

Analysis of U.S. Unemployment Rate (1948-2023)

Sarah Lew (sarahlew@ucsb.edu)

2024-05-29

Contents

| | |
|---|-----------|
| Abstract | 1 |
| Introduction | 2 |
| Data | 3 |
| Methodology | 4 |
| Method 1: SARIMA(p,d,q)x(P,D,Q) model | 4 |
| Method 2: Spectral Analysis | 4 |
| Results | 5 |
| Results from SARIMA (p,d,q) x (P,D,Q) | 5 |
| Results from Spectral Analysis | 7 |
| Conclusion and Future Study | 9 |
| References | 10 |
| Appendix | 11 |
| Appendix A: | 11 |
| Appendix B: | 11 |
| Appendix C: | 11 |
| Appendix D: | 12 |
| Appendix E: | 14 |
| Appendix F: | 15 |

Abstract

The U.S. unemployment rate is a valuable metric that reflects the changing labor market dynamics and its broader implications for other economic trends. We use data from the U.S. Bureau of Labor Statistics to uncover seasonality patterns, apply data transformations in ARIMA modeling, and utilize spectral analysis on unemployment rate data from 1948 to 2023. The project identifies a model to forecast the rate for future points, which is a valuable resource for strategic planning and decision making during times of major economic change (specifically the COVID-19 pandemic). Our results indicate that we can successfully identify the appropriate SARIMA model for forecasting and understand the underlying seasonality of unemployment rate.

Introduction

Unemployment rate is a significant measure of economic health as changes in rate can provide insights on changes in the labor market and socio-economic conditions. This project focuses on understanding and predicting unemployment rate in the United States from 1948 to 2023 using time series analysis. By having a better understanding on past unemployment trends in relation to changes over time and external factors, we can uncover long term trends that can be found about the U.S. labor market. Time series analysis techniques can provide further understanding of variations of unemployment rate, and forecasting can aid those such as economists or businesses in planning.

It is important to note that major events that have impacted the economy during this time, such as the COVID-19 pandemic, which can affect how the model characterizes changes in rate. Another notable economic event, the global financial crisis in 2008, has also had a significant impact on labor markets and thus unemployment rates. Events such as these have resulted in disruptions on the economy such as unprecedented closures and job losses.

The data collected from 1948 to 2023 still showcases a period of various policy shifts, recessions, and worldly events that can be used to understand underlying patterns and trends. Researchers can use this information to determine proper measures to mitigate and recover from economic declines. Specifically, policy makers can use unemployment forecasts to determine methods that will reduce future job losses and also provide resources for relief measures. Those who are interested in unemployment forecasts can use this information to make informed decisions on how to navigate the labor market for themselves and their career choices.

In this project, I will be implementing two time series methods to analyze the dataset, including the SARIMA model and Spectral Analysis in RStudio. Specifically, I will be applying the Box-Jenkins Approach for the ARIMA model. For the first model, I will also be predicting 12 points ahead to understand implications of such models on the U.S. Unemployment rate. The goal of this project is to have a holistic understanding of the unemployment rate over time, which will be valuable in making optimal economic decisions.

Data

The dataset used in this project is the “U.S. Unemployment Rate Data,” which covers historical data published from the U.S. Bureau of Labor Statistics. It ranges from Jan. 01, 1948 to December 2023. There are 912 observations, and the frequency is monthly data. I will be analyzing the units as U.S. unemployment rate, which are positive values. The data was collected by <https://fred.stlouisfed.org/series/UNRATENSA>, which is the Federal Reserve Economic Data (FRED). Specifically, the series is from the ‘Current Population Survey (Household Survey)’ that was conducted.

This particular dataset is significant as it captures a period of time that is marked by various changes including recessions, policy changes, and also the 2020 pandemic. In addition, it covers other significant times in U.S. economic history from times of stability, financial crisis (2008) and post recovery periods. Such diversity in economic phases will provide a comprehensive exploration of factors on unemployment patterns. Analyzing the data will allow for uncovering trends on how these events played a role in unemployment over these past decades.

The two variables used in the dataset are `DATE` and `UNRATENSA`.

- `DATE`: the day, month, and year of the specific corresponding unemployment rate from 1948
- `UNRATENSA`: representation of the number of unemployed as a percentage of the labor force.

The labor force data is restricted to those who are over sixteen years old, within the 50 states or District of Columbia. It also excludes those who are in institutions or are on active duty.

When the unemployment rate is high, that indicates a time where a large percent of the labor force cannot find employment despite actively seeking. Economists typically consider a rate above 5% to be relatively high in the US economy, but this can differ depending on various conditions. On the contrary, if the unemployment rate is low, it marks a period of positive economic health, increased consumer confidence and economic growth.

Below we can observe the first 10 observations from their respective months and years.

Table 1: Dataset Observations

| DATE | UNRATENSA |
|------------|-----------|
| 1948-01-01 | 4.0 |
| 1948-02-01 | 4.7 |
| 1948-03-01 | 4.5 |
| 1948-04-01 | 4.0 |
| 1948-05-01 | 3.4 |
| 1948-06-01 | 3.9 |
| 1948-07-01 | 3.9 |
| 1948-08-01 | 3.6 |
| 1948-09-01 | 3.4 |
| 1948-10-01 | 2.9 |

For the analysis, we will first read in the dataset and rename columns for use. To tidy the data, I have removed all missing values and converted the date variables given into a value that can be used for the time series. In addition, I extracted the year and month values to be used for our analysis. We can now create a time series from the data frame that will be used for our two methods.

Methodology

The two methods applied will be SARIMA and Spectral Analysis. After interpreting the results and deriving a conclusion, I will forecast the next 12 data points using SARIMA.

Method 1: SARIMA(p,d,q)x(P,D,Q) model

We will utilize the Box Jenkins Approach when applying the SARIMA(p,d,q) model. This consists of first plotting the time series original data, transforming the data using Box-Cox or log transform, and then interpreting the ACF and PACF plots. To achieve stationarity, which is critical for applying time series analysis techniques, we will transform the data.

To understand the ACF or PACF plots, we will analyze significant points and determine if there is any seasonality present. Specifically, the ACF plot will reveal whether or not the model estimated could be a moving average MA(q) model; the PACF plot indicates if the model can be autoregressive AR(p). An autoregressive (AR) model will incorporate past data points to predict within the same time series. The moving average (MA) model is a smoothing technique that applies the averages of points in the time series to highlight long term trends.

An example of MA(q) is when the ACF plot drops off after lag q, and this is the same for AR(p) when considering a PACF plot after lag p. To consider an ARMA(p,q) model, there will be gradual decay in both the PACF and ACF plots.

