A LOOMING CRISIS:

THE U.S. UNEMPLOYMENT RATE IN FIGURES

A time series analysis of the U.S. unemployment rate and

the initial impact of the COVID-19 pandemic

RMIT University-Melbourne

Hardi Winata (S3794905), Cha Riofloridо (S3789114), Keavatey Srun (S3767615), Meng Heng Ea (S3716021)

Table of Contents

List of Figures

List of Tables

INTRODUCTION

In the last 70 years, the U.S. has been through multiple economic recessions. Two of these significant interventions were the dot-com bubble in 1995-2001 and the Great Recession/Global Financial Crisis in 2008-2010. The impact has been consistently seen on the increased unemployment rates. As it is a lagging economic indicator, the effect usually starts to increase only after a recession has begun and will continue upward even during the recovery period.

The time series analysis aims to model the unemployment rates from 2000 to 2019 and determine its future trajectory considering these two historical interventions. However, as the world faces its latest recession from the coronavirus pandemic, the analysis also assesses its initial impact on the world's largest economy.

It is imperative to note that since the World Health Organization only issued the first warning against COVID-19 in January 2020, and the unemployment rate is a lagging indicator, the forecasts will only be based until the December 2019 figures. The ten-month predicted values would then be compared to the actual and available published rates for 2020 to identify the accuracy of the models.

METHODOLOGY

Dataset sourced from Kaggle contains the U.S monthly unemployment rate from January 1948 to December 2019. For the time series analysis, only the subset dataset consisting of the unemployment rates from the last two decades were explored.

Before any modeling or transformations were applied, the time series plot was carefully examined to determine any influencing factors that can affect the model and forecast. As the economy of the U.S. faced several interventions with varying degrees, the time series analysis was approached based on two scenarios. The first approach aimed to analyze the entire time series from 2000 to 2019. The second approach was to split the series two parts—before and after the peak of a significant intervention i.e., Global Financial Crisis (GFC). The resulting forecasts were then compared against the published unemployment rates for 2020 to identify the accuracy of the models.

Both deterministic trends and stochastic trend models were fitted to identify the best model. Tests for normality and stationarity were applied accordingly.

Deterministic Trend Modelling covered the fitting of the linear and quadratic trend models. Criteria such as the principle of parsimony, higher adjusted R^2, lower residual standard error, and randomness were taken into consideration. Residual diagnostics were also presented to support the decision of the model with the best-fit.

Stochastic Trend Modelling accounted for the use of the model specification tools such as the autocorrelation plot (ACF), partial autocorrelation plot (PACF), extended autocorrelation plot (EACF), and Bayesian information criterion (BIC) table. The objective is to identify the value of the parameters p, d, and q of the candidate models. The parameters of the candidate models were estimated using the maximum likelihood (ML) instead of the conditional sum-of-squares (CSS) method. Residual analyses were also assessed to eliminate any models not conforming with criteria of the goodness of a model. The remaining candidate models were then subjected to the Akaike information criterion (AIC), Corrected Akaike information criterion (AICc), and BIC test. The best-fitted model was used to forecast for the next ten months.

The null hypothesis H_0 and alternative hypothesis H_1 of the significance tests used were as follow:

Shapiro-Wilk Test for Normality

H_0: $\rho >= 0.05$ Distribution of the series is normal

H_1: $\rho < 0.05$ Distribution of the series is not normal

Correlation Analysis

H_0: $r = 0$ There is no association

H_1: $r \neq 0$ A non-zero correlation could exist

Augmented Dickey-Fuller Unit-Root Test for Stationarity

H_0: $\rho >= 0.05$ Time series data is non-stationary

H_1: $\rho < 0.05$ Time series data is stationary

Constant Variances Test

H_0: $\rho >= 0.05$ Variance is constant

H_1: $\rho < 0.05$ Variance is non-constant

Runs Test for Independence

H_0: $\rho >= 0.05$ The residuals are independent

H_1: $\rho < 0.05$ The residuals are dependent

Box-Ljung Test for the Residual Analysis

H_0: $\rho >= 0.05$ Error terms are uncorrelated

H_1: $\rho < 0.05$ Error terms are correlated

All statistical tests were based on a 5% significance level. RStudio, with supplementary packages, was used to conduct the analyses and generate all visualizations in the report. Corresponding R Codes were appended in Appendix A.

RESULTS

Chapter 1: Monthly U.S. Unemployment Rate from 2000 to 2019

1.1. Descriptive Analysis

Figure 1 shows the time series plot of the U.S monthly unemployment rate from January 2000 to December 2019. According to the historical figures, the U.S unemployment started to rise at the beginning of 2001 until 2002. There were fluctuations during 2002-2003 before the unemployment rate began to decrease in 2004. This increase in unemployment rates can be attributed to the dot-com bubble, which crashed in 2000, leading to job losses during this period (Hayes, 2019). Not long after that, the unemployment rates started to increase again in late 2008, this time caused by the 2008 Global Financial Crisis, reaching its peak of 10% in October 2009 as 2.6 million jobs were lost at this time (Uchitelle, 2009).

Overall, the time series plot is bimodal as it has two peaks—one in mid-2003 and another in late 2009. Besides, there seems to be slightly changing variance with non-constant mean in this series, indicating non-stationarity.

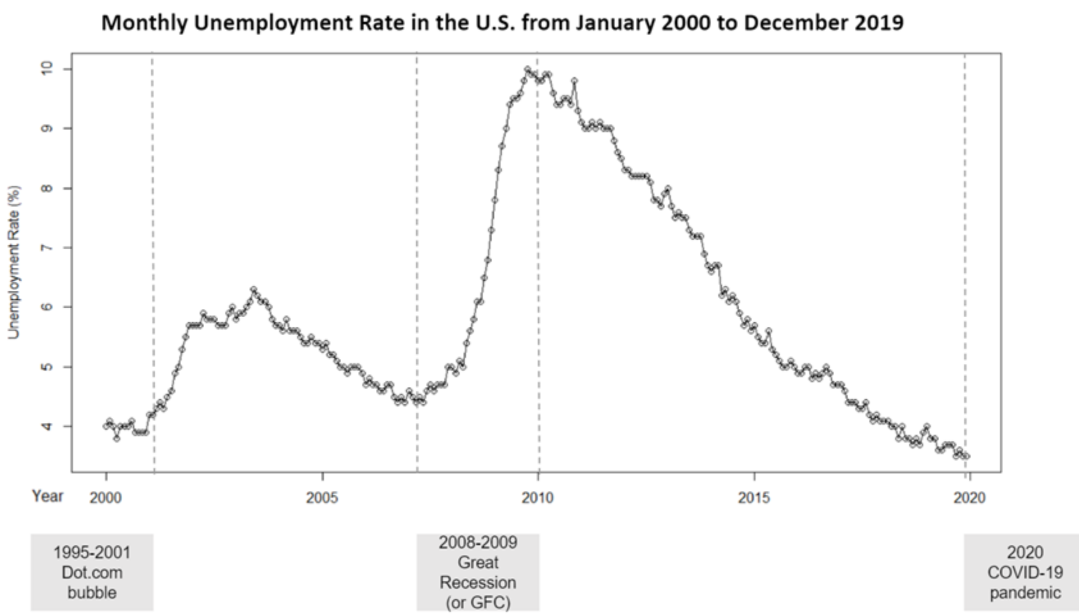

Figure 1. Time series plot of the monthly U.S. unemployment rate from Jan 2000 to Dec 2019

It is worth noting that values in the consecutive months tend to be close to one another, which suggests that the values might be related to each other. To further investigate this, the unemployment rate was plotted against the previous month's values shown in Figure 2. The scatter plot shows a positive correlation between consecutive values within the series. All value points seem to be on the same straight line, and each value is very close to each other, especially the values in the first half of the line. With R = 0.996, it indicates a very strong linear relationship between these values. It means the unemployment rate in one month is heavily influenced by the unemployment rate of the previous month.

