Analysis of the U.S. Unemployment Rate

STA457 Final Project

Ying Xiong (1004795885)

2022-04-17

Contents

1 Abstract **2**

2 Introduction **2**

3 Statistical Methods **3**

4 Result **7**

5 Discussion **11**

6 Reference **12**

1 Abstract

Unemployment rates have long since been an accurate indicator of economic output and the general welfare in most economic research. Persistence of high unemployment rates reveals that the country's economy is in turmoil and inefficient, suggesting instant government interventions for stabilization. The recent pandemic has resulted in soaring unemployment rates in prior periods, where predictions on the future rates become essential during this time. Through exploring the monthly U.S. unemployment dataset from the U.S. Bureau of Labor Statistics, I adopted a seasonal ARIMA model to examine the trend and make forecasts for the short-term future. The results suggest strong seasonality in the month and an overall downward trend in the U.S. unemployment rates, which indicates that the U.S. economy is steadily recovering from the pandemic though the rate is still higher compared to before the pandemic. Moreover, the mid and end of the year tend to suffer from high unemployment, which may be caused by the influx of graduates. The spectral analysis shows no significant spectral peak to identify the characteristic cycle in the data under the 95% confidence interval. Overall, the government may implement campaigns to speed up recovery from the pandemic by incentivizing companies to hire new grads and increase production. One limitation of this paper comes from the apparent outliers in the model diagnostic, advising model modification, or data processing in further analysis.

Keywords: U.S. Unemployment Rate, Time Series Forecasting, Real time, Spectral Analysis

2 Introduction

Previous economic studies have demonstrated that the unemployment rate can effectively evaluate a country's economy. According to the widely accepted Okun's law in macroeconomics, there is a negative relationship between a country's output and its unemployment rate, affecting the growth rate of real GDP (Wen, 2012). In particular, a persistently high unemployment rate will instill pessimism into the market and potentially translate into economic downturns and stagnations if not handled appropriately (Levine, 2013). Hence, the analysis of unemployment trends and predictions of future movements are of particular interest, especially during periods of turbulence, such as the global pandemic. This paper aims to build a time-series model to forecast future American unemployment rates ten-periods after the March 2022 cutoff date, the latest data available. For model construction, I extracted the monthly U.S. unemployment rate data from the U.S. Bureau of Labor Statistics. Due to significant changes in American monetary and fiscal policies as well as the general economic situation since the turn of the century, all of which impact the long-term unemployment rate, I have opted to select only data after the year 2000, yielding 267 observations in total. Similar to the steady decline in the unemployment rate forecasts in the following years from FOMC (Federal Reserve, 2022), my analysis appears that the unemployment rate will continue to trend downwards for the next ten months while exhibiting seasonal fluctuations as usual, which is a signal of economic resurgence after pandemic.

3 Statistical Methods

This section provides a detailed discussion and justification of the data analysis methods employed for model construction and the model selection process.

First, I generate a time series plot of the monthly U.S. unemployment rate using the raw data with its ACF and PACF plots to inspect for anomalies. In *Figure 1*, the original data is non-stationary with downward trends and seasonal fluctuations. Simultaneously, the sample ACF plot displays a slow decay to zero, indicating the need for differencing. *Figure 2* plots the differenced data, which appears stationary around zero with some seasonality. The seasonal subseries show persistence in the seasons through fluctuations occurring every 12 months. Hence, I employ another twelfth-order difference and plot the transformed data in *Figure 3*, $\nabla\nabla_{52}$x, where the data now possesses the properties of a stationary process with constant mean and variances around zero. I will fit a SARIMA model utilizing the processed data.