For non-seasonal time series data, we will consider (p,d,q) and for seasonal data we will also consider SMA(Q) and SAR(P) when fitting the model. The difference (d) will be equal to 0 if no differencing has taken place, and will be equal to 1 if the data has been differenced once. However, the seasonal difference (D) will be 1 if we seasonally difference it. We will then estimate the parameters used for the model ARIMA(p,d,q) or SARIMA(p,d,q)x(P,D,Q) if there is seasonality. Using these results, we will plot model diagnostics based on residual plots.

The diagnostic plots to consider are: standardized residuals, ACF of residuals, Normal Q-Q plot of standardized residuals, and p-values for the Ljung Box Statistic. To analyze these, we will first consider the standardized residuals and determine if it appears stationary. As for the ACF of residuals, we will determine if there are any significant spikes. The Q-Q plot will reveal if the residuals follow a normal distribution if they are along the line. The p-values for Ljung Box Statistic will test whether the residuals are independent or have an auto correlated relationship.

We will then select the best model through these diagnostic tests and through comparison of Akaike Information Criteria (AIC). The AIC is an estimator of prediction error, and a lower AIC value indicates a higher quality model.

Method 2: Spectral Analysis

Unemployment follows a cycle related to business and fluctuations in the economy. As an economy reaches a peak or experiences growth, the rate of cyclical unemployment will be low.

We will compute the raw and log spectrum with a periodogram. The periodogram will display the relative strengths of different frequencies that can explain the variation or cyclic behavior in the time series. It is a rough sample estimate of the spectral density, calculated using the Fourier Transform. After plotting these spectrums, we will observe any significant peaks that are apparent. We will also explore non parametric estimation smoothing using centered moving averages (Daniell Kernel) and tapering. I also computed the confidence interval of the top 3 dominant frequencies from the raw spectrum to examine any significance.

Results

Results from SARIMA (p,d,q) x (P,D,Q)

First, we will plot the original time series.

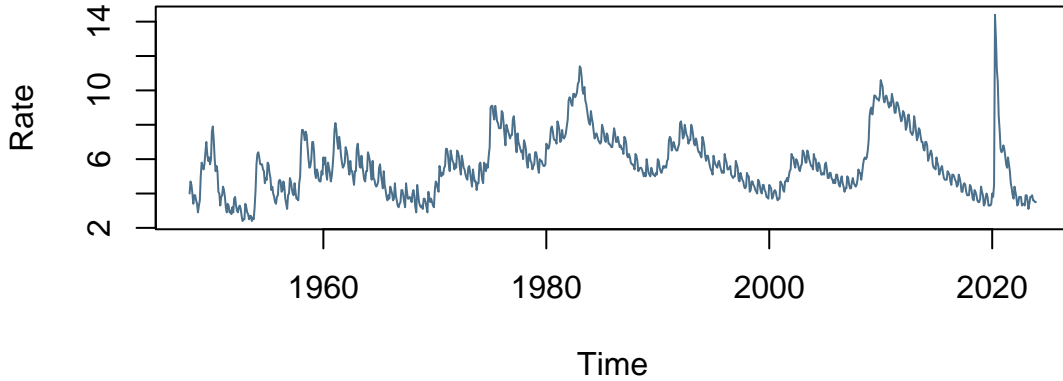
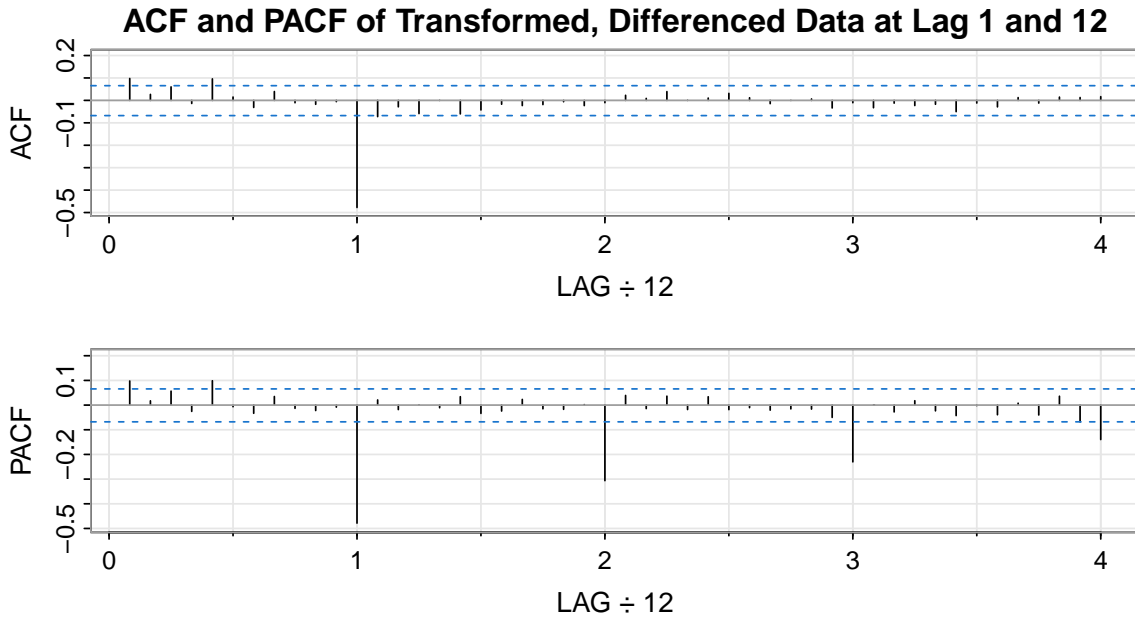


Figure 1: Unemployment Rate in US from 1948 - 2023

The plot has fluctuations and a sharp increase as there are changes in observations. I transformed the data using the log transform to stabilize the variance, and we can visualize the normality in Appendix A. The difference is $d = 1$, as we have taken the log difference of the original data to ensure stationarity.

After observing the ACF and PACF plot of the log differenced data in Appendix B, we can see there are significant spikes at multiples of 12. This highly suggests a seasonality, so we will now take a difference at $S = 12$ to remove the seasonal component. This will ensure a stationary series for the (S)ARIMA model to fit.



We can first examine the non-seasonal components. The ACF plot has some spikes at lags 1 and 5, though it cuts off after a significant peak lag 12. As for the PACF plot, there are also points outside the confidence interval at lags 1 and 5. The rest of the peaks are at larger lags of multiples of 12. We can then consider MA and AR models with $q = 1, 5$ and $p = 1, 5$ respectively.

Now, we can analyze the seasonal components of the plots. The ACF cuts off after lag 12, and we can assume SMA(1). The PACF decays at lags following multiples of 12 including 24, 36, and 48. We can then consider SAR(1), SAR(2) and SAR(3). Though there is a peak at lag 48, I have decided not to include it as it is not as strong as the others. Recall that $D = 1$ and $d = 1$, but we will be applying `sarima` to the transformed, differenced data.

We will consider models including: SARIMA(1,0,1) x (1,0,1), SARIMA (1,0,1) x (1,0,2), and SARIMA(5,0,1) x (1,0,3) on the already differenced data.

Below we can examine the diagnostic plots for the SARIMA(1,0,1) x (1,0,1)[12].

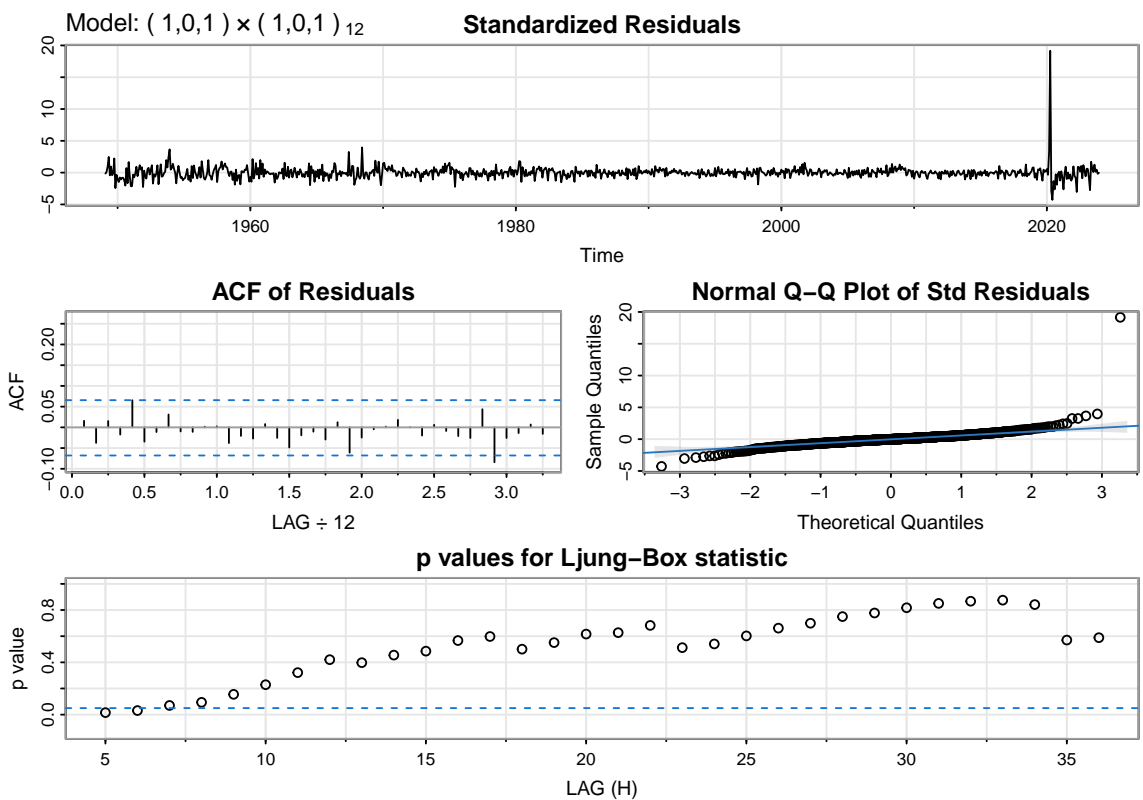


Figure 2: Diagnostic Plots

For this SARIMA(1,0,1)x(1,0,1)[12] model, there are no apparent patterns in the data and it is relatively stationary, with the exception of 2020. This makes sense as it corresponds to the recession during the COVID-19 pandemic, so the unemployment rate peaks significantly. The ACF plot tails off, and there are not any significant points of residuals. We can observe that the Q-Q plot appears to be relatively normal as it follows the straight line. There are outliers in the beginning of the plot and a significant outlier at the highest theoretical quantile, but it still reasonably follows a normal distribution. As for the p-values for the Ljung-Box statistic, there are two lower p-values that falls close to being statistically significant at earlier lags. Overall, as the lags increase, the p-values

increase as well and are greater than 0.05. This indicates the values are not autocorrelated.

Model Selection We can now decide which model to select for forecasting using AIC, AICc, and BIC. For the SARIMA(1,0,1)x(1,0,1)[12] model, the values are: AIC = -2.583814, AICc = -2.583739, BIC = -2.551769 . This model considers seasonality and has the lowest criteria values compared to all other models attempted. Another good potential estimate is the SARIMA (0,0,1)x(1,0,2) model, as it also has small AIC values. However, upon observation of its Ljung Box Statistic plot, the early p-values appear more significant.

Therefore, we will select SARIMA(1,0,1)x(1,0,1) model on the differenced and transformed data with the consideration of the comprehensive diagnostic plots and AIC value.

Forecasting Now we can use the estimated model on the original data and forecast the next 12 data points (months) in the future. Since we are using the original data with our SARIMA model, we will set $d = 1$ and $D = 1$ to account for differencing.

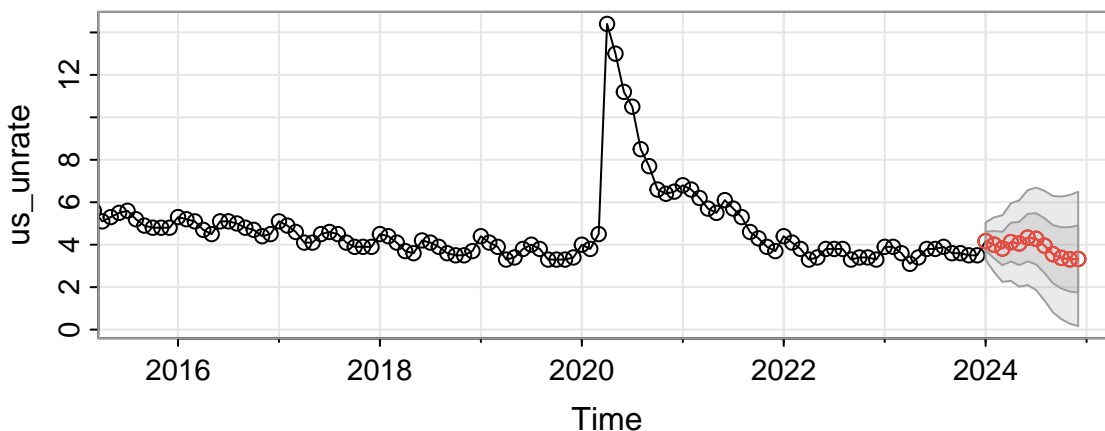
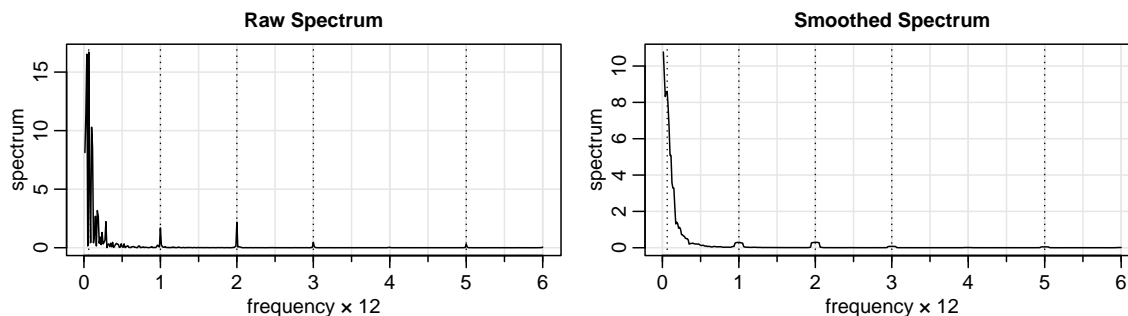
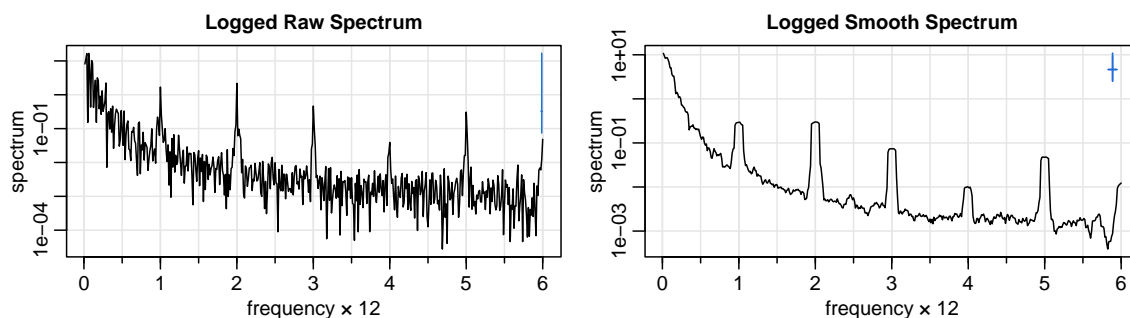


Figure 3: Forecasting US Unemployment Rate for Next 12 Months

Results from Spectral Analysis

We can plot the raw and smoothed periodogram of our original data. We will use a smoothing method for nonparametric estimation with the Daniell kernel of parameter m . I have decided to use the parameter 4, as it will create a smoothed value at time t . In addition, we will use tapering methods to which will reduce bias. I have chosen a value of 0.1, which means that 10% of the length of the time series at its ends will be tapered.





As we can see from the raw spectrum of the unemployment time series, there are many peaks close to a frequency of 0. There are significant frequencies that correspond to annual effects. For example, there is a notable peak at $\omega = 1\Delta = 1/12$ cycles per month (one cycle per year). Note that the frequency axis is scaled by a factor of 12. There is also a notable peak at frequency $2\Delta = 2/12$ cycles a month.

The minor peaks following these more significant ones can be found at $3/12$ and $5/12$, though they are not as obvious on the smoothed spectrum. These are the harmonics of the annual cycle $\omega = k\Delta$ for $k = 2, 3, 4..$, which typically occur when there is a periodic non-sinusoidal component.

Regarding the log and smoothed log periodograms, there is also an additional peak at 4Δ . This differs from the raw spectrum as it is not as prominent there. Upon analysis of the smoothed spectra, we can see that the sharper peaks are less prominent. Smoothing can reduce the noise in the periodogram, but causes these peaks to appear less pronounced.

Table 2: Dominant Frequencies in the Time Series

| Series | Frequency | Period | Spectrum | Lower | Upper |
|-----------------|-----------|---------|----------|--------|----------|
| Top Frequency A | 0.0625 | 16.0000 | 16.7107 | 4.53 | 660.0374 |
| Top Frequency B | 0.0375 | 26.6667 | 16.5201 | 4.4784 | 652.5091 |
| Top Frequency C | 0.0250 | 40.0000 | 11.3208 | 3.0689 | 447.1477 |
| Frequency 1 | 1.0000 | 1.0000 | 1.6882 | | |
| Frequency 2 | 2.0000 | 0.5000 | 2.1761 | | |
| Frequency 3 | 3.0000 | 0.3333 | 0.4645 | | |
| Frequency 5 | 5.0000 | 0.2000 | 0.3066 | | |

We can identify the first three dominant frequencies of the raw spectrum, and their respective periods (Frequencies A, B, C). I've also found the periods for the significant peaks as well as their respective frequency, spectrum, and confidence interval bounds as shown in Table 2. I also computed the confidence interval at the yearly cycle $1/12 = 80/960$, which is $(0.4576484, 66.68073)$.

From the top three frequencies, we can see that Frequency A is 0.0625 with a period of 16. Similarly, we can also see that Frequency B has a high period of 26. However, upon observation of their respective confidence intervals, we cannot establish significance between these top frequencies A, B, and C. This is because their spectrum values lie within the upper and lower confidence bounds of the other top frequencies. For example, the frequency A has a spectrum value of 16.17, which lies within the bounds of the second and third top frequencies.

Overall, the peak at 1Δ corresponds to a strong periodic component that repeats every 12 months (yearly) at a spectrum of 1.6882. This is reasonable as unemployment rates may follow an annual cycle as it exhibits some seasonal variation.

Conclusion and Future Study

Through the two methods, SARIMA and Spectral Analysis, we were able to forecast the next 12 months of U.S. Unemployment rate and explore its frequencies.

From the SARIMA model, we applied the Box-Jenkins methodology on the original data. This included plots from the original data as well as the ACF and PACF plots of the transformed, differenced data. In order to reduce the variability and seasonality, we transformed the data by log and took the difference at lags 1 and 12. Appendix A displays the histograms of the logged data and original data, showing that normality was reached after taking the log transform. We can see that the data also appears to be more stationary. Following this, we were able to estimate various SARIMA models based on analysis of the PACF and ACF plots. The ACF and PACF plots of the first differenced, log transformed plots are shown in Appendix C. Our results indicated that the best model with the lowest AIC and appropriate diagnostic plots was the SARIMA(1,1,1)x(1,1,1)[12] model. The model diagnostics of other estimated models are depicted in Appendix D. Note that if we are applying the `sarima` function on the transformed, differenced data, we will use SARIMA(1,0,1)x(1,0,1)[12]. However, we will set $d = 1$ and $D = 1$ when applying to our original data. This model had an AIC value of -2.583814.

With this SARIMA model, we predicted the next 12 data points for our original data. In this case, we found that the prediction appears accurate as the values correspond to the pattern of previous unemployment rates (shown in Appendix E). The model was able to capture the seasonal components of our dataset.

Using Spectral Analysis, we were able to determine frequencies that appear strong or significant using a raw and smoothed periodogram. We discovered that there were many strong frequencies close to 0, though these were not as significant as the peaks at 1 and 2.

We applied smoothing methods such as tapering and the Daniell kernel, which reduced the heights of peaks but also showed the most prominent ones. Through the periodograms, we found that many of the spectral densities were described by lower frequencies. However, there were significant peaks at frequencies of 1, 2, 3 and 5. This is reasonable as unemployment rates may typically follow an annual cycle from the start of the year to the holiday period at the end of the year.

To understand the significance of frequencies closer to 0, which were the highest peaks, I also computed the confidence intervals. This was useful in assessing their significance in relation to other peaks, as well as interpreting their periods and spectrums. Understanding these frequencies and cycles from the unemployment rate data is useful for policymakers and economists to understand future trends and to better allocate resources when faced with economic changes.

For future study, it would be intriguing to further refine the SARIMA model and add additional variables for analysis. Predictors such as demographic information or location would be helpful in forecasting accuracy as it would provide specific insights into unemployment rate for that particular category. In addition, further spectral analysis on specific economic crises and how those have impacted unemployment rate could provide key information as well.

Overall, this project utilized time series analysis through SARIMA and Spectral Analysis to forecast and uncover underlying seasonality in U.S. unemployment rate. These findings have significant implications on future planning in regards to the changing U.S. economy and its employment rates.

References

Shumway, Robert H., and David S. Stoffer. Time Series Analysis and Its Applications: With R Examples. Springer International Publishing, 2017.

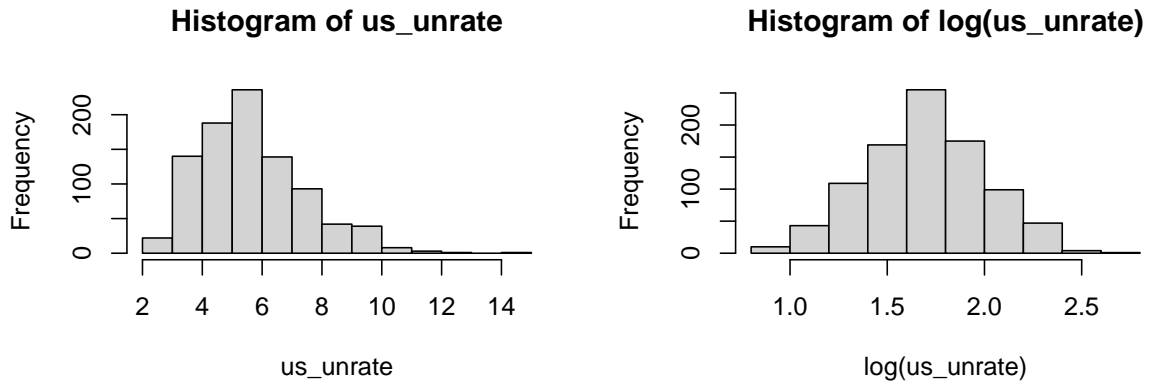
“Unemployment Rate (UNRATENSA) | FRED | St. Louis Fed.” Federal Reserve Economic Data | FRED, <https://fred.stlouisfed.org/series/UNRATENSA>. Accessed 7 June 2024.

“Time Series.” 1.5 Spectral analysis | timeseRies, <https://lbelzile.github.io/timeseRies/spectral-analysis.html>. Accessed 7 June 2024.

Appendix

Appendix A:

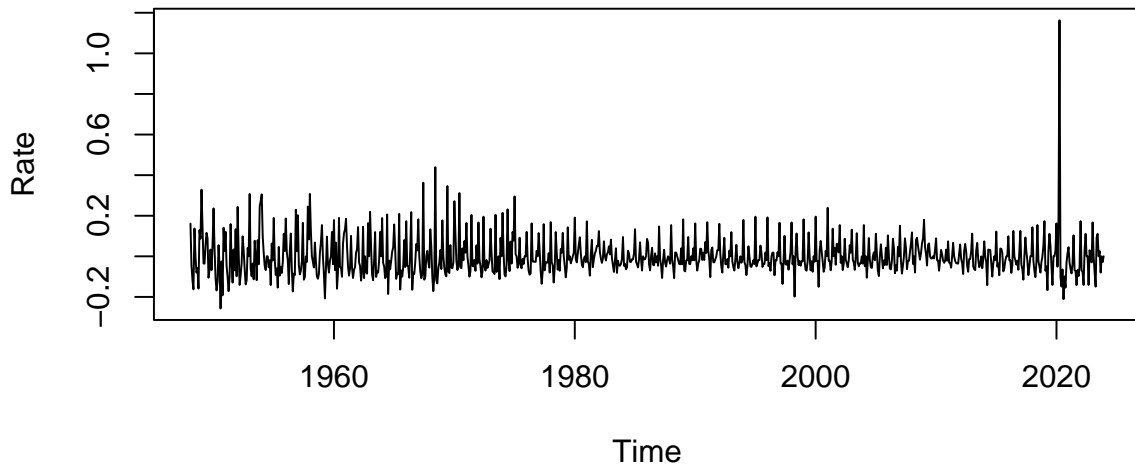
Histograms of Time Series, Histogram of Logged Time Series to show normality



Appendix B:

Plot of Transformed Data

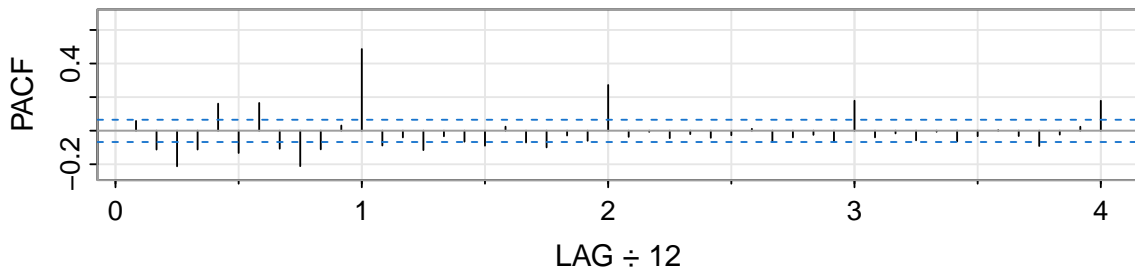
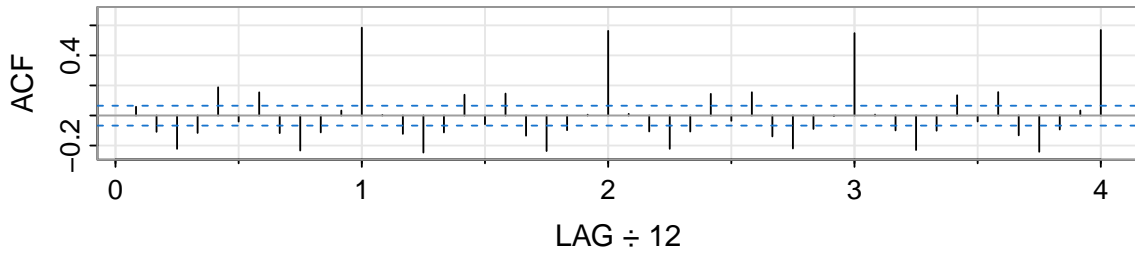
Transformed Unemployment Rate in U.S. from 1948 – 2023



Appendix C:

ACF and PACF of First Differenced, Transformed Plots

Series: transform_unrate

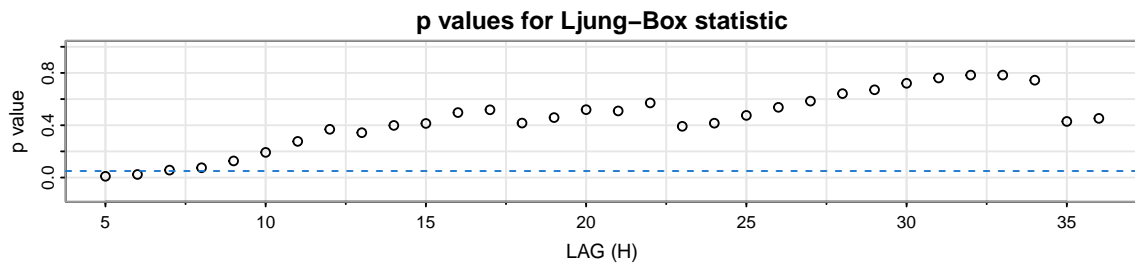
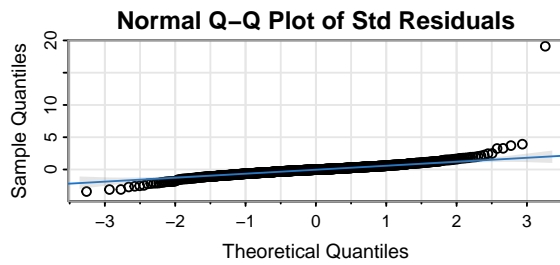
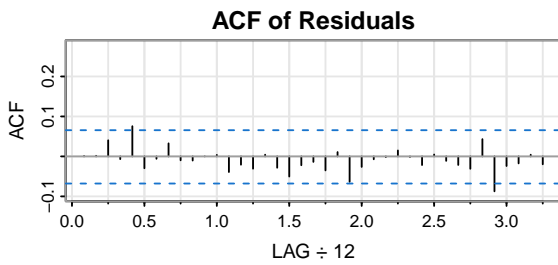
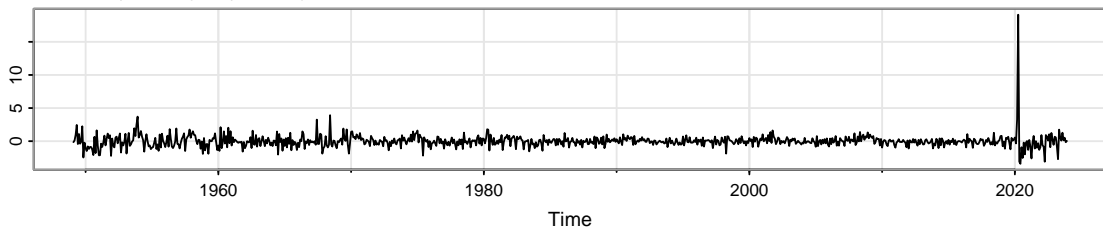


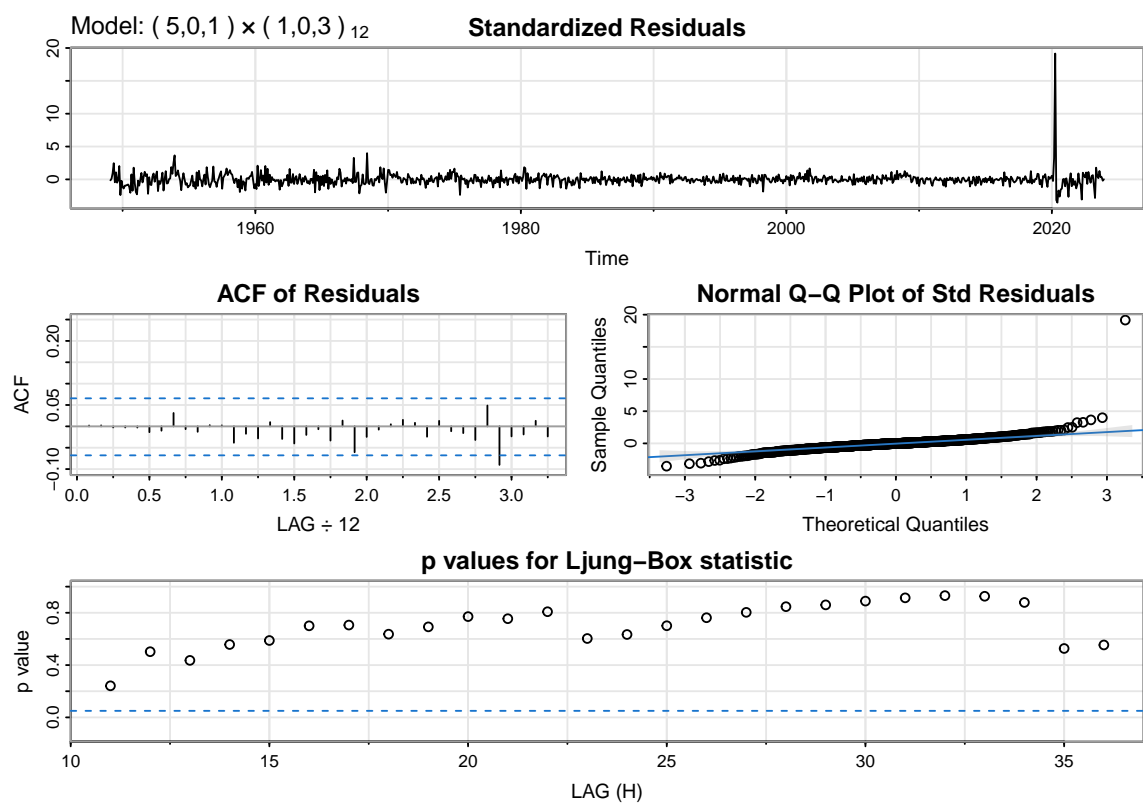
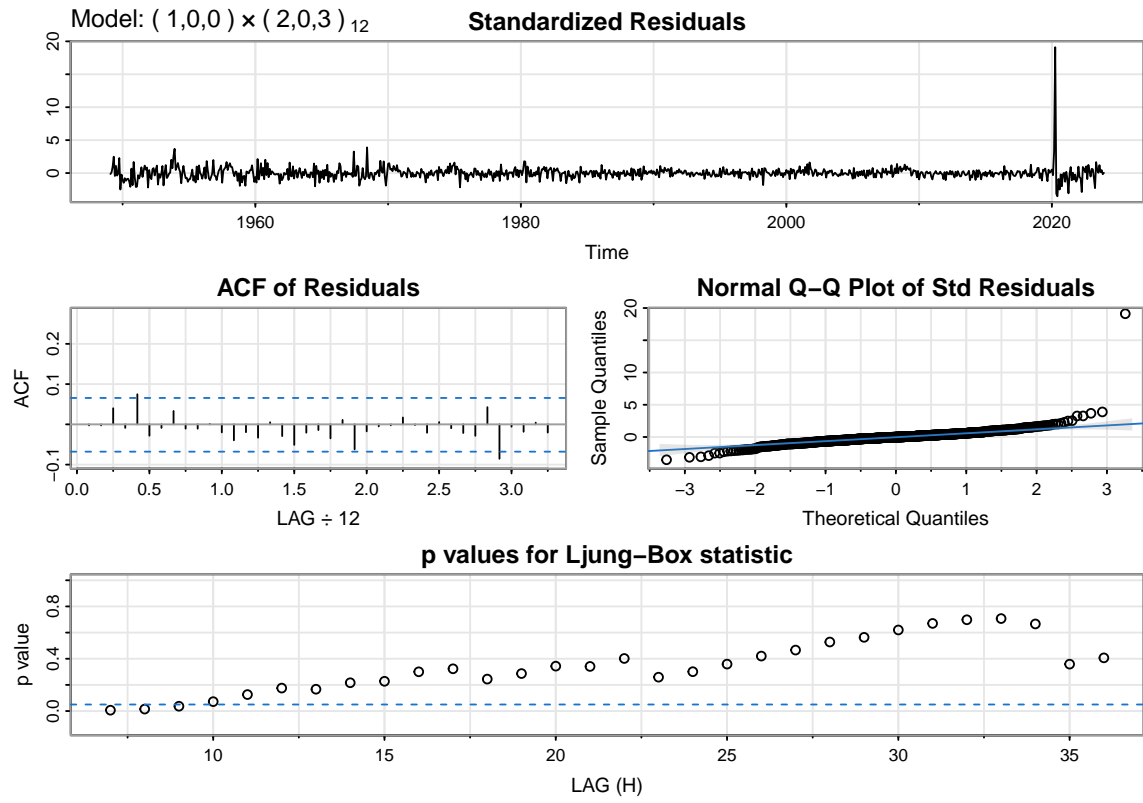
Appendix D:

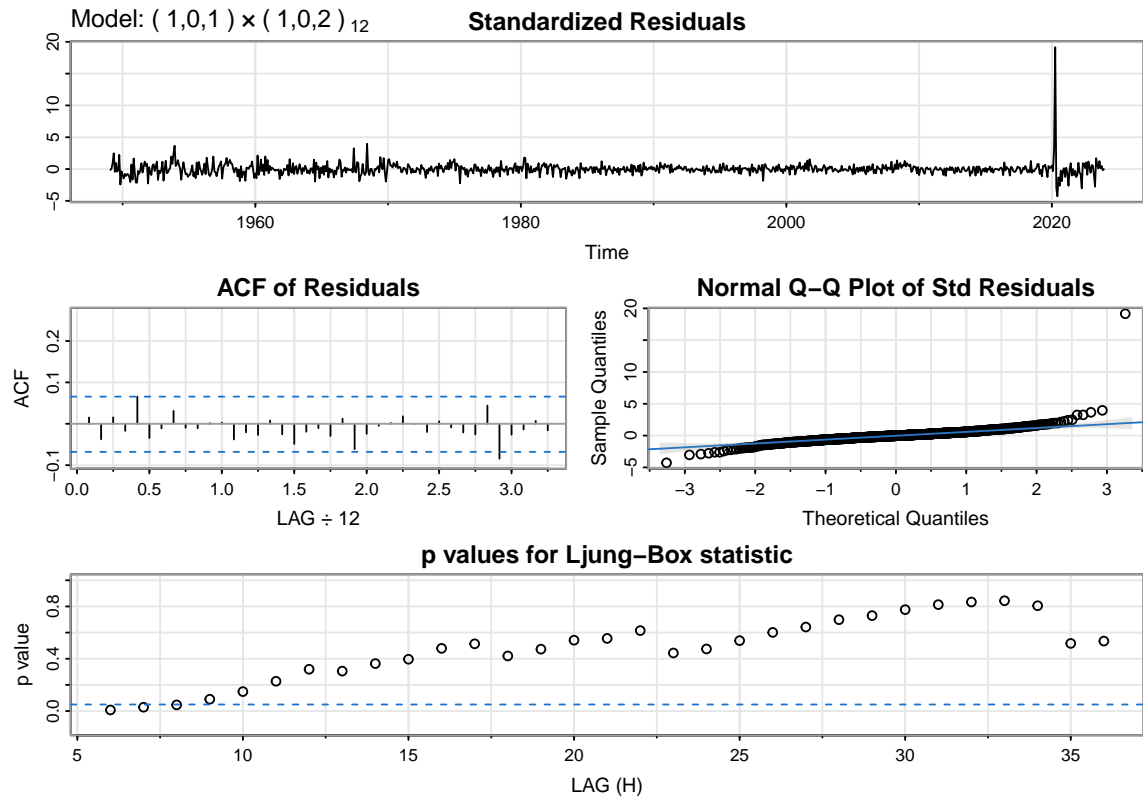
Fitting other potential models.

Model: $(0,0,1) \times (1,0,2)_{12}$

Standardized Residuals







Appendix E:

Predicted Future Data with SARIMA model

| ## | Prediction | PI.95..Lower.Bound | PI.95..Upper.Bound |
|-------|------------|--------------------|--------------------|
| ## 1 | 4.162717 | 3.305865 | 5.019568 |
| ## 2 | 4.080553 | 2.846718 | 5.314389 |
| ## 3 | 3.972031 | 2.474126 | 5.469937 |
| ## 4 | 4.325229 | 2.621126 | 6.029331 |
| ## 5 | 4.327137 | 2.454101 | 6.200173 |
| ## 6 | 4.631576 | 2.616295 | 6.646856 |
| ## 7 | 4.630576 | 2.493386 | 6.767767 |
| ## 8 | 4.387470 | 2.144500 | 6.630441 |
| ## 9 | 4.062761 | 1.727170 | 6.398352 |
| ## 10 | 3.949676 | 1.532421 | 6.366931 |
| ## 11 | 3.925423 | 1.435771 | 6.415076 |
| ## 12 | 3.982950 | 1.428833 | 6.537068 |

Appendix F:

R Code:

```
unrate <- read_csv("data/UNRATENSA.csv")

# removing missing values
unrate <- unrate[complete.cases(unrate), ]

knitr::kable(head(unrate, n = 10), "simple",
              caption = "Dataset Observations")

# converting into date
unrate$Date <- as.Date(unrate$DATE)

unrate$Year <- year(unrate$Date)
unrate$Month <- month(unrate$Date)

unrate <- unrate %>% filter(Date <= as.Date("2023-12-31"))

# Creating time series
ts_unrate <- ts(unrate$UNRATENSA, start = c(min(unrate$Year),
                                           min(unrate$Month)), frequency = 12)

# Create a data frame for plotting
plot_unrate <- data.frame(Date = time(ts_unrate), Value = ts_unrate)

# define time series from data frame
us_unrate <- as.ts(plot_unrate$Value)

# results:
# plotting original time series
plot(us_unrate, col = "skyblue4",
     ylab = "Rate", xlab = "Time")

# transform data and difference at lag 1
par(mfrow = c(2,2))
transform_unrate <- diff(log(us_unrate))
# difference at 12
diff12=diff(transform_unrate, 12)
acf2(diff12, main = "ACF and PACF of Transformed,
                  Differenced Data at Lag 1 and 12")

# sarima for best model
sarima(diff12, 1, 0, 1, 1, 0, 1, 12) # AIC = -2.583814

# predicting next 12 points
sarima.for(us_unrate, 12, 1, 1, 1, 1, 1, 1, 12)

nextn(912)
par(mfrow = c(2,2))
```

```

# raw spectrum
unrate.per = mvspec(us_unrate, taper = 0.1, log = "no", main =
                    "Raw Spectrum")
abline(v = c(0.0625, 1, 2, 3, 5), lty = 3)

k = kernel("daniell", 4)
# raw smoothed
unrate.smo = mvspec(us_unrate, taper = 0.1, kernel = k, log = "no",
                    main = "Smoothed Spectrum")
abline(v = c(0.0625, 1, 2, 3, 5), lty = 3)

par(mfrow = c(2,2))
# log spectrum
unrate.log.per = mvspec(us_unrate, taper = 0.1, log = 'y',
                        main = "Logged Raw Spectrum")
abline(lty = 2)

# log smoothed
unrate.log.smo = mvspec(us_unrate, taper = 0.1, kernel = k, log = 'y',
                        main = "Logged Smooth Spectrum")
abline(lty = 2)

#Identify the first three dominant frequencies for unrate series
freq <- unrate.per$details[order(unrate.per$details[,3],decreasing = TRUE),]
# freq[1,];freq[2,];freq[3,]

# extracting significant peaks
one <- unrate.per$details[80,] # 80/ 960 = 0.08333 = 1/12
two <- unrate.per$details[160,]
three <- unrate.per$details[240,]
five <- unrate.per$details[400,]

##95% CIs for the dominant frequencies
unrate.u1 = 2*freq[1,3]/qchisq(.025,2)
unrate.l1 = 2*freq[1,3]/qchisq(.975,2)
unrate.u2 = 2*freq[2,3]/qchisq(.025,2)
unrate.l2 = 2*freq[2,3]/qchisq(.975,2)
unrate.u3 = 2*freq[3,3]/qchisq(.025,2)
unrate.l3 = 2*freq[3,3]/qchisq(.975,2)

# calculating confidence intervals
# unrate.per$spec[80]
# U = qchisq(.025,2)
# L = qchisq(.975,2)
# 2*unrate.per$spec[80]/L # 0.4576484
# 2*unrate.per$spec[80]/U # 66.68073

# creating table
dominant_freq <- data.frame(
  Series = c("Top Frequency A", "Top Frequency B", "Top Frequency C",

```



```

      "Frequency 1", "Frequency 2", "Frequency 3", "Frequency 5"),
Frequency = c(freq[1, 1], freq[2, 1], freq[3, 1],
              one[1], two[1], three[1], five[1]),
Period = c(freq[1,2], freq[2,2], freq[3,2],
            one[2], two[2], three[2], five[2]),
Spectrum = c(freq[1,3], freq[2, 3], freq[3, 3],
              one[3], two[3], three[3], five[3]),
Lower = c(round(unrate.l1, 4), round(unrate.l2, 4),
           round(unrate.l3, 4), " ", " ", " ", " "),
Upper = c(round(unrate.u1, 4), round(unrate.u2, 4),
           round(unrate.u3, 4), " ", " ", " ", " "))

knitr::kable(dominant_freq, "simple", caption =
             "Dominant Frequencies in the Time Series")

# histograms of original and logged time series
par(mfrow = c(1,2))
hist(us_unrate)
hist(log(us_unrate))

# plot of transformed, differenced at lag 1
plot(diff(log(us_unrate)), main="Transformed Unemployment Rate in U.S.
      from 1948 - 2023",
      ylab = "Rate")

# acf and pacf of first differenced, transformed plots
acf2(transform_unrate)

# estimation of other models
sarima(diff12, 0, 0, 1, 1, 0, 2, 12) # AIC = -2.583618
sarima(diff12, 1, 0, 0, 2, 0, 3, 12) # AIC = -2.579977
sarima(diff12, 5, 0, 1, 1, 0, 3, 12) # AIC = -2.578607
sarima(diff12, 1, 0, 1, 1, 0, 2, 12) #AIC = -2.581589

# future predictions
pred1 <- sarima.for(us_unrate, 12, 1, 0, 1, 1, 0, 1, 12, plot = FALSE)
year = c(1:10)
# the 5% upper and lower prediction interval
upper = pred1$pred+qnorm(0.975)*pred1$se
lower = pred1$pred-qnorm(0.975)*pred1$se
(data.frame("Prediction"=pred1$pred,"PI 95% Lower Bound"=lower,
            "PI 95% Upper Bound"=upper))

```