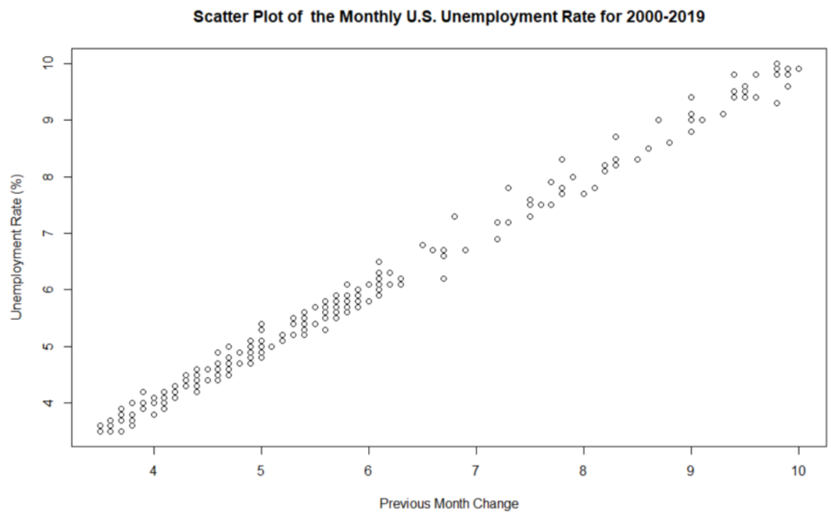

Figure 2. Scatter plot between the monthly data and previous month change for 2000-2019

There is no indication of the seasonal or cyclical trend in this time series as ACF of the time series displayed in Figure 3 does not show any wavy pattern. However, the relationship between consecutive points mentioned earlier suggests that there is an autoregressive behavior in this time series. Furthermore, the fluctuations could be an indication of a moving average behavior happening in the series.

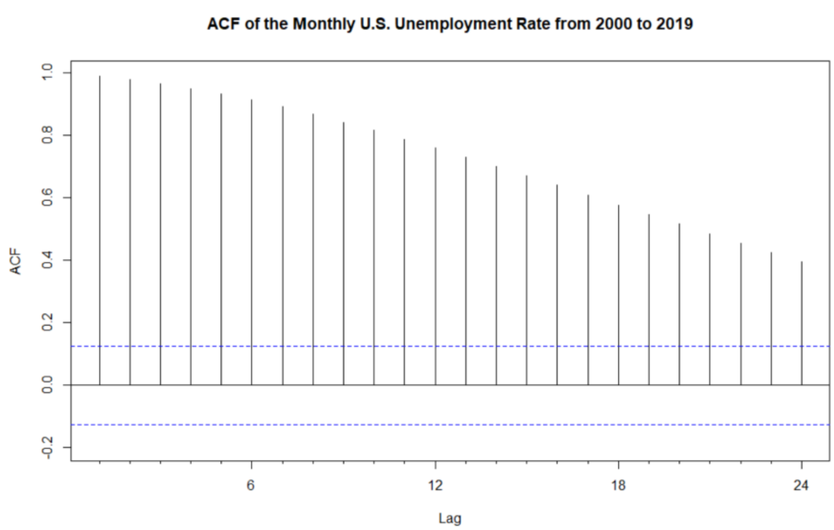

Figure 3. ACF for the 2000-2019 time series

To stress-test the accuracy of the forecast and assess all possible models, deterministic trends were fitted for the dataset.

1.2. Linear Trend Model

A deterministic linear trend model takes the following form, $\mu = \beta_0 + \beta_1$ where β_0 is the intercept and β_1 is slope of the linear trend. As shown in Table 1, the *p*-values of both intercept (0.788) and slope (0.901) are found to be higher than α = 0.05, indicating its insignificance.

	Estimate	Std. Error	t-value	p-value
$\hat{\beta}_0$	10.982452	40.778568	0.269	0.788
$\hat{\beta}_1$	-0.002539	0.020288	-0.125	0.901

Quantity	Values
Residual standard Error	1.815
R^2	6.578e-05
Adjusted R^2	-0.004136
F-statistic	0.01566

Table 1. Least squares regression estimates of a deterministic linear trend for 2000-2019

The linear trend model resulted in an R^2 of 6.578e-05 and an adjusted R^2 of -0.004. Since these two values are close to 0, the linear trend could not explain any of the variations in this series. It is further supported by the time series plot in Figure 4, which shows that the fitted line does not capture the mean values in the series.

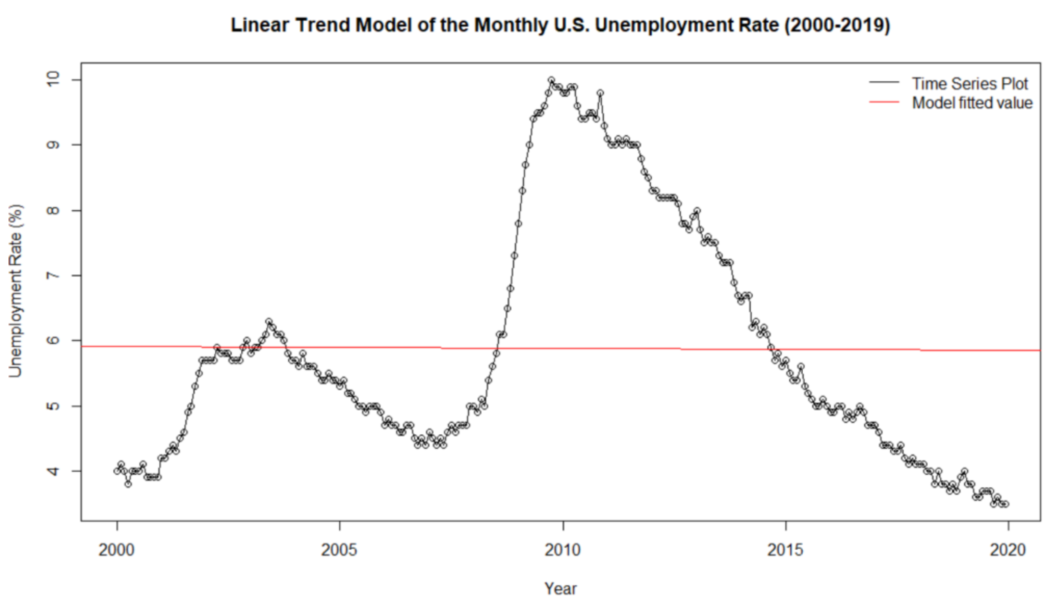

Figure 4. Time series plot with a fitted linear trend model for 2000-2019

Based on the results above, there is no need to analyze the residuals of linear trend model, as it can be concluded that the linear trend model is not a good model for the dataset.

1.3. Quadratic Trend Model

A deterministic quadratic trend model can be defined as $\mu = \beta_0 + \beta_1 t + \beta_2 t^2$ where β_0 is intercept, β_1 is the linear trend, and β_2 is the quadratic trend in time.

	Estimate	Std. Error	t-value	p-value
$\hat{\beta}_0$	-1.721e+05	1.131e+04	-15.21	<0.01
$\hat{\beta}_1$	1.713e+02	1.126e+01	15.21	<0.01
$\hat{\beta}_2$	-4.261e-02	2.800e-03	-15.21	<0.01

Quantity	Values
Residual standard Error	1.293
R^2	0.4941
Adjusted R^2	0.4899
F-statistic	115.7

Table 2. Least squares regression estimates of a deterministic quadratic trend for 2000-2019

The *p*-values of intercept (<0.01), linear trend (<0.01) and quadratic trend (<0.01) are statistically significant at $\alpha = 0.05$ (Table 2). The R^2 of 0.4941 and adjusted R^2 of 0.4899 suggests that the quadratic model can explain approximately half of the variation in the series. In comparison to the linear trend model, the fitted line of the quadratic trend seems to capture the mean values, and the data set better, as displayed in Figure 5. The residual standard error of the quadratic trend is also lower than the linear trend (1.293 < 1.815).