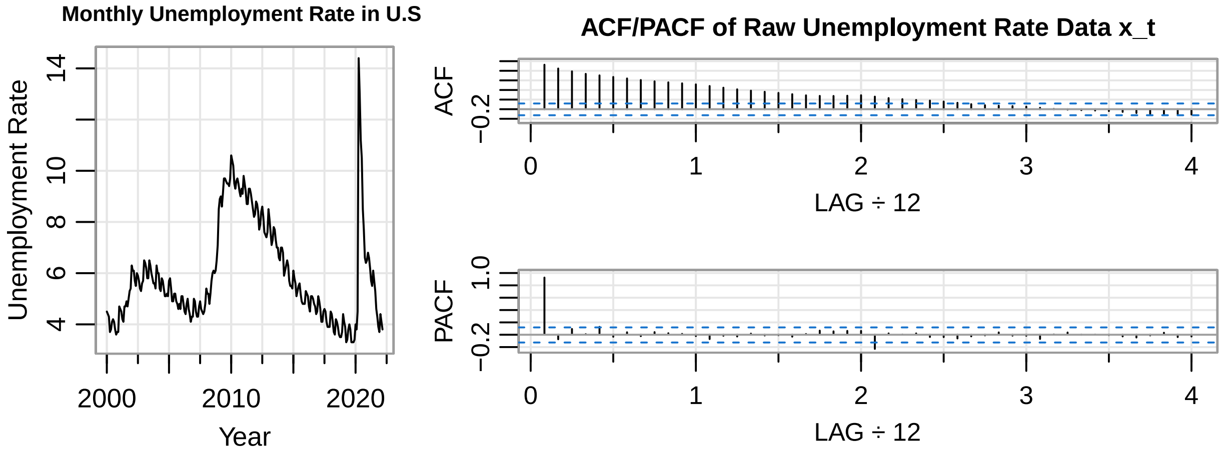

Figure 1: Analysis of the Raw Data for U.S. Unemployment Rate (2000-2022)

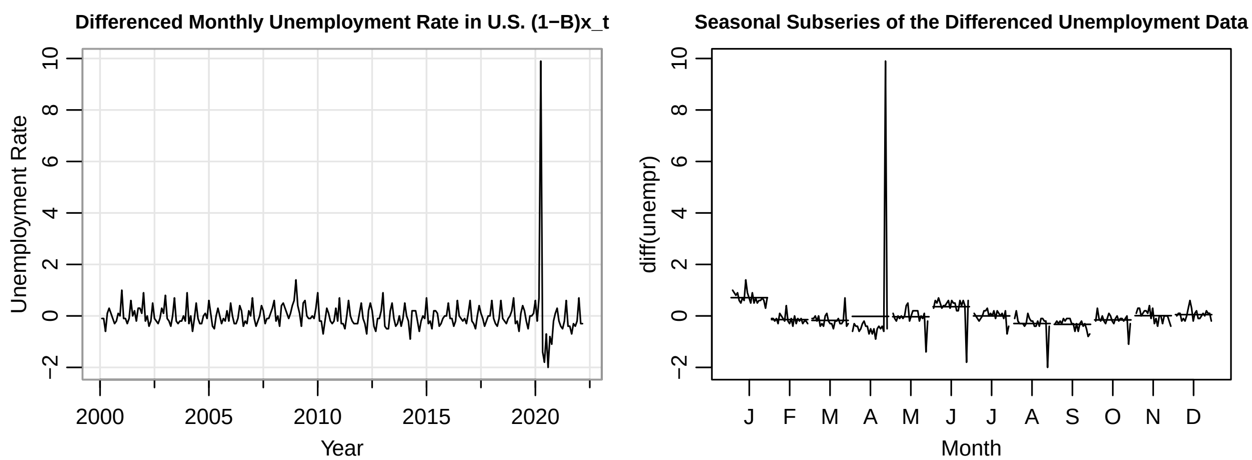

Figure 2: Analysis of Data with First Difference (d = 1)

Figure 3: Plot of Differnced Data (d = 1, D = 1)

To identify the seasonal and non-seasonal components of the model, I generate the ACF/PACF plot of the transformed data and display it in *Figure 4*.

As shown, at the seasons, ACF cuts off at lag = 1s where s = 12, while PACF appears to tail off at 1s, 2s, ... These suggest a seasonal Moving Average (1) model, where P = 0, Q = 1 with D = 1 in the seasons (s = 12). For the non-seasonal component, the dependence orders are less clear. One may argue that ACF cuts off after lag 2 while PACF is tailing off within the season and propose a model of ARIMA(0,1,2)x$(0,1,1)_{12}$. Another may propose the opposite where ACF is tailing off within the season and PACF cuts off after lag 2 and suggest a model of ARIMA(2,1,0)x$(0,1,1)_{12}$.

