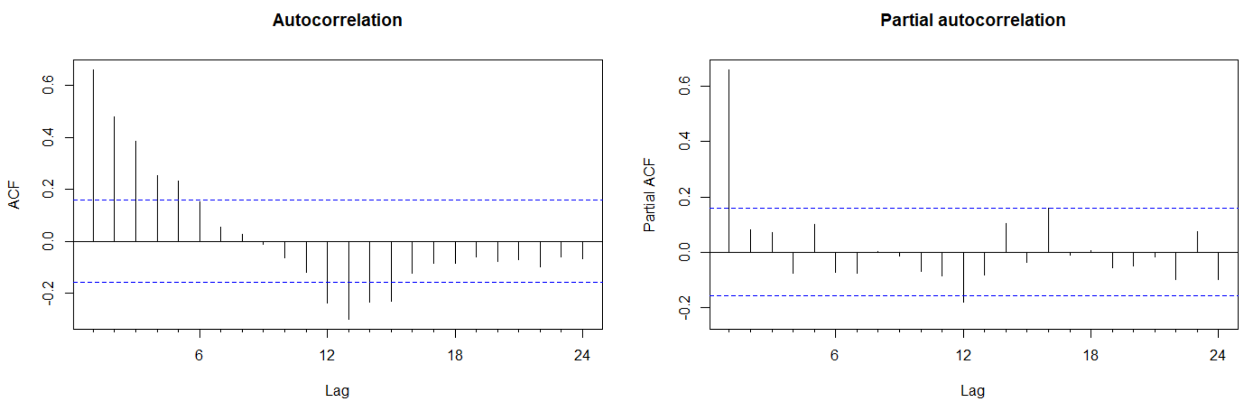
In this part, we tried to model our data with TSLM (Time Series Linear Model). We split the data into training and test sets, with the training data covering the period up to January 1, 2019, and the test data starting from that date onward. Initially, we built a TSLM using all available external regressors: the federal funds effective rate, CPI, civilian labor force level, GDP, real GDP, average weekly hours, average hourly earnings, GDP growth, and the S&P 500 index (SPX). Additionally, we included trend and seasonality as explanatory variables. The number of COVID-19 case data was not included as a regressor, as it was not present in the training set (prior to 2019). However, the influence of COVID-19 in the test data was likely captured indirectly through the other regressors, as it influenced the economy greatly.

The initial model, which included trend and seasonality, had an AIC of 155.6776. To optimize the model, we performed stepwise feature selection by removing regressors one by one, based on their p-values ($\Pr(>|t|)$). If a regressor had a p-value higher than 0.05, it was excluded from the model. After this process, the final model had an AIC of 130.7255. The features removed were seasonality, federal funds rate, real GDP, SPX, average weekly hours and the intercept. Most probably those features were in correlation with some other feature, so they did not have statistical significance. For example, real GDP obviously correlates with GDP. All the remaining features are statistically significant.

This optimized (best) TSLM was then used to predict the test data. This plot shows the predictions of best TSLM:

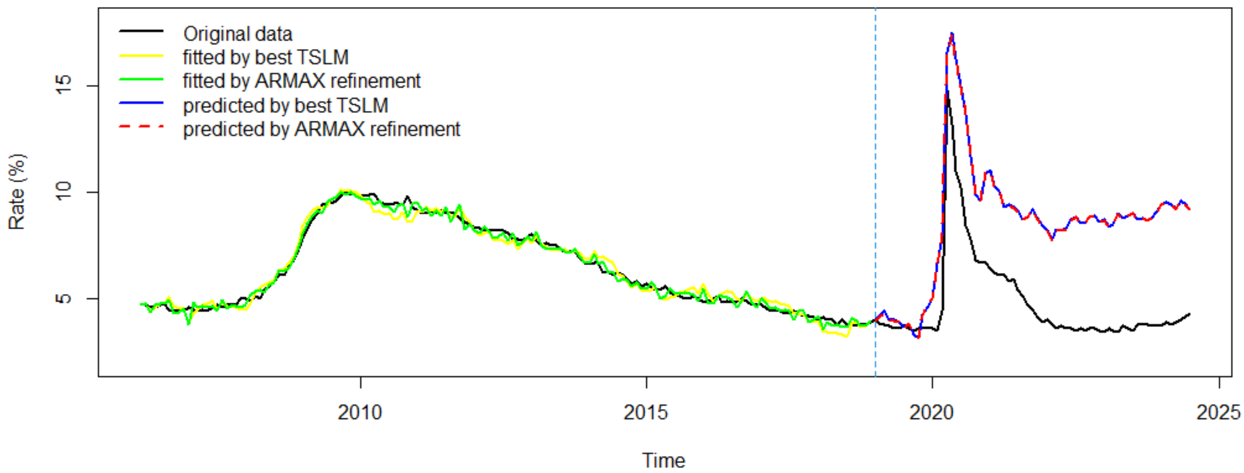


However, an analysis of the residuals showed that they were not white noise, indicating the presence of unexplained structure in the data:

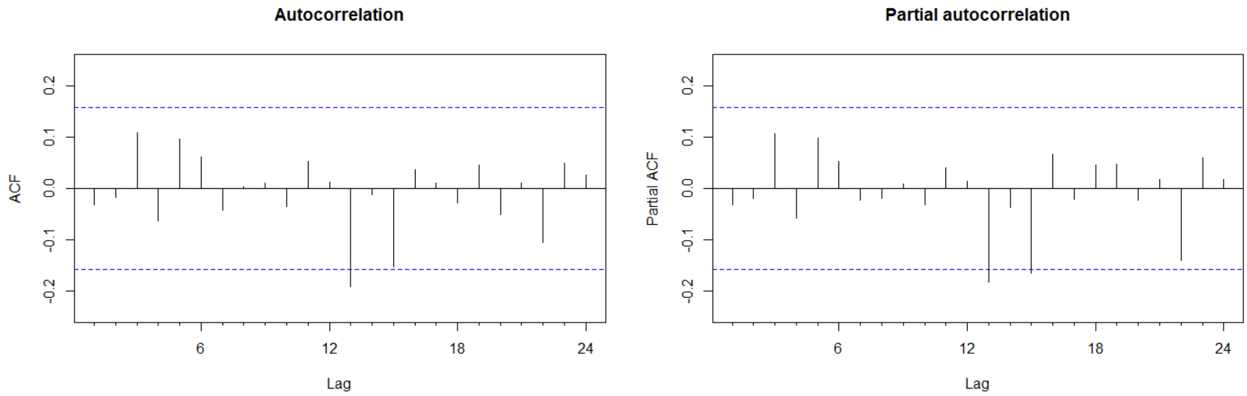


To fix this, we refined the model using ARMAX. We used the *auto.arima* function, which selected an ARIMA(1,0,0)(1,0,2)[12] model. This model reduced the AIC to 29.94. It captures both seasonal and non-seasonal components. Specifically, the non-seasonal part (1,0,0) indicates that the unemployment rate depends linearly on its value from one month ago, and the seasonal part (1,0,2)[12] shows also seasonal behavior, taking into account patterns that occurred 12 months ago.

We used the refined TSLM model for predictions on the test set. The difference between the predictions of the initial and refined models was minimal, with the mean squared error (MSE) of the refined model being 17.51279. The final model captures the dynamics of the unemployment rate, but not too well and there is still room for improvement.



The residuals of the refined model are white noise, so all the patterns in the data were captured successfully:



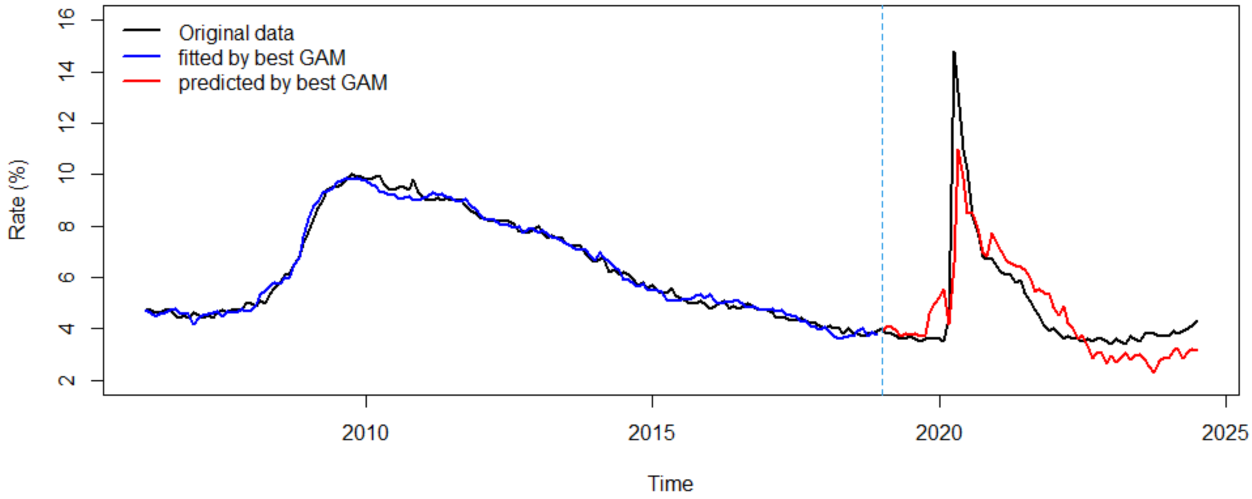
GAM

Furthermore, we applied a Generalized Additive Model (GAM) to predict the unemployment rate, following a pipeline similar to the one used for TSLM. The data was split into train and test in the same way, with training data covering the period up to 2019 and the test data covering the time period since January 2019. The initial baseline model included all the available external regressors: the federal funds effective rate, CPI, civilian labor force level, GDP, real GDP, average weekly hours, average hourly earnings, GDP growth, and the SPX. As with the TSLM pipeline, we did not include the data on the number of COVID-19 cases since it is unavailable in the training set (pre-2019). Additionally, we modeled the dependency on time (trend).

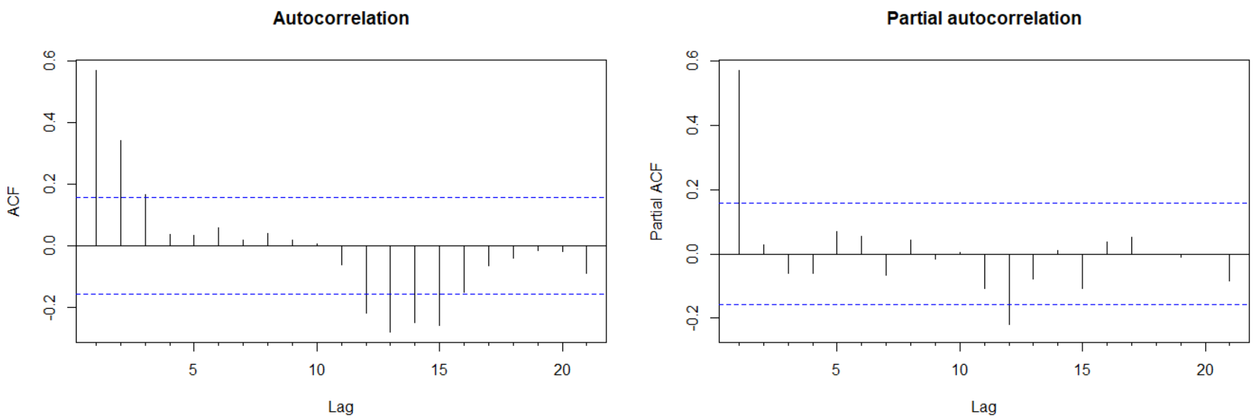
The baseline model achieved an AIC of -29.93031. To optimize the model, we performed stepwise selection by removing regressors one by one based on their p-values, removing those with p-values greater than 0.05. We had to exclude only average weekly hours and managed to obtain an improved model with an AIC of -31.13073.

Next, we focused on the selection of the smoothing degree. We tested degrees in the range from 1 to 20 on the test data and evaluated the models based on mean squared error (MSE). The best smoothing degree was found to be 3 (while the default is 4).

Predictions by best GAM (here smoothing degree is 3):

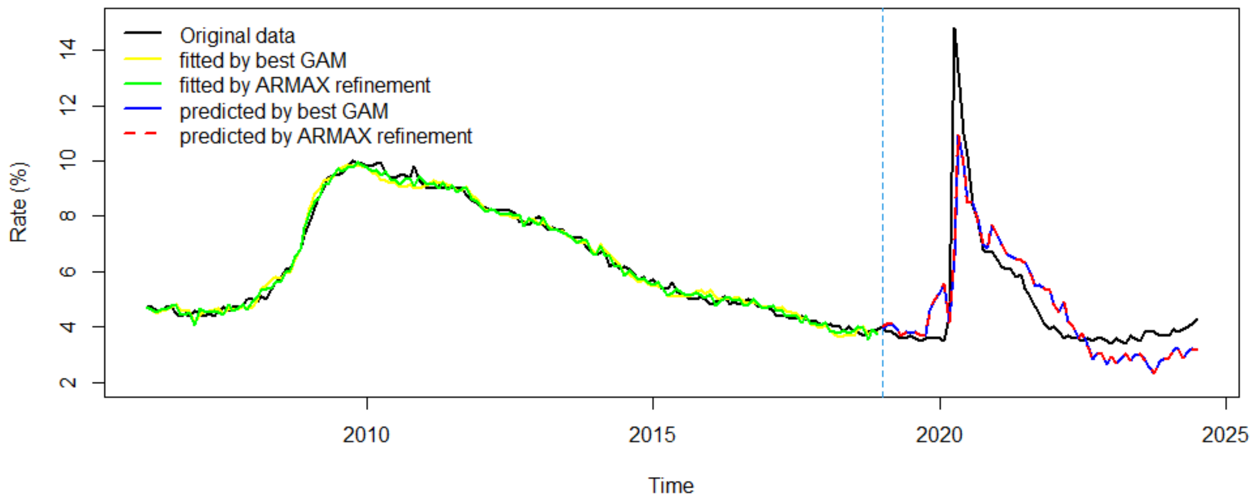


An analysis of residuals showed they were not white noise:

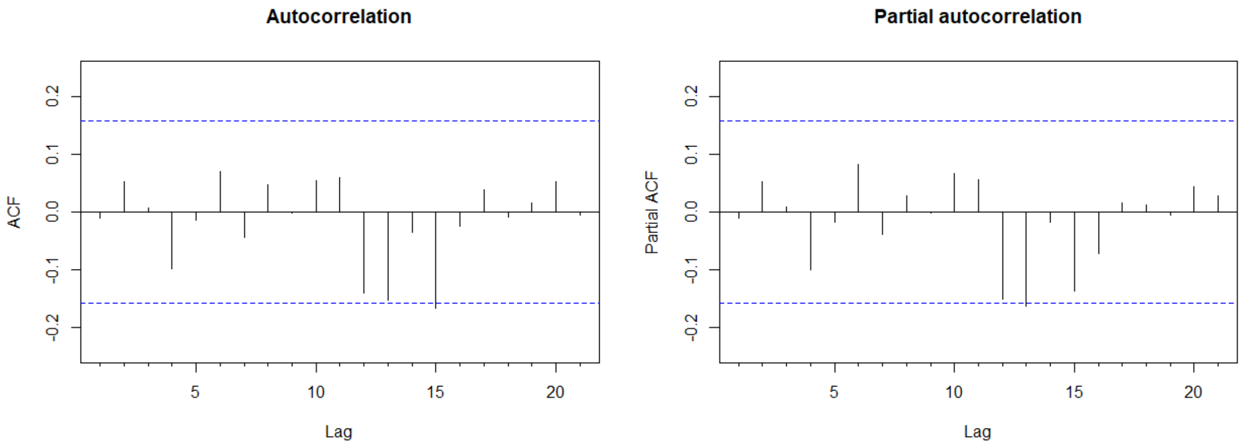


To refine the model with ARMAX, we applied *auto.arima* again, which selected an ARIMA(1,0,0) model for the residuals with AIC equal to -89.43. This model again captures the dependency on the previous month, but no seasonal dependencies. The MSE of the refined model on test data is 1.773305.

Comparison between the predictions of the baseline and the refined models:



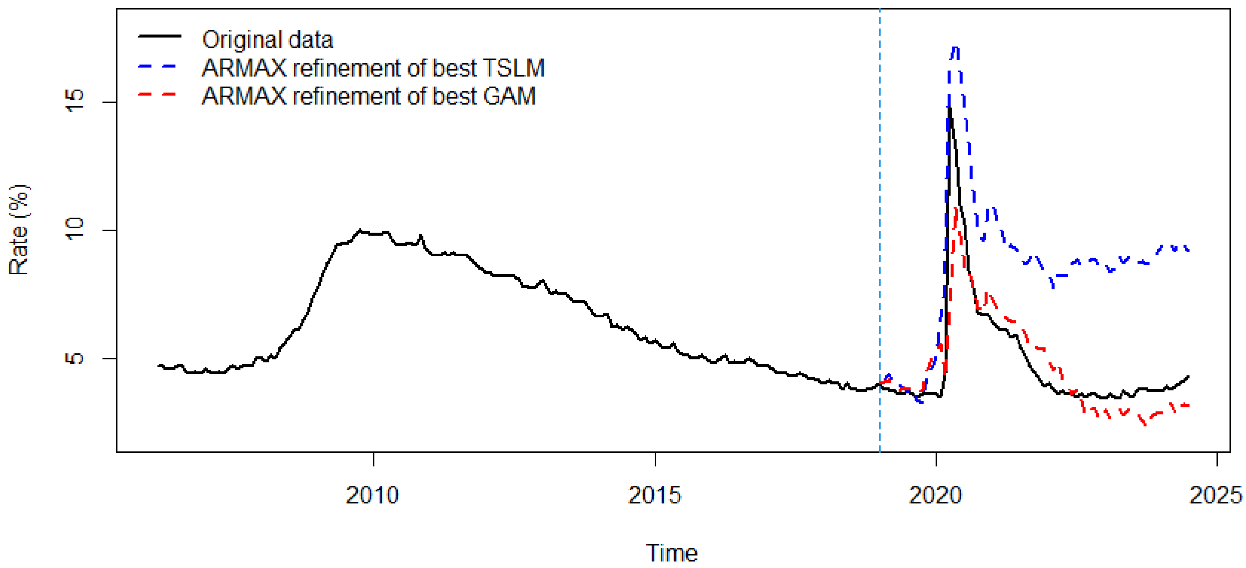
Again, the difference in predictions is barely visible, but refinement ensures that no patterns were left in the residuals:



Now, the residuals are white noise.

TSLM and GAM comparison

Moreover, we performed a comparison between predictions of TSLM and GAM. For this purpose, we used the final refined version of the models. Both models were trained on the same data, so the comparison should be fair enough:



As we can see from the plot, ARMAX refinement of best GAM predicts the unemployment rate much better. This can also be noticed from their MSE results: MSE of refined TSLM is 17.51279, which is 10 times greater than the MSE of refined GAM, which is only 1.773305. Therefore, we can conclude that GAM is more suitable in this case.

GBM

Additionally, we also tried to apply a gradient boosting model to predict the unemployment rate. As before, we split the data into training and test sets, with the training data covering the period up to January 1, 2019, and the test data starting thereafter.

To find the best-performing GBM model, we conducted a grid search over the following hyperparameters:

- Number of trees: 500, 1000, 2000, 5000
- Tree depth: 1, 3, 5, 7, 10
- Learning rate (shrinkage): 0.001, 0.005, 0.01, 0.02, 0.05, 0.1
- Minimum observations per node: 3, 5, 7, 10, 12, 15

For reproducibility, we set a fixed random seed before fitting each model. This ensured consistent results across runs and allowed for a fair comparison between models.

The hyperparameters of the best model are the following:

- Number of trees: 500
- Tree depth: 1
- Shrinkage: 0.05
- Minimum observations per node: 15

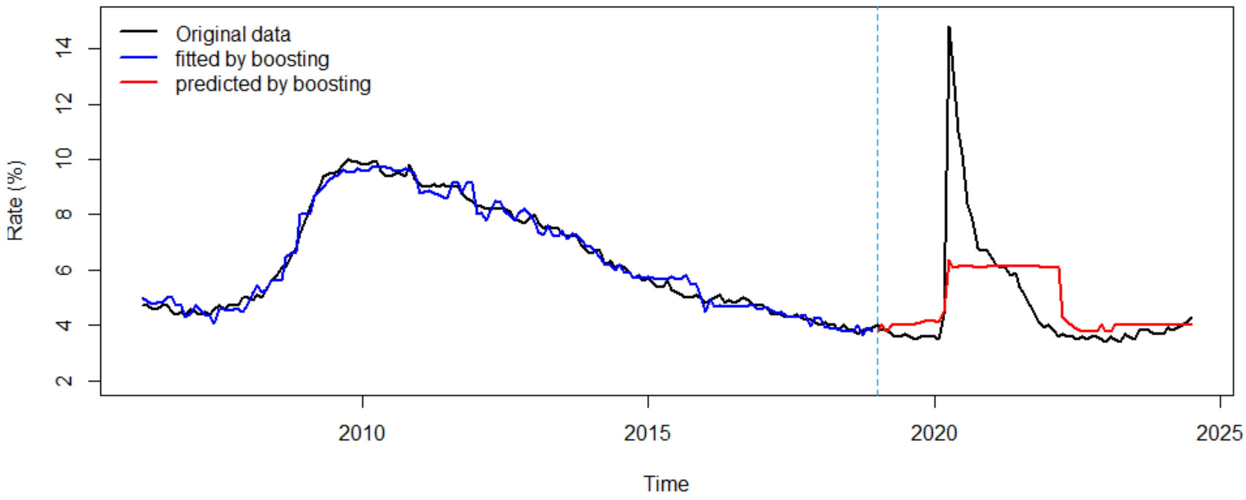
On the test set this model achieved an MSE of 3.103044. The best predictions were obtained by limiting the model to the first 117 trees. This suggests that the model might overfit if too many trees are used. Also, we can see that in this case shallow trees (with depth of 1) allow us to achieve better results.

Influence of the regressors in the best model:

Variable	Relative influence
Federal funds effective rate	50.0487646
SPX	33.6661442
Average hourly earnings	6.2699607
CPI	4.1561656
GDP	2.1259952
Average working hours	1.6144796
Civilian labor force level	0.9937536
GDP growth	0.8601526
Real GDP	0.2645839

This table shows that the most influential regressors are the federal funds effective rate and SPX, average hourly earnings, CPI, GDP and average working hours also have some effect and the others' influence is lower than 1%. Likely the boosting model finds high correlation between the effective rate and the unemployment rate, and between the SPX and the unemployment rate as well.

However, the model performs poorly on the test data:



Although the MSE of the model is quite low, it fails to predict the COVID-19 peak and overall, the pattern of the shock caused by the pandemic. This might be due to the fact that gradient boosting models are not designed for sequential or time series data. Boosting model treats each observation in the data as independent and does not explicitly account for the sequential or temporal nature of the data. In time series, values at one time step are typically influenced by prior time steps, particularly during abrupt changes like the COVID-19 peak.

Conclusion

In this project, we tried different tools for time series analysis presented in the course. However, we left the Bass Models out, because we do not think the unemployment rate can be modelled with a diffusion process. Bass Model requires the assumption that the product, which is modelled, has a limited life cycle, so eventually it should decline to zero. However, the unemployment rate would probably never go to zero (unfortunately) and overall, the unemployment rate does not have any clear upper or lower limit. Also, unemployment is not exposed to imitation dynamics. Therefore, Bass Models are not suitable in this case.

Among the models considered, GAM showed probably the best ability to model the unemployment rate, given additional external data. TSLM also showed substantial results, but its accuracy in predictions is limited, probably due to the simpler nature of this model. Moreover, the prophet model produces results of significant quality as well. Exponential smoothing based models could not cope with this task, because of the nature of our data. These models are more suitable in the case of expressed seasonality or trend, which we do not have in our data. ARIMA models do not predict well on their own but perfectly perform the refinement of other models. And finally, gradient boosting models also did not manage to perform the task with good quality, again, because of being originally designed for other purposes.