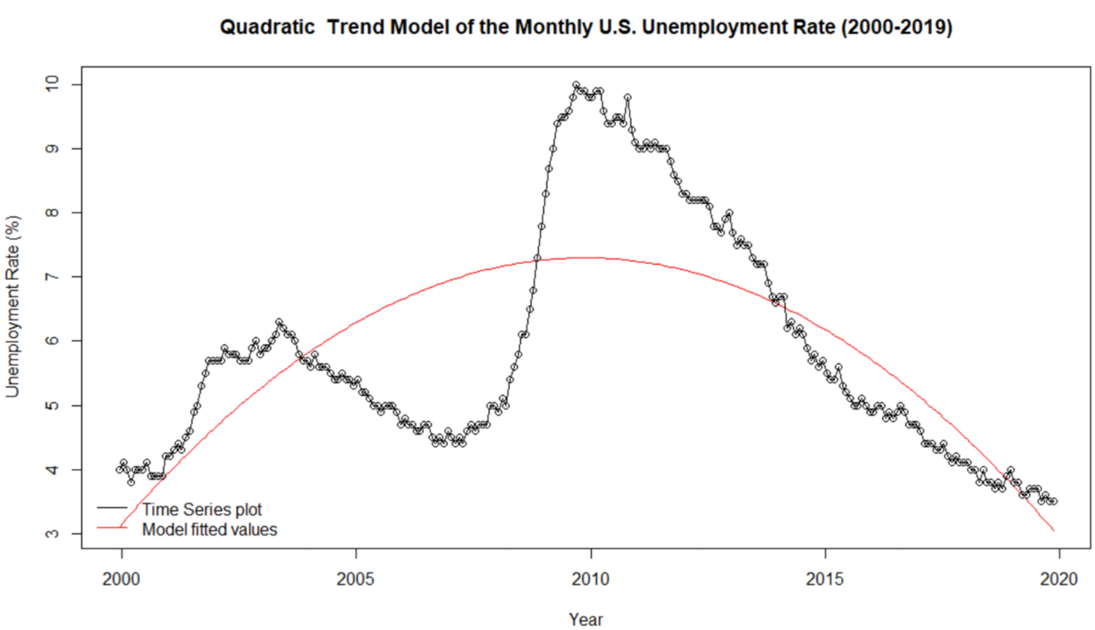

Figure 5. Time series plot with a fitted quadratic trend model for 2000-2019

The residuals were analyzed to confirm the goodness of fit of the quadratic trend model (Figure 6). The residual plot of the quadratic trend model shows the standardized residuals are in between the bounds of -2 and 2. However, there are u-shape patterns displayed, and most residuals are hanging together, indicating that it is not of a white noise process. The Residuals vs. Fitted Values plot does not show randomness since the standardized residuals do not spread around evenly. The test of normality also confirms the standardized residuals are not normally distributed with a *p*-value of 0.0007. Lastly, the ACF plot of the standardized residuals shows that it is not independent of each other, as there are significant autocorrelations across all lags.

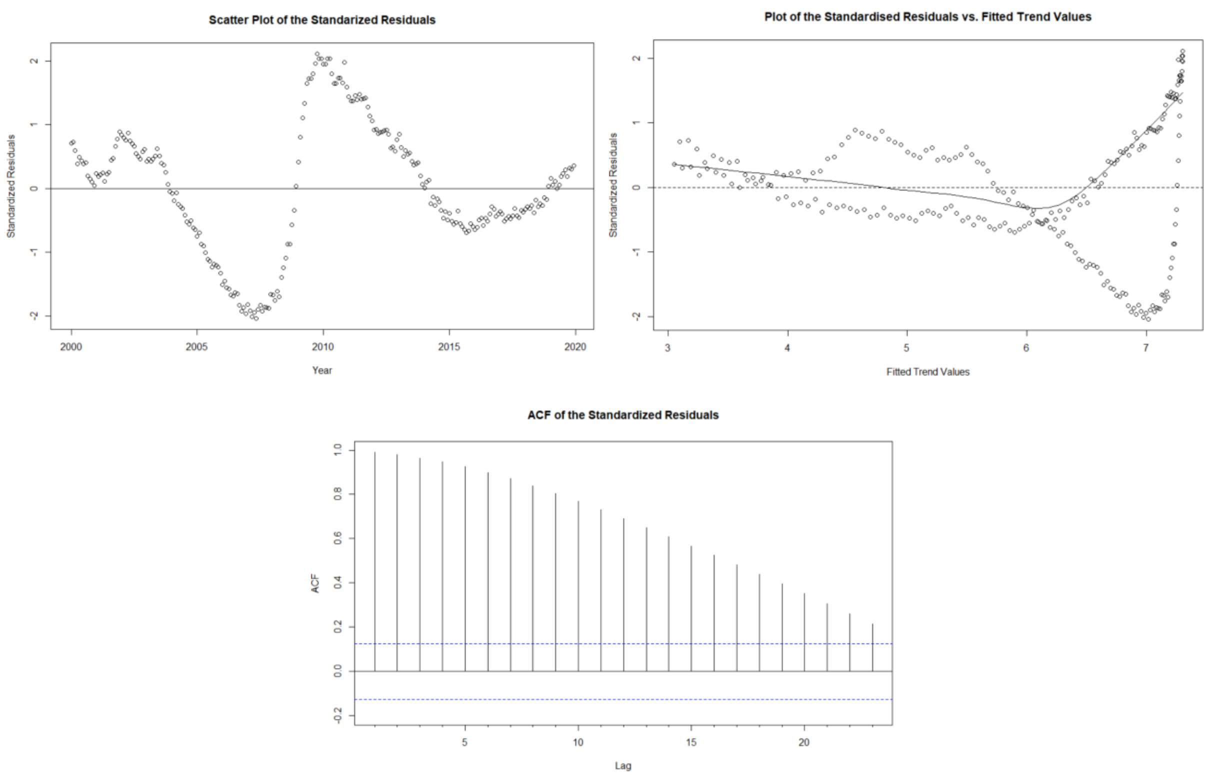

Figure 6. Standardized residuals for the fitted quadratic trend model for 2000-2019

In summary, it indicates that the quadratic trend model is not an excellent model to account for all patterns in the data adequately.

As both deterministic trend models were not identified to be the best fit, the stochastic trend model via ARIMA modeling will be fitted in the next section.

1.4. Stochastic Trend Model

1.4.1. Test of Normality and Stationarity

ACF and PACF were displayed to visualize the pattern of the series. The ACF and PACF plots in Figure 7 show a slow decay pattern in ACF, while PACF shows an exponential decay after one significant correlation at the first lag. With the slow decay of significant correlations in the ACF plot, it indicates non-stationarity in the time series. The Augmented Dickey-Fuller Unit-Root Test was also used to check the stationarity of the data, which returned a *p*-value 0.544, which is higher than α = 0.05. The ADF test results mean that we will fail to reject the null hypothesis, confirming the non-stationarity of the time series.

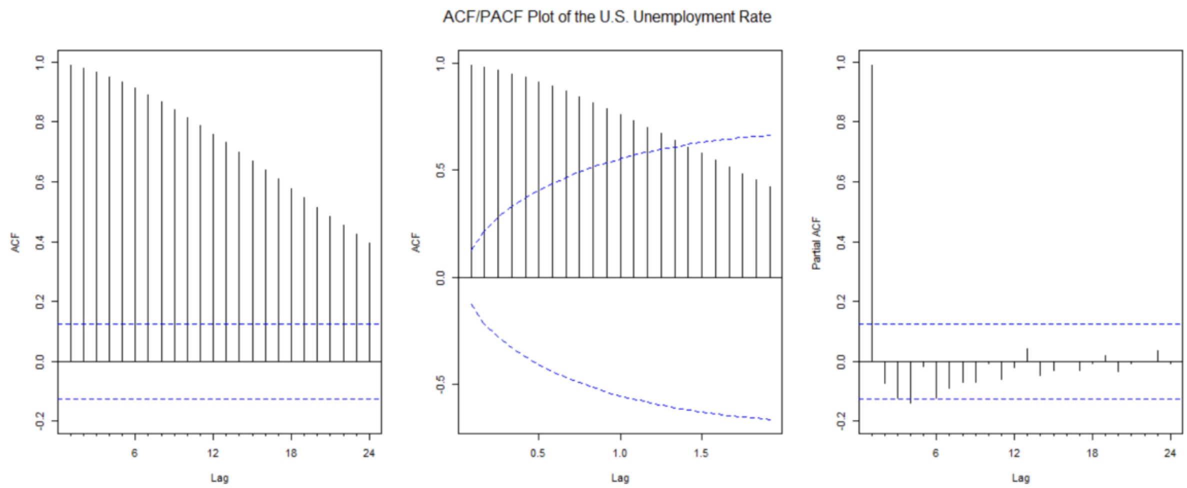

Figure 7. ACF and PACF plots of the U.S. unemployment rate for 2000-2019

A Shapiro-Wilk test was used to check the normality of the time series data. The result of the test returned a *p*-value of <0.01 (i.e., 2.696e-12), which is lower than 0.05 supporting the non-normality of the series. Therefore, the Box-Cox transformation was applied. To find the optimal value of the lambda (λ) parameter in the Box-Cox transformation, a relative log likelihood plot was used. From the plot in Figure 8, the lambda equals to -0.709 was chosen.

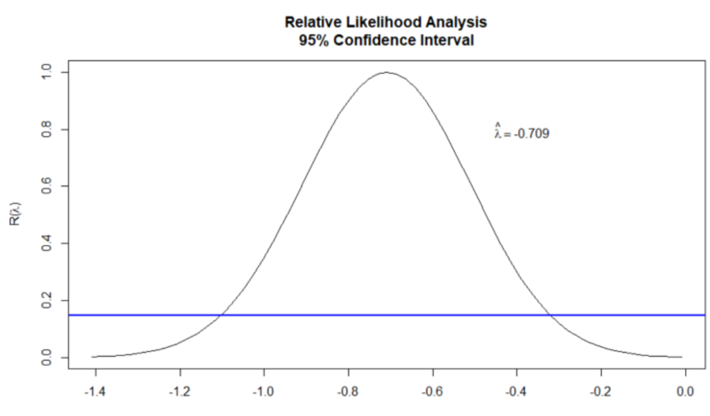

Figure 8. Log Likelihood plot for 2000-2019

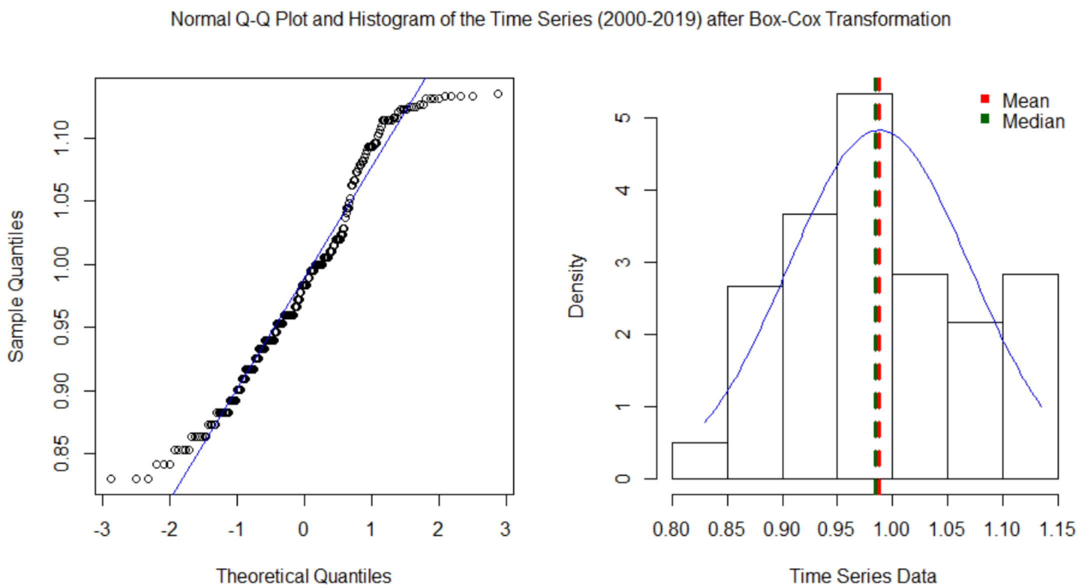

Figure 9. Normal Q-Q plot and histogram after Box-Cox transformation for 2000-2019

Figure 9 illustrates the histogram and normal Q-Q plot of the data after the Box-Cox transformation. Both plots still show non-stationarity. The histogram does not show a perfect symmetry while the normal Q.Q. plot having both tails going away from the normal quantile line. The *p*-value of Shapiro Wilk's test after Box-Cox transformation resulting in <0.01 (i.e., 4.454e-06), confirmed non-normality in the transformed data set.

The transformation has slightly improved the normality of the data, as indicated by its respective *p*-values. Hence, further analysis was done using the transformed data.

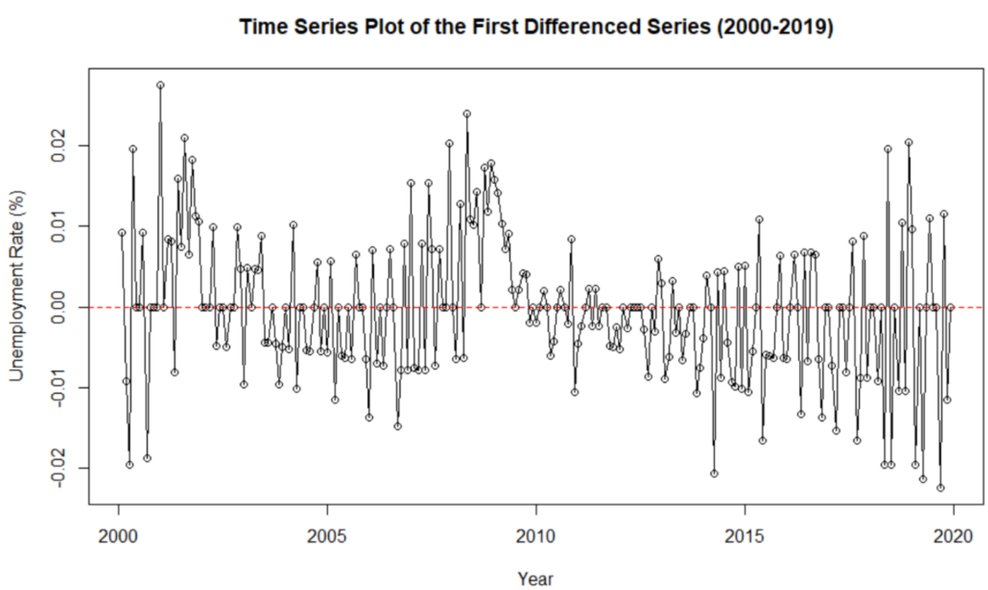

Figure 10. Time series plot of the transformed series (2000-2019) after first differencing

The first differencing was applied to de-trend the data. ADF test resulted in non-stationarity with a *p*-value of 0.394, failing to reject the null hypothesis. After the first differencing, the trend seems to have disappeared, as shown in Figure 10, barring the intervention points in the early 2000s and 2008-2010 period. According to the result from the ADF test, *p*-value returns <0.01, which is statistically significant at 5%, which confirms stationarity of the time series.

1.4.2. Model Specification

Since the series had been de-trended, the specification of the models can be checked to determine the value of parameters p and q. ACF/PACF plot, EACF plot, and BIC plot were used to specify the orders of p and q in the model.

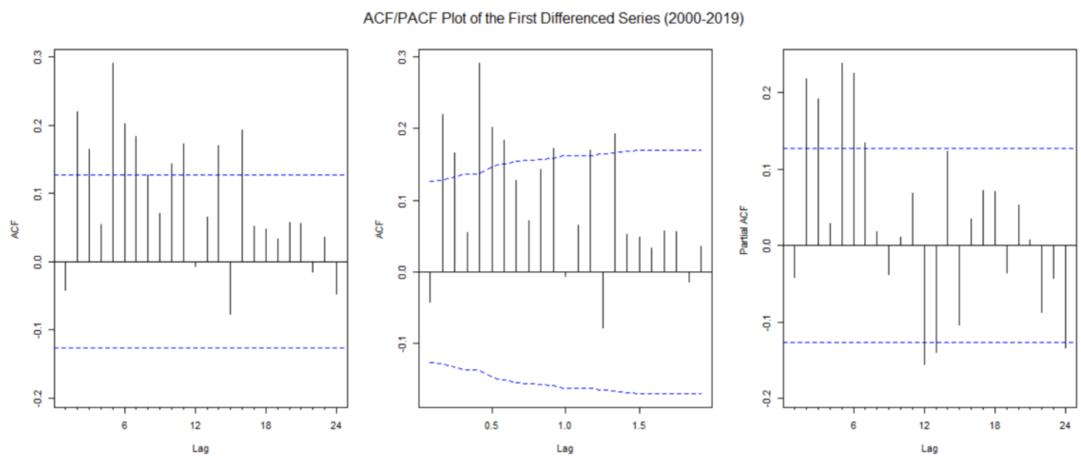

Figure 11. ACF and PACF of the transformed series (2000-2019) after first differencing

The ACF and PACF plots in Figure 11 have significant autocorrelations at various lags, indicating both A.R. (p) and M.A. (q) components in the dataset. The ACF with the modified interval bounds shows that there are 6 significant autocorrelations. Hence, the possible value of q is 6. PACF displays a damped sine wave with 5 significant autocorrelations, indicating p = 5. Given this, a tentative candidate ARIMA model was identified to be ARIMA(5,1,6). Because ACF and PACF plots are not the most effective method to determine ARMA models, the EACF table is used to identify the order of p and q further.

AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	o	x	x	o	x	x	x	o	o	x	x
1	x	x	o	o	x	o	o	o	o	o	x
2	x	x	o	o	x	x	o	o	o	o	x
3	o	x	x	o	x	o	o	o	o	o	o
4	o	x	x	o	x	o	o	o	o	o	o
5	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	o	x	o	x	o	o	o	o
7	o	x	o	o	o	o	x	o	o	o	o
8	x	x	x	o	o	o	o	x	o	o	o
9	x	x	o	o	o	o	o	o	o	o	o
10	x	x	o	o	o	o	o	o	o	x	o

Table 3. EACF table of the transformed series (2000-2019) after first differencing

Possible vertices in the EACF Table 3 have been highlighted in yellow. The row corresponding to p = 1 and q = 5 shows a possible vertex, despite an x mark at q = 10. A second vertex at p = 3 and q = 5 is also highlighted. The ones in blue are the models surrounding the vertices, which were also considered. The possible candidate models from the EACF table are ARIMA(1,1,5), ARIMA(3,1,5), ARIMA(3,1,6) and ARIMA(4,1,5).

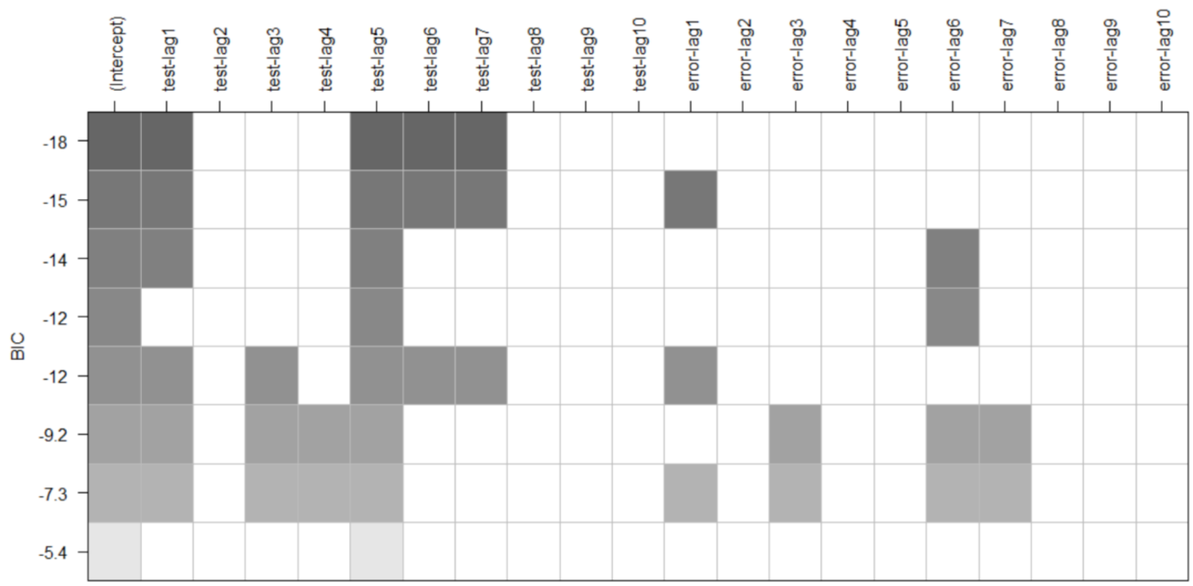

Table 4. BIC table of the transformed series (2000-2019) after first differencing

Table 4 highlights the Bayesian Information Criterion. The most shaded columns at the top row in the BIC table represent the best model while the one in the second row represents the second-best model. The columns 'test-lag' indicates the order of the autoregressive process, whereas the columns 'error-lag' indicates the order of moving average process. According to this, the shaded columns top row correspond to AR (1), AR (7), and MA (0), while the shaded columns in the second row correspond to AR (1), AR (7), and MA (1). Thus, ARIMA (1,1,0), ARIMA (7,1,0), ARIMA (1,1,1), and ARIMA (7,1,1) were included in the set of possible candidates.

A set of candidates of ARIMA model is the intersection of the set of possible candidates resulted in ACF, PACF plots, EACF, and BIC tables. Therefore, the set of candidate models are {ARIMA (1,1,5), ARIMA (3,1,5), ARIMA (3,1,6), ARIMA (4,1,5), ARIMA (5,1,6), ARIMA (7,1,0) and ARIMA (7,1,1)}.

Candidate models	ACF/PACF plots	EACF plot	BIC table
ARIMA (1,1,5)		√	
ARIMA (3,1,5)		√	
ARIMA (3,1,6)		√	
ARIMA (4,1,5)		√	
ARIMA (5,1,6)	√		
ARIMA (7,1,0)			√
ARIMA (7,1,1)			√

Table 5. Summary of *the* candidate *ARIMA* models

1.4.3. Model Fitting and Parameter Estimation

Maximum Likelihood Estimates (MLE) is considered the best parameter estimator and very efficient for a large sample data. As the dataset of this study consists of n=240, it is large enough to rely on the MLE method.

After the orders of the autoregressive and moving average of elements of ARMA models were specified, parameters of the specified tentative candidate models were estimated based on MLE with a coefficient significance test.

The summary of results from the coefficient test is as follows. For the full results, refer to Appendix B.

- Models retained for further residual analysis:

  ARIMA (1,1,5) – All coefficients of this model are statistically significant except for the coefficients of MA (3) and MA (4).

  ARIMA (4,1,5) – All coefficients of this model are statistically significant except for the coefficients of MA (4) and MA (5).

- Models excluded from the residual analysis:

  ARIMA (3,1,5) – All coefficients of this model are statistically insignificant except for the coefficients of AR (1), MA (1) and MA (5).

  ARIMA (3,1,6) – The coefficient test produced N.A.s.

  ARIMA (5,1,6) – The coefficient test produced N.A.s.

  ARIMA (7,1,0) – Only coefficients of AR (3), AR (4), and AR (7) are insignificant. Since there are only 4 significant parameters, the model is not retained.

  ARIMA (7,1,1) – Only coefficient of AR (5) and AR (6) are statistically significant.

All in all, five out of the seven candidate models have insignificant parameters. Therefore, the possible candidates which can be proceeded include {ARIMA (1,1,5) and ARIMA (4,1,5)}.

1.4.4. Residual Analysis

This section focuses on the residual analysis of the two candidate models identified above. The criteria for determining the goodness of a model, according to Hyndman and Athanasopoulos (2018) are as follow:

- Residuals should have a zero mean to avoid forecast bias.
- Residuals should have a constant variance.
- Residuals are normally distributed.
- Residuals should be uncorrelated and have a resemblance to the properties of white noise. Otherwise, there is a need to re-evaluate the estimations.

ARIMA (1,1,5)

Figure 12 illustrates the plots of the residuals. The top left plot shows the standardized residuals fluctuates around zero. It also shows some randomness since the residuals somehow spread around the rectangular. The plot seems to suggest no trend presented. The independence of the residuals can be confirmed by the Runs test where the *p*-value is 0.138 > 0.05, meaning the null hypothesis that the residuals are independent cannot be rejected. Moreover, the histogram of standardized residuals looks symmetric and Q.Q. plot shows all points follow a straight line. The Shapiro–Wilk test returned a *p*-value of 0.0595, which is on the line. Normality can somehow be assumed for residuals of this model. There is no significant autocorrelation in the ACF of standardized residuals. The Ljung-Box Test indicated the p-value of 0.503 > 0.05, confirming all correlations equals to zero. These results suggest ARIMA (1,1,5) has properties of white noise and can be an appropriate fitting model.

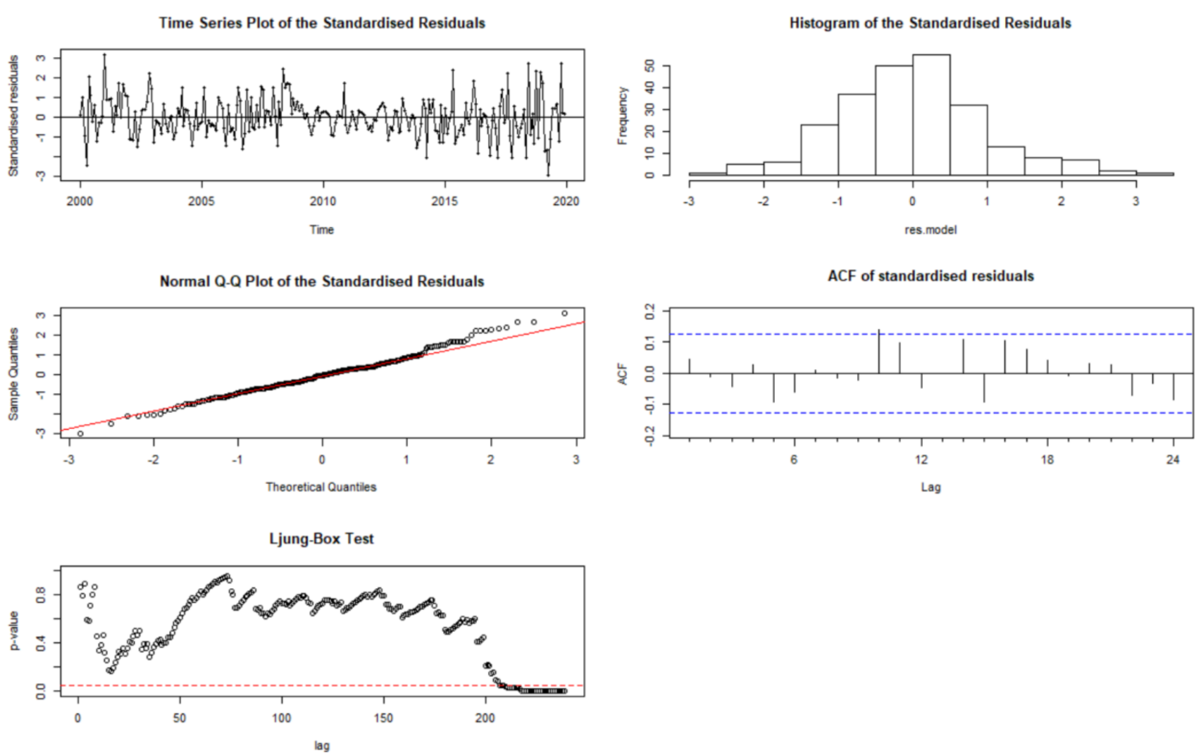

Figure 12. Residual analysis of ARIMA (1,1,5) for 2000-2019

ARIMA (4,1,5)

Similarly, the residual analysis of ARIMA(4,1,5) shows similar result to that of ARIMA (1,1,5). As shown in Figure 13, the standardized residuals fluctuate around zero. It also shows some randomness since the residuals somehow spread around the rectangular. The Runs test supports this claim with a *p*-value greater than 0.5, suggesting there is not enough statistical evidence to reject the independence of the residuals. The histograms are not perfectly symmetric and Q.Q. plots indicate points follow a close straight line with tails go a bit off the curve, which makes the normality is quite questionable. However, the Shapiro-Wilk Test of the standardized residuals returned *p*-value > 0.05 for the three models indicating they follow the normal distribution. Likewise, there is also no significant autocorrelation presented in the ACF of standardized residuals, which is confirmed by the Ljung-Box Test (p-value > 0.05). Based on the results, ARIMA(4,1,5) has the properties of white noise and can also be an appropriate fitting model.

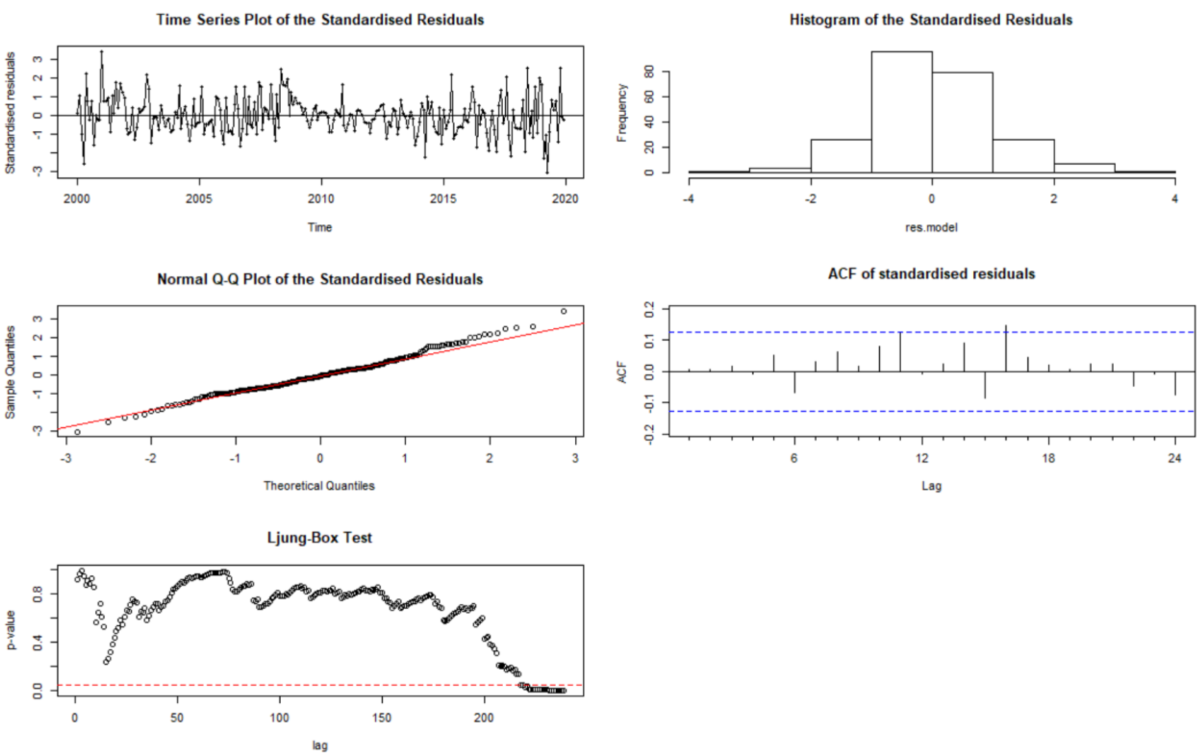

Figure 13. Residual analysis of ARIMA (4,1,5) for 2000-2019

1.4.5. Model Selection

After residual diagnostic check confirmed independence, normality, and no significant autocorrelation of the four models, the values of Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were computed and ranked to select the best model.

AIC assesses the quality of each model relative to one another. The lowest AIC can be considered as closest to the truth. BIC is an estimator of a function of the posterior probability, which means that the lowest BIC can be deemed to be more likely the true model. Corrected Akaike Information Criterion (AICc) was also used in addition to AIC.

Model	Df	AIC	AICc	BIC
model.115	7	-1639.917	-1641.432	-1615.582
model.415	10	-1639.053	-1640.088	-1604.288

Table 6. Model comparison for the 2000-2019 series

As indicated in the results (Table 6), it can be concluded that ARIMA (1,1,5) is the best model having the smallest AIC, AICc, and BIC values as opposed to ARIMA (4,1,5).

1.4.6. Overfitting

After choosing an adequate model, slightly general models or “close by” models that contain the original model should be fitted to detect anomalies in terms of the goodness of fit.

ARIMA (1,1,5) was chosen as the best based on AIC and BIC values. Thus, ARIMA (2,1,5) and ARIMA (1,1,6) are overfitting of ARIMA (1,1,5). However, the coefficient of AR (2) of ARIMA (2,1,5) was found insignificant as the *p*-value of the coefficient of AR (2) equals to 0.07, larger than 0.05 (See Appendix B). For ARIMA (1,1,6), coefficient’s p-value of MA (6) = 0.072 > 0.05, so it is not statistically significant either. It confirms ARIMA (1,1,5) is an adequate and the best model of the series. For the results of the coefficient tests, refer to Appendix B.

1.5. Ten-month Forecast

ARIMA (1,1,5) was used to predict the unemployment rate in the U.S for ten months from January 2020 to October 2020. The predicted values in millions with 95% confidence intervals are displayed in Table 7. The unemployment rate forecasted a slight increase in the first month, decreases a bit in the second month, and then rises gradually for the next eight months. Compared to the actual data, the forecast seems to capture the real scenario for the first two months in 2020.

However, the global pandemic of COVID-19 has abruptly changed the situation. In March 2020, the U.S. unemployment rate rose to 4.4%, higher than the predicted value of 3.5%. Then, it steeply increased to 14.7% in April 2020.

Figure 14 illustrates a slightly upward trend of the unemployment rate for the first ten months versus the real data of the first four months of 2020. Based on visual analysis, the forecast does not seem to follow the general downtrend of unemployment – possibly due to the two intervention points in the time series – which suggests that the model is not adequate.

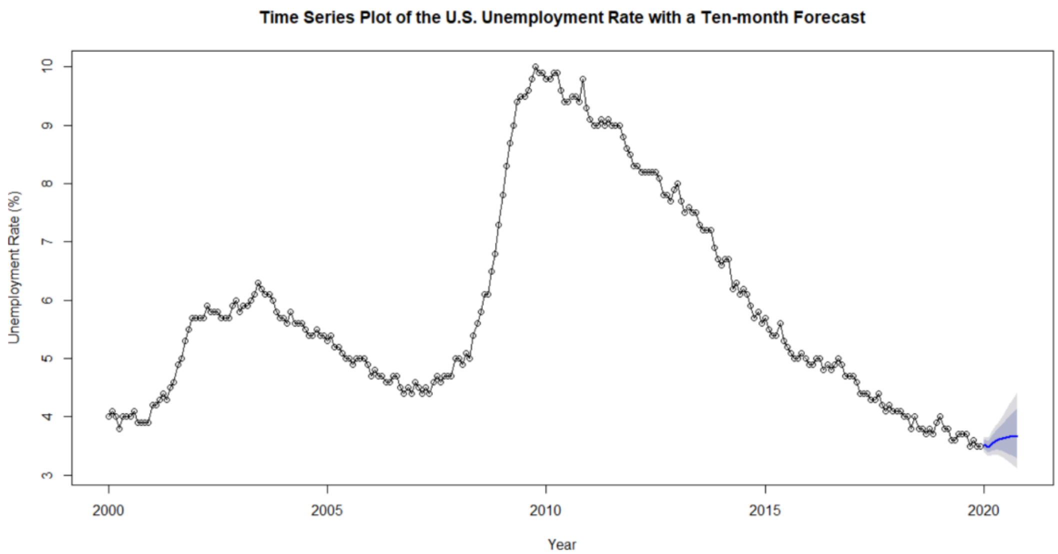

Figure 14. Ten-month forecast for the 2000-2019 time series

Year	Forecast	Low 95%	High 95%	Actual Data
January 2020	3.525	3.400	3.658	3.6
February 2020	3.486	3.334	3.650	3.5
March 2020	3.540	3.348	3.752	4.4
April 2020	3.582	3.349	3.845	14.7
May 2020	3.615	3.349	3.919	
June 2020	3.637	3.317	4.013	
July 2020	3.651	3.269	4.117	
August 2020	3.661	3.216	4.222	
September 2020	3.667	3.163	4.326	
October 2020	3.671	3.113	4.426	

Table 7. Ten-month forecast for the 2000-2019 time series and actual 2020 U.S. unemployment rate figures

To overcome this, the next chapter will focus on modeling the time series from 2010 to 2019 to get rid of the effects of the intervention points. The ten-month forecast will then be compared to see how it performs.

Chapter 2: Monthly U.S. Unemployment Rate from 2010 to 2019

2.1. Descriptive Analysis

Following the recommendation from Chapter 1: Monthly U.S. Unemployment Rate from 2000 to 2019, Chapter 2 will focus on modeling and forecasting the U.S. unemployment rates after the 2008 Global Financial Crisis. Data used for the modeling will be from January 2010 to December 2019. The resulting best-fitting model would then be chosen to forecast the unemployment rates for the next ten months, mainly January 2020 to October 2020.

The mean and standard deviation of this time series is 6.22% and 2.06%, respectively. Figure 15Figure 15 shows the time series plot of the monthly U.S. unemployment rate from January 2010 to December 2019. There appears to be an apparent decreasing trend in general and a non-constant mean that indicates non-stationarity. Moreover, there are no drastic changes in variance during this period.

The fluctuations in this series suggest that there is moving average behaviour in this time series. It is notably seen that the neighboring observations tend to cluster very closely together, which indicates that their values might be related. In other words, this is indicative of extreme autoregressive behaviour happening in the time series.

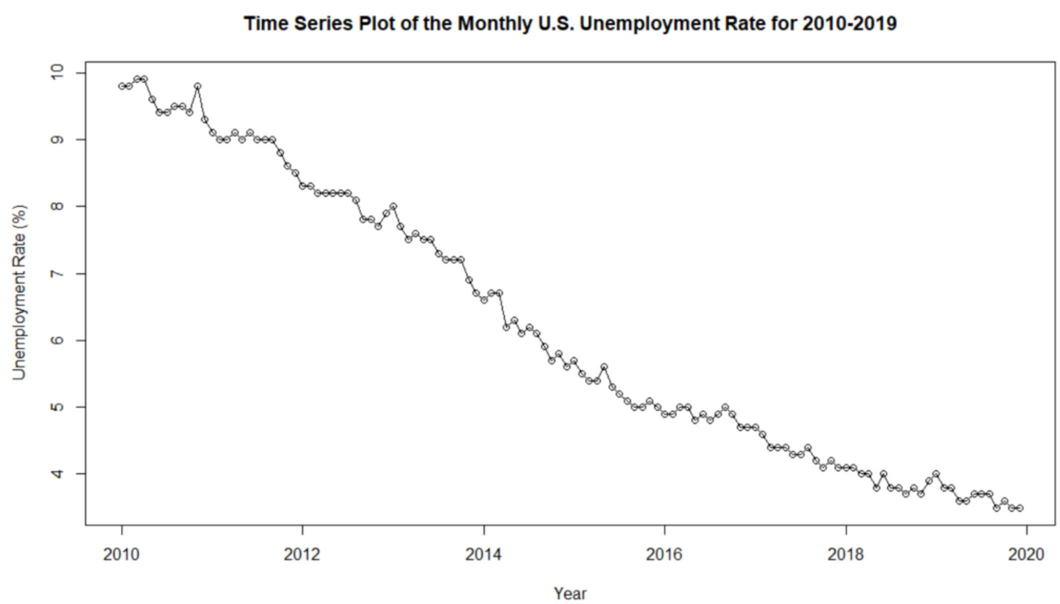

Figure 15. Time series plot of the monthly U.S. unemployment rate from Jan 2010 to Dec 2019

To further investigate this, the unemployment rate was plotted against the previous month's values as shown in Figure 16. The results show a very strong positive correlation of 0.998, which suggests that the unemployment rate in one year depends heavily on the unemployment rate in the previous year.

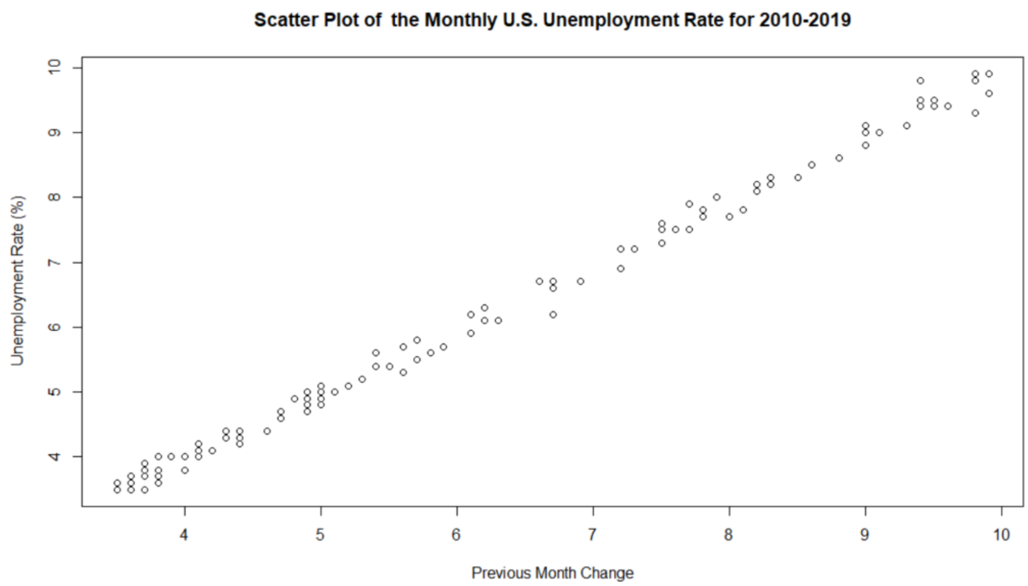

Figure 16. Scatter plot between the monthly data and previous month change for 2000-2019

The ACF plot of the time series shows a slow decay in the significant autocorrelations, indicating supporting the conclusion of non-stationarity upon the earlier visual analysis of the time series (Figure 17). Besides, no wave pattern suggests the non-existence of a cyclical trend within this time series.

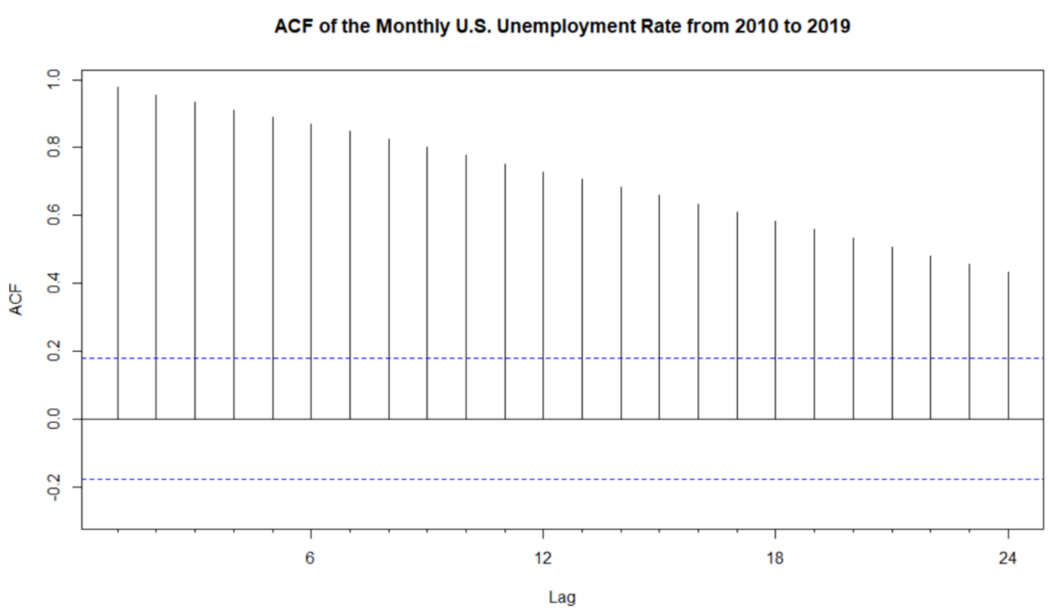

Figure 17. ACF for the 2010-2019 time series

2.2. Linear Trend Model

Similar to Chapter 1, a linear trend model will be fitted to ensure that all model types are considered in the modeling process. The linear trend model can be written as $\mu_t = \beta_o + \beta_1 t$, where β_o and β_1 stands are the intercept and the slope of the model, respectively.

	Estimate	Std. Error	t-value	p-value
$\hat{\beta}_0$	1418.28165	23.43948	60.51	<0.01
$\hat{\beta}_1$	-0.70079	0.01163	-60.24	<0.01

Statistics	Values
Residual Standard Error	0.3678
R^2	0.9685
Adjusted R^2	0.9682
F-statistic	3629

Table 8. Least squares regression estