STA457 Final Project: Analysis of the U.S. Unemployment Rate Fluctuations

Yiwen Li 1005749219

April 9, 2022

Abstract

The unemployment rate is an important indicator of the economic status of the country and can also influence the policymaking of the government. The main purpose of this report is to study the fluctuation pattern of the monthly unemployment rate in the United States from January 1948 to November 2016 and to make forecasts for the future U.S. unemployment rates. After stabilizing the time series, we first propose two models, SARIMA(3,1,0)x$(0,1,1)_{12}$ and SARIMA(2,1,1)x$(0,1,1)_{12}$. By model selection and model diagnostics, we chose SARIMA(2,1,1)x$(0,1,1)_{12}$ as the final model to predict the U.S. unemployment rate for a 10-month period starting from December 2016. According to the model forecast, while the US unemployment rate remains slightly volatile, the overall U.S. unemployment level will show a slow downward trend since December 2016.

Keywords: *Time series, Seasonal autoregressive integrated moving average model, Unemployment rate, Forecasting*

Introduction

The unemployment rate is an important indicator of the economic status of a country. The unemployment rate measures the number of unemployed people as a percentage of the labor force, while the labor force is defined as the total number of unemployed and employed people[1]. When the economy is sluggish, especially after a negative economic shock, a decrease in market demand may lead to an increase in the unemployment rate. For the government, an increase in the unemployment rate means a decrease in tax revenue and also an increase in the cost of unemployment benefits, thus may worsen the government's financial conditions. Therefore, one of the important macro objectives for the government is to stabilize the unemployment rate at a relatively low level[2]. In this report, we are interested in the fluctuation pattern of the U.S. unemployment rate. The time-series data we choose for the analysis is the 'UnempRate' dataset in the 'astsa' package in R[3]. This dataset records the monthly U.S. unemployment rate in percentage from January 1948 to November 2016. This time series includes 827 observations and is published by the U.S. Bureau of Labor Statistics[4]. The main purpose of this report is to analyze the past trends in the U.S. unemployment rate and build a model to forecast future fluctuations, helping the U.S. government to assess future economic conditions and provide some effective information for policy-making.

Statistical Methods

Data Stationarity

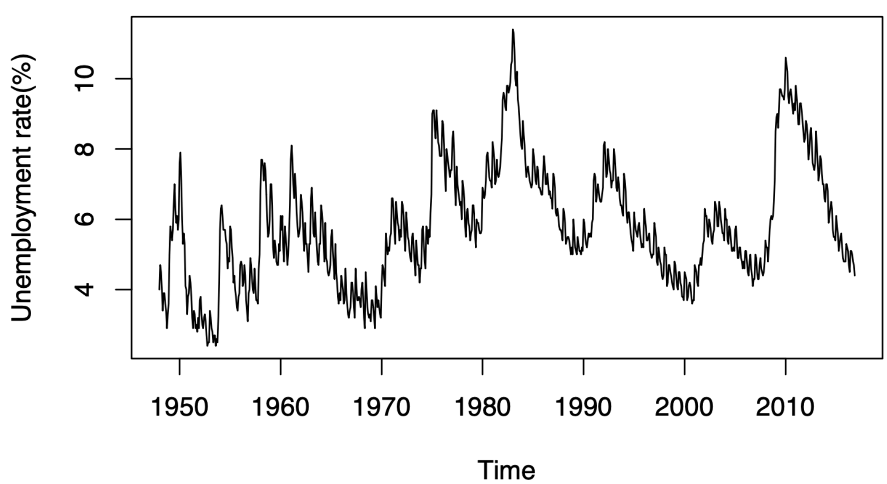

Fig.1: *U.S. Unemployment Monthly Rate from 1948 to 2016*

Fig.1 shows the initial fluctuation patterns of the U.S. unemployment rates from 1948 to 2016. At first glance, the process is not stationary since the mean of the series is not constant over time.

According to Fig.1, the U.S. unemployment rates do fluctuate over time and overall there is a slight upward trend. Notably, during the deep recessions of the early 1980s and of 2007–2009, the U.S. unemployment rate rose sharply, reaching roughly 10%[5]. Also, we observed that there are yearly cyclical rises and falls in the unemployment rate, indicating a potential existence of seasonality.

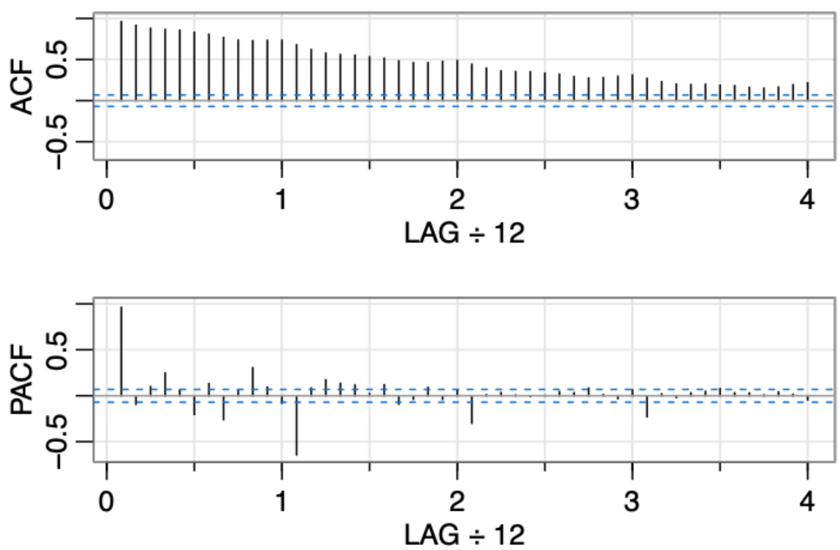

Fig.2: *Sample ACF and PACF of the U.S. unemployment rate series*

Based on Fig.2, it is clear that there is a slow decay in the sample ACF. Therefore, we concluded that the original process is not stationary and differencing is needed. To stabilize the mean of the series, we first difference the data once. However, we observed that except for the slow decay trend, the ACF shows a rough peak at lag 1s(s=12), indicating that the dependence on the past tends to occur strongly at lag 1s(s=12). Thus, to account for both the non-stationary and the seasonal fluctuations, a twelfth-order difference will also be applied after taking the first difference.

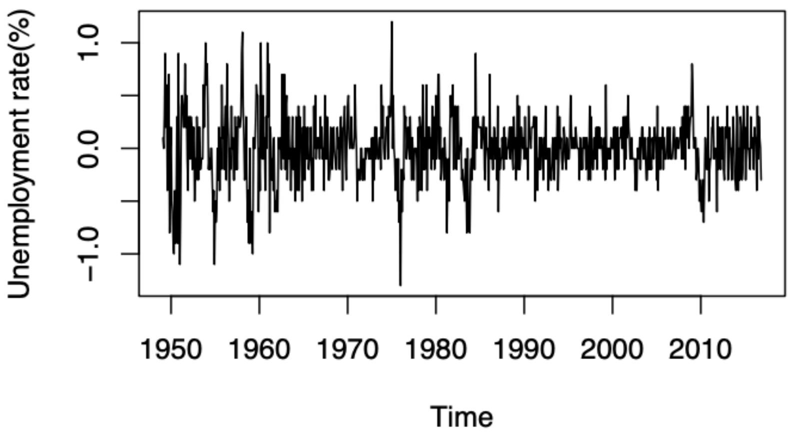

Fig.3: *U.S. unemployment Rate after differencing*

From Fig.3, it is clear that after detrending the data, the mean and variance of the series stay constant over time, meaning that the series is now stationary and is ready for model fitting. To make forecasts of future unemployment rates, an SARIMA(p, d, q)x(P, D, Q)$_s$ model will be fitted on the series. In our case, the model will exhibit the series at the multiple of the yearly seasonal period s=12 months. Since we have applied both differencing and seasonal difference once, d=D=1 in all proposed models. To determine both the seasonal(P, Q) and non-seasonal model components(p, q), the sample ACF and PACF of the detrended series are plotted.

Models Building

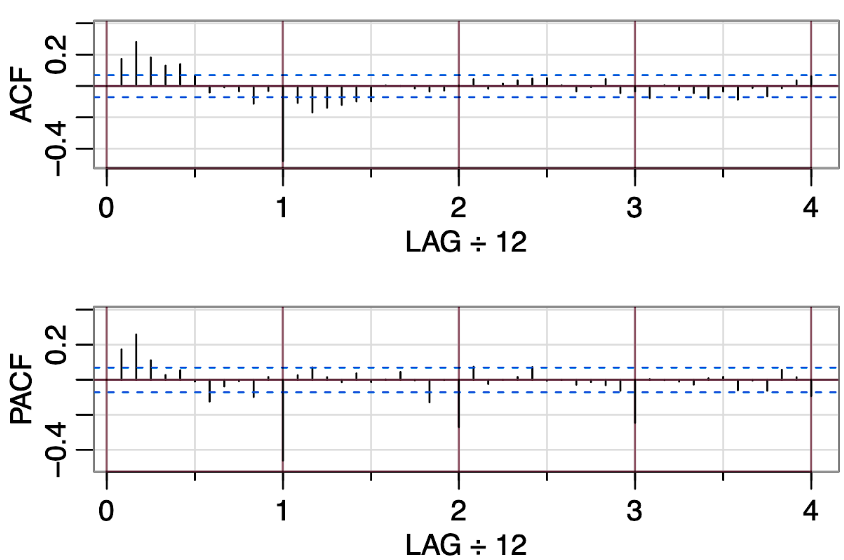

Fig.4: *Sample ACF and PACF of the U.S. unemployment rate after differencing*

The sample ACF and PACF of the monthly U.S unemployment rate after differencing are plotted in Fig.4. For the seasonal component, the ACF is cutting off at lag 1s(s=12), whereas the PACF is tailing off. Thus, we suggest an SAR(1) model, P=0,Q=1, in the season(s=12). Inspecting the lower lags to determine the non-seasonal component, the PACF is cutting off at lag 3 and the ACF is tailing off, suggesting that p=3 and q=0 in the SARIMA model. Thus, we proposed that the U.S. unemployment rates follow an ARMA(3, 0) process within the seasons. However, rather than focus on one model, we also suggest both ACF and PACF are tailing off, where ACF shows a relatively slower decay than the PACF. This suggests an ARMA(2,1) model within the seasons, p=2 and q=1. Therefore, the two candidate models will be SARIMA(3,1,0)x$(0,1,1)_{12}$ and SARIMA(2,1,1)x$(0,1,1)_{12}$. Next, we will fit the two models on the unemployment series to test the significance of the model parameters and perform model diagnostic tests to investigate their performances. We will also compare the BIC value of the candidate models.

Results

Significance tests and model diagnostics

Table.1: *Table of* $SARIMA(3,1,0) \times (0,1,1)_{12}$ *model summary*

parameter	Estimate	SE	t-value	p-value
ar1	0.1148	0.0351	3.2678	0.0011
ar2	0.2023	0.0345	5.8651	0.0000
ar3	0.0900	0.0350	2.5709	0.0103
sma1	-0.7674	0.0256	-30.0169	0.0000

Table.2: *Table of* $SARIMA(2,1,1) \times (0,1,1)_{12}$ *model summary*

parameter	Estimate	SE	t-value	p-value
ar1	0.5897	0.1105	5.3353	0.000
ar2	0.1342	0.0465	2.8826	0.004
ma1	-0.4831	0.1090	-4.4324	0.000
sma1	-0.7676	0.0254	-30.2459	0.000

For both SARIMA(3,1,0)x$(0,1,1)_{12}$ and SARIMA(2,1,1)x$(0,1,1)_{12}$ models, p-values for all parameters are smaller than 0.05, meaning that all parameters in both candidate models are significant. Therefore, we will not drop any parameter from the two models. Next, we perform the model diagnostic tests on both models to investigate their performances on the series.

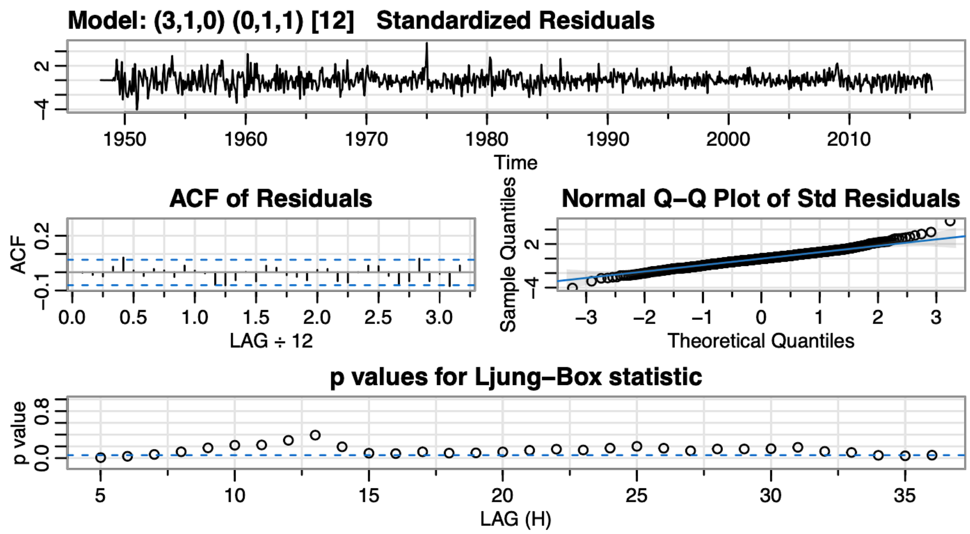

Fig.5: *Diagnostics of the residuals from* $SARIMA(3,1,0) \times (0,1,1)_{12}$ *fit on unemployment rate*

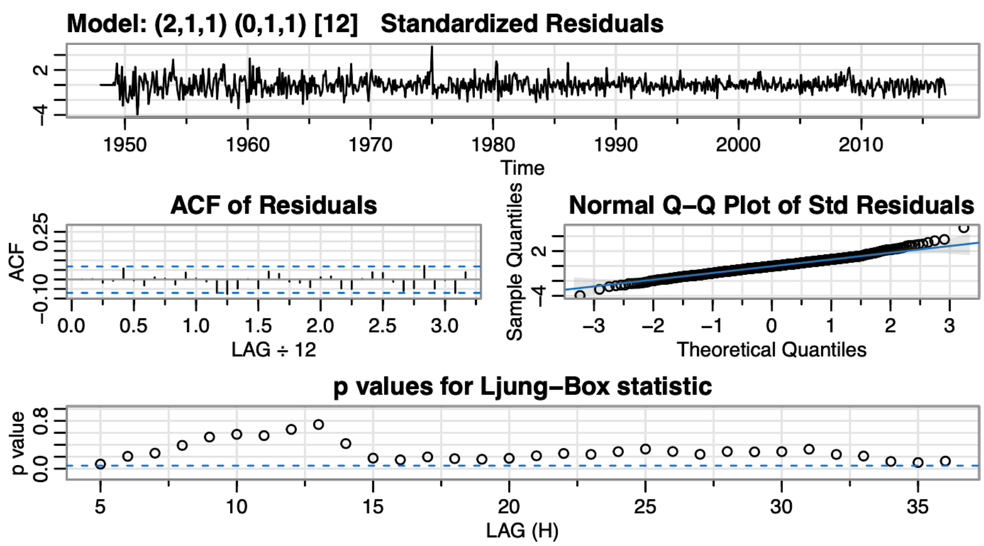

Fig.6: *Diagnostics of the residuals from* $SARIMA(2,1,1) \times (0,1,1)_{12}$ *fit on unemployment rate*

Fig.5 shows the diagnostic results of SARIMA(3,1,0)x$(0,1,1)_{12}$. According to Fig.5, the standardized residual plots show no specific pattern and fluctuate randomly around 0. Also, there are no significant spikes in the ACF residual plot, so we conclude that the randomness assumption is satisfied. From the normal QQ plot of residuals, the normal assumption for the model is satisfied as except for a few outliers in the tails, most of the points lie on the straight line. However, some of the p-values for Ljung-Box statistics are below the significance level for several lags shown in the plots, rejecting the null hypothesis that the residuals are independent.

Overall, the standardized residual seems to follow a normal distribution with zero mean and constant variance, but the Ljung-Box statistics indicate that the residuals are not independent and identically distributed, suggesting that the SARIMA(3,1,0)x(0,1,1)$_{12}$ model does not fit well.

Fig.6 shows the diagnostic result of SARIMA(2,1,1)x(0,1,1)$_{12}$. Similar to the SARIMA(3,1,0)x(0,1,1)$_{12}$ model, the standardized residual seems to follow a normal distribution with zero mean and constant variance. However, compared to the SARIMA(3,1,0)x(0,1,1)$_{12}$ model, the p-values for Ljung-Box statistics of SARIMA(2,1,1)x(0,1,1)$_{12}$ are higher than the significance level for most lags, indicating that we do not reject the null hypothesis that the residuals are independently distributed. Thus, we suggest that SARIMA(2,1,1)x(0,1,1)$_{12}$ seems to fit well on the unemployment data.

Model selection

Table.3: *Table comparing the BIC value of candidate models*

Model	BIC
$SARIMA(3,1,0) \times (0,1,1)_{12}$	0.006670833
$SARIMA(2,1,1) \times (0,1,1)_{12}$	0.002620143

Table.4: *Table of $SARIMA(2,1,1) \times (0,1,1)_{12}$ model summary*

parameter	Estimate	SE	t-value	p-value
ar1	0.5897	0.1105	5.3353	0.000
ar2	0.1342	0.0465	2.8826	0.004
ma1	-0.4831	0.1090	-4.4324	0.000
sma1	-0.7676	0.0254	-30.2459	0.000

Based on Table.3, the BIC score for SARIMA(2,1,1)x(0,1,1)$_{12}$ is lower. Combining with the results of the diagnostics in the previous section, all evidence support that SARIMA(2,1,1)x(0,1,1)$_{12}$ should be chosen as the final model. Therefore, based on the estimate parameters in Table.4, the final model can be expressed as:

$$(1 - 0.5897B - 0.1342B^2)y_t = (1 - 0.4831B)(1 - 0.7676B^{12})w_t$$

where $y_t = (1 - B)(1 - B^{12})x_t$. All p-values of the coefficients of the final model are significant. In order to provide a more intuitive interpretation in the real context, we substitute y_t in the model expression for x_t. The expanded model is $x_t = 1.5897x_{t-1} - 0.4555x_{t-2} - 0.1342x_{t-3} + x_{t-12} - 1.5897x_{t-13} + 0.4555x_{t-14} + 0.1342x_{t-15} + w_t - 0.4831w_{t-1} - 0.7676w_{t-12} + 0.3708w_{t-13}$. Therefore, we conclude that the U.S. unemployment rates from 1 month, 2 months, 3 months, 12 months, 13 months, 14 months, and 15 months ago will all have an impact on the current U.S. unemployment rates. Based on the fitted model, there are positive relationships between the current U.S. unemployment rate and the U.S. unemployment rates from 1 month, 12 months, 14 months, and 15 months ago. Also, the coefficient at lag 1, 12, and 13 are relatively larger, meaning that the unemployment rate at 1 month, 12 months, and 12 months ago may have stronger relationships with the current unemployment rate.

Forecasting

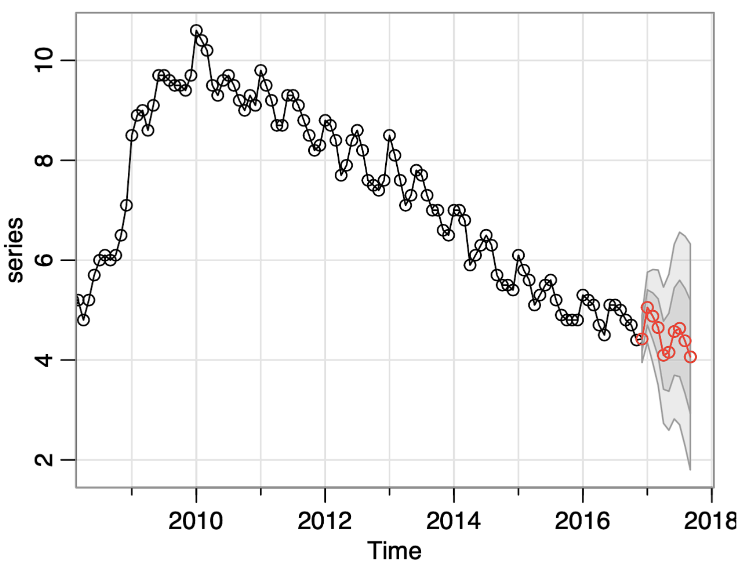

Fig.7: *Ten months forecast using the* $SARIMA(2,1,1) \times (0,1,1)_{12}$ *model on the unemployment series*

Table.5: *95% CI Interval for the Unemployment Rate Predictions*

Time	Forecast	Lower bound	Upper bound
Dec 2016	4.421514	3.959382	4.883645
Jan 2017	5.052758	4.363500	5.742017
Feb 2017	4.877218	3.961792	5.792644
Mar 2017	4.649057	3.518913	5.779201
Apr 2017	4.095737	2.760799	5.430676
May 2017	4.154975	2.625351	5.684600
Jun 2017	4.570231	2.855666	6.284795
Jul 2017	4.632799	2.742535	6.523064
Aug 2017	4.383510	2.326175	6.440845
Sep 2017	4.062609	1.846188	6.279031

We used the final model to predict the U.S. unemployment rates for 10 months starting from Dec 2016. Based on Fig.7, the U.S. unemployment rate will remain modestly volatile going forward, with the overall level remaining between 4% and 6%. We can find that since the Great Recession shock during 2007-2009 that pushed the unemployment rate to a peak of 10%, the unemployment rate has continued to decline from 2010 onwards[5]. And according to the model forecasts, the U.S. unemployment rate will continue to show an overall slow decline over the next 10 months starting from December 2016. According to Table.5, all of our predicted unemployment rate over the 10-months period fall within the generated confidence interval, and the confidence interval for the U.S. unemployment rate become wider as time progresses. In Dec 2016, we are 95% confident that the true U.S. unemployment rate will be between 3.959% and 4.884%, while in Sep 2017, we are 95% confident that the true U.S. unemployment rate will be between 1.846% and 6.279%.

Spectral Analysis

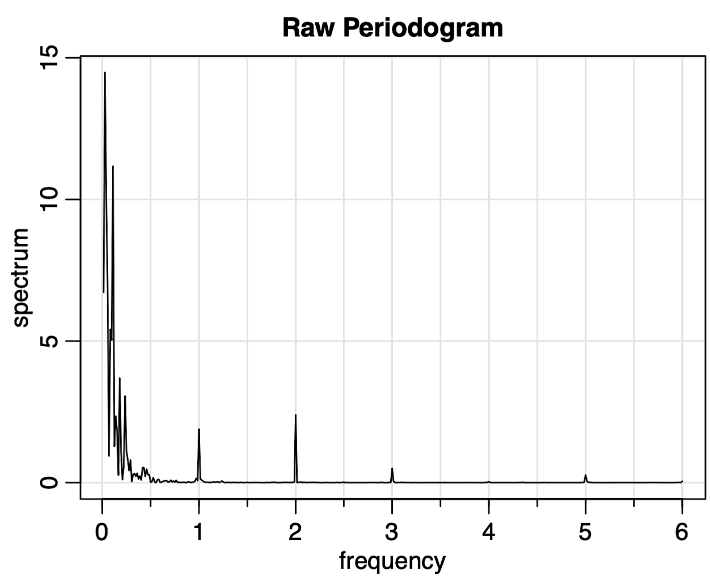

Fig.8: *Periodogram of the U.S. Unemployment series, where the frequency axis is labeled in multiples of 1/12*

Table.5: *95% CI Interval for the Unemployment Rate Predictions*

Dominant Frequency	Dominant Period	Spectrum	Lower	Upper
0.0278	36.0000	14.4829	3.926097	572.0440
0.1111	9.0000	11.1720	3.028562	441.2704
0.0417	24.0000	9.9625	2.700685	393.4977

As shown in Table.6, the three predominant periods for the unemployment series are 36 years, 9 years, and 24 years sequentially. In other words, there will be a dominant period around every 36 years, 9 years, and 24 years on average. The 95% confidence interval for each predominant period indicates the interval that we are 95% confident the spectrum will fall within. However, the confidence intervals for all three periods are so wide that each interval captures all three periodogram ordinates shown in the table. Therefore, we cannot establish the significance of the three dominant peaks of the unemployment series.

Discussion

This report aims at analyzing the fluctuations of U.S. unemployment rates and predicting the future unemployment rates based on the past data. **According to the model predictions, the U.S. unemployment rate will remain modestly volatile and will continue to show an overall slow decline over the next 10 months starting from December 2016. This could be a sign that the general status of the U.S. economy will slowly improve.** Therefore, the U.S. government could consider gradually tightening unemployment benefits policies and allocating funds to other social components, such as education and urban development. However, there are some limitations to our analysis. Firstly, the SARIMA model we have chosen may not be the best model for predictions. In this analysis, we relied on the sample ACF and PACF plots to build the model. Therefore, the choice of model is subjective and there may exist other models that can fit better on the series. Also, the real situation may deviate from our forecast given the existence of future uncertainties. Unexpected economic shocks or government policies could cause unusual fluctuations in the unemployment rate. Therefore, future studies can combine the unemployment rate and national GDP data to further analyze how the unemployment rate will change after the economy receives a sharp shock or the government adjusts its economic policies.

Bibliography

[1] Team, T. I. (2022, March 28). What is the unemployment rate? Investopedia. Retrieved April 16, 2022, from https://www.investopedia.com/terms/u/unemploymentrate.asp

[2] Macroeconomics. Corporate Finance Institute. (2021, February 22). Retrieved April 16, 2022, from https://corporatefinanceinstitute.com/resources/knowledge/economics/macroeconomics/

[3] David Stoffer (2021). astsa: Applied Statistical Time Series Analysis. R package version 1.14. https://CRAN.R-project.org/package=astsa

[4] U.S. Bureau of Labor Statistics. (n.d.). Bureau of Labor Statistics Data. U.S. Bureau of Labor Statistics. Retrieved April 16, 2022, from https://data.bls.gov/timeseries/LNU04000000/

[5] OpenStax. (n.d.). Patterns of unemployment. Principles of Economics. Retrieved April 16, 2022, from https://opentextbc.ca/principlesofeconomics/chapter/21-2-patterns-of-unemployment/

[6] Report prepared by R: R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/./