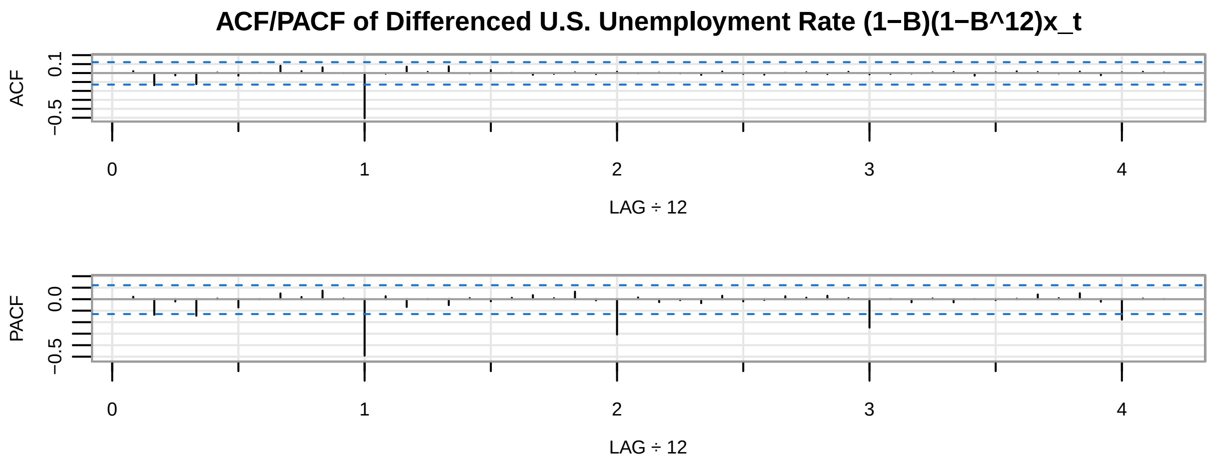

Figure 4: ACF/PACF of Differnced Data (d = 1, D = 1)

Figure 5 and *Figure 6* store the residual diagnostics of the two models. Browsing through the plots, we can claim that both models fit the data quite well. For the plots of the standardized residuals, there are no obvious patterns except for the spike around 2020, which is reasonable due to the special event of pandemic. The ACFs of the standardized residuals exhibit no apparent departure from the model assumptions. Both normal Q-Q plots of the residuals show that the assumption of normality is reasonable through straight lines, with some exceptions of outliers at the tails. Last, the p-values for Ljung-Box statistics almost all lie above the significance level with few violations in the smaller lags, which means we do not reject the null hypothesis of independent residuals. Therefore, the diagnostic plots for the two models are comparable with similar results, where the residuals seem independent and normal with mean zero and constant variance. I will further compare the models through their AIC, BIC, and AICc, where smaller values in these measures are the main criteria.

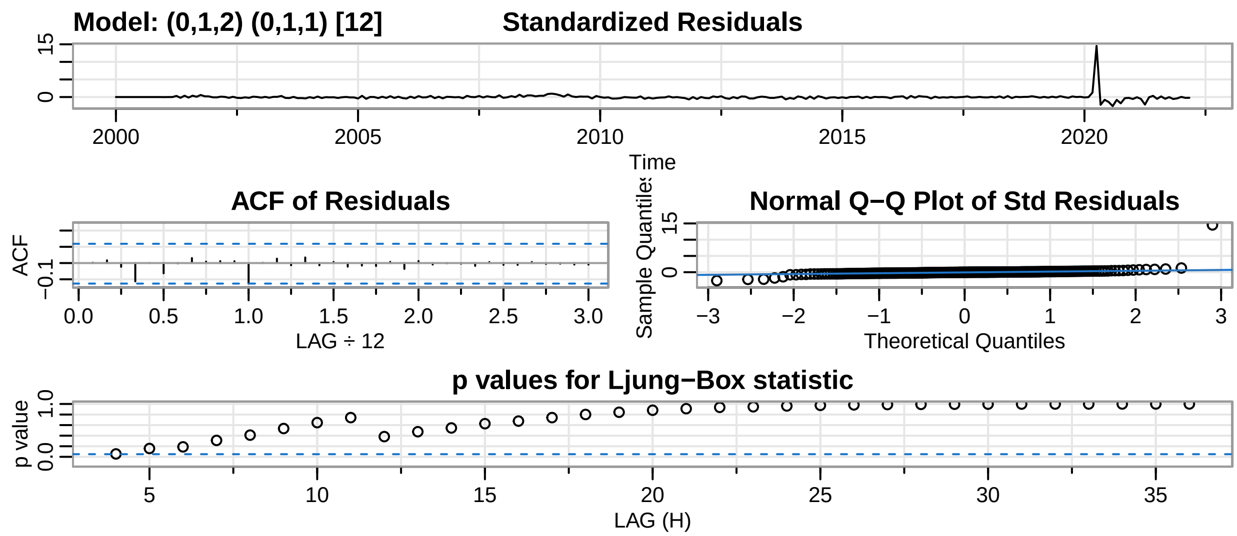

Figure 5: Residual Diagnostic of Model ARIMA(0,1,2)x(0,1,1)_12

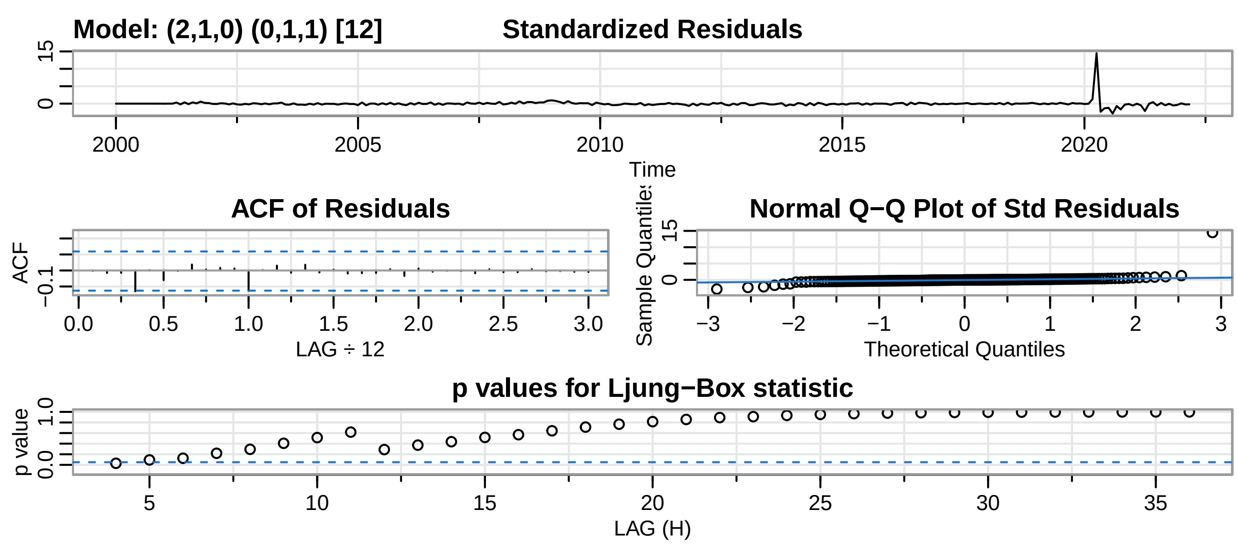

Figure 6: Residual Diagnostic of Model ARIMA(2,1,0)x(0,1,1)_12

As exhibited in *Table 1*, the ARIMA(0,1,2)x$(0,1,1)_{12}$ model has comparatively smaller results in all three measures. With similar performance in the diagnostics, I choose to fit an ARIMA(0,1,2)x$(0,1,1)_{12}$ model as the most appropriate model and apply it for further analysis. In specific, s = 12 represents a 12-month seasonal trend; p = 0 and P = 0 indicate no autoregressive term; d = 1 and D = 1 stand for one difference with one seasonal difference imposed on the series; q = 2 and Q = 1 stands for moving average terms.

Table 1: AIC, BIC, AICc Comparisons for Two Proposed Models

Model	AIC	BIC	AICc
ARIMA(0,1,2)x(0,1,1)_12	2.25686	2.31256	2.25723
ARIMA(2,1,0)x(0,1,1)_12	2.26127	2.31698	2.26165

The formula expression is listed below.

$$(1-B)(1-B^{12})x_t = w_t(1+\theta_1 B+\theta_2 B^2)(1+\Theta_1 B^{12}),$$

where x_t is the time series, standing for U.S. unemployment rate at time t; w_t is white noise that follows N(0, σ_w^2), standing for randomness; θ_1, θ_2, and Θ_1 are the coefficients for the moving average terms.

4 Result

This section presents the findings from the constructed model through analysis of parameter estimates, application in prediction, and spectral analysis.

Table 2: Parameter Estimates of $ARIMA(0,1,2)*(0,1,1)_{12}$

	Estimate	SE	p.value
ma1	0.0267	0.0633	0.6732
ma2	-0.1583	0.0716	0.0278
sma1	-0.8507	0.0505	0.0000

Table 2 presents the parameter estimates of the final model proposed above, with their

standard error and p values. As illustrated, θ_1 (ma1) has a p-value of 0.67, which is explicitly much larger than the threshold of 0.05, indicating statistically insignificant. Both p values of θ_2 and Θ_1 stay below the cutoff, with 0.03 and 0 correspondingly. Hence, the model parameters are statistically significant except for θ_1. With the estimates of θ_2 = -0.1583 and Θ_1 = -0.8507, the mathematical expression of the model is listed below:

$$(1-B)(1-B^{12})\hat{x}_t = w_t(1-0.1583B^2)(1-0.8507B^{12})$$

To interpret, I write the fitted mode in difference equation form:

$$\hat{x}_t = x_{t-1} + x_{t-12} - x_{t-13} + w_t - 0.1583w_{t-2} - 0.8507w_{t-12} + 0.1347w_{t-14}$$

As illustrated, the current unemployment rate is closely associated with the level of unemployment in the previous month as well as one year ago with minor negative influences from the randomness two months before and one year ago. In other words, the seasonality trend is critical in the measure of U.S. unemployment rates, and the current value is negatively related to the unemployment rate one year ago, indicating an overall downward trend when compared year-on-year.

With the fitted model discussed above, I predict the U.S. unemployment rate ten-month ahead. The forecasts with corresponding 95% prediction intervals are depicted in *Table 3*. To better demonstrate the model predictions, I plot the time series with red dots being the model forecasts, shown in *Figure 7*. As illustrated, the model predicts fluctuated rates for the following ten months, with a minor rebound in the first three months (2022.6) followed by gradually declining rates and elevations in the last two months (2022.12 – 2023.1). We can find the forecasts mimic the seasonal variations as before, where we may expect further stable decline if continue the forecasts. To extend, the U.S. unemployment rates tend to reduce in the first season and then rebound a little in mid-year and keep sliding until the end of the year. In general, the unemployment rates have an overall downward trend when compared year-on-year. Meanwhile, the 95% prediction intervals widen as time pass, while maintaining a stable and narrow range, suggesting relatively precise forecasts.

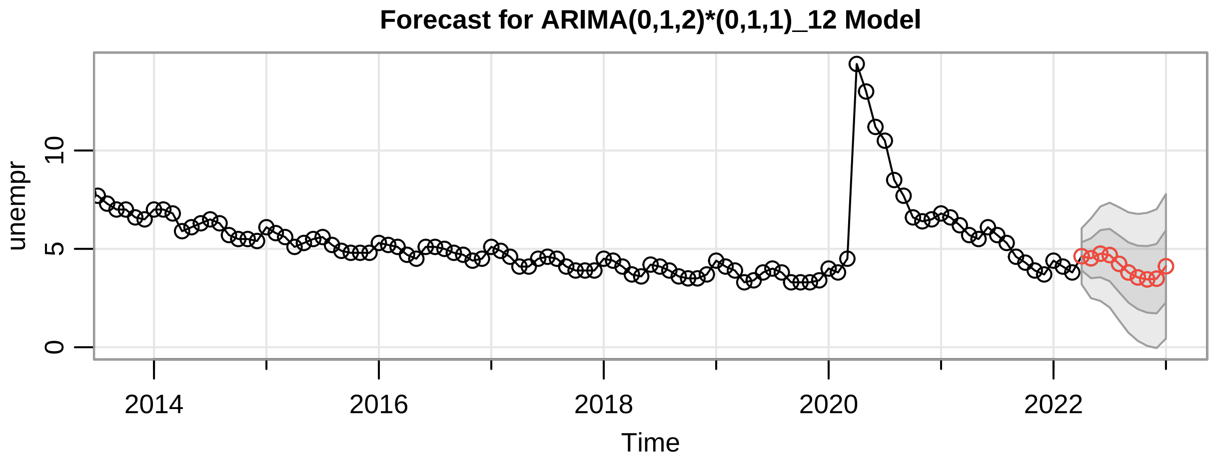

Figure 7: Prediction of U.S. Unemployment Rate for the upcoming 10 Months

Table 3: Ten-Month Prediction of U.S. Unemployment Rate using f $ARIMA(0,1,2)*(0,1,1)_{12}$

Time	Forecast	Upperbound	Lowerbound
1	4.623409	6.018820	3.22799799
2	4.521535	6.508629	2.53444127
3	4.754849	7.105570	2.40412742
4	4.684719	7.294802	2.07463704
5	4.239162	7.057288	1.42103686
6	3.798067	6.796471	0.79966298
7	3.548051	6.710038	0.38606458
8	3.448744	6.763148	0.13434081
9	3.484953	6.943562	0.02634319
10	4.114506	7.710817	0.51819556

Furthermore, I implement a spectral analysis on the original data to identify the predominant periods through a periodogram as *Figure 8* and store the first three predominant periods with 95% confidence intervals in *Table 4*. As shown, the peaks exhibit at spectrums of 12.57, 10.43, and 4.11, and cycles occur at 22.5, 12.25, and 5.625 months, where the second is consistent with the seasonality in months. However, according to the 95% confidence intervals, we cannot claim any of the three frequencies as a statistically significant peak since they all lie in the confidence intervals of the other dominant frequencies. In other words, there are not much information can be derived from this analysis due to the extremely wide range of confidence intervals.

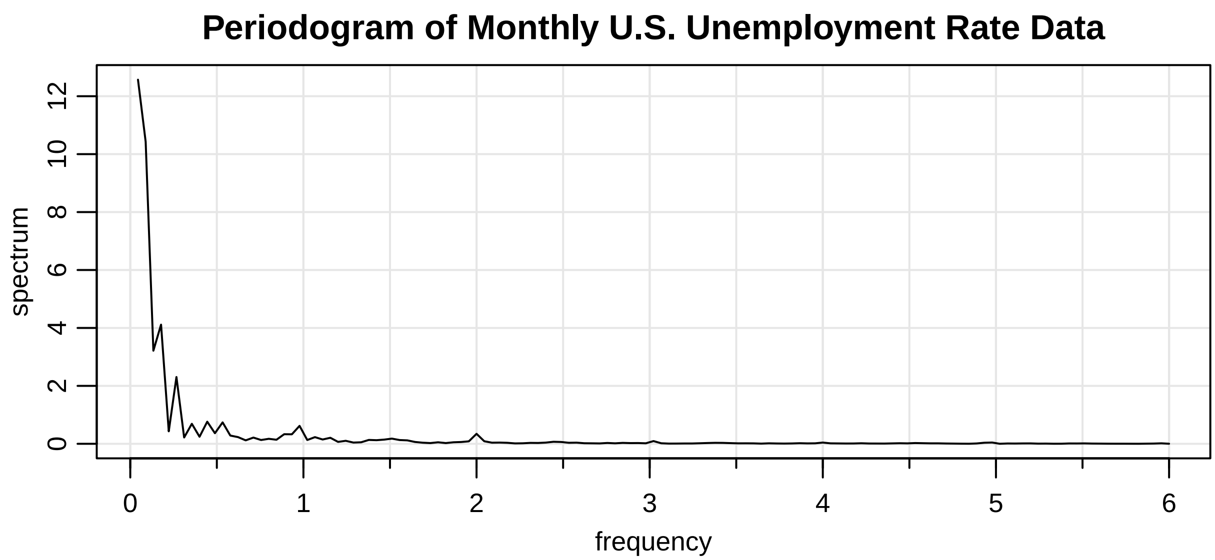

Figure 8: Periodogram of Monthly U.S. Unemployment Rate Data

Table 4: First Three dominent periods of Monthly U.S. Unemployment Rate Data

Series	Dominant_Freqency	Period	Spectrum	Lowerbound	Upperbound
1	0.0444	22.500	12.5716	3.40797	496.5517
2	0.0889	11.250	10.4271	2.82663	411.8485
3	0.1778	5.625	4.1144	1.11535	162.5101

5 Discussion

To sum up, the U.S. unemployment rate shows a downward trend with fluctuations in seasons and will continue to decline in the upcoming 10-months period, signifying economic recovery. It matches with the economic situations in the real world since we can find all the soaring unemployment rates take place when there's a specific global event such as the financial crisis in 2008-2009 and the recent COVID-19 pandemic in 2020. As the event's impacts ease over time, the market will stabilize and revive with elevating demands and recruitment, which will reduce unemployment steadily. Seasonality is another critical factor, where the unemployment rates attain peaks in the mid and end of the year. One potential explanation is that these are the time points where a lot of students graduate. Once they graduate from colleges, universities, or high schools, most will start to actively look for jobs, which will enlarge the population of unemployed compared to the labor force population. Since the predicted unemployment level is still higher than before the pandemic, the related government authorities may introduce campaigns that stimulate the production to accelerate the economic recovery, especially on recruitment of the new grads.

The primary limitation of this paper is derived from the imperfect residual diagnostics of the model. The plots imply the existence of outliers in the data, which may result from the outbreak of the COVID-19 pandemic. As I did not modify the data or the model to cope with the outliers, the final fitted model may be subject to reduced reliability. According to past academic research, simple time series cannot quickly adjust to changing conditions, which may not produce credible and precise estimations in uncertain times. (Castle, Clements & Hendry, 2016) For further research, one may collect more data on the recovery times after the pandemic and fit a model with no outliers to make future predictions. Another direction is to make modifications to the model such as non-linear models or inclusion of other potential factors so that it would incorporate the fact of pandemics and hence produce reliable results.

6 Reference

1. Castle, J.L., Clements, M.P. & Hendry, D.F. An Overview of Forecasting Facing Breaks. J Bus Cycle Res 12, 3–23 (2016). https://doi.org/10.1007/s41549-016-0005-2

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

3. Hao Zhu (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. https://CRAN.R-project.org/package=kableExtra

4. Levine , Linda. Economic Growth and the Unemployment Rate. Congressional Research Service, 7 Jan. 2013, https://sgp.fas.org/crs/misc/R42063.pdf.

5. "Summary of Economic Projections." Federalreserve.gov, BOARD OF GOVERNORS of the FEDERAL RESERVE SYSTEM, 16 Mar. 2022, https://www.federalreserve.gov /monetarypolicy/files/fomcprojtabl20220316.pdf.

6. Wen, Yi, and Mingyu Chen. "Okun's Law: A Meaningful Guide for Monetary Policy?" Research.stlouisfed.org, 8 June 2012, https://files.stlouisfed.org/files/htdocs/publicatio ns/es/12/ES_2012-06-08.pdf.

7. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

8. Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, Implementing Reproducible Computational Research. Chapman and Hall/CRC. ISBN 978-1466561595

9. Yihui Xie (2015) Dynamic Documents with R and knitr. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963

10. Yihui Xie (2022). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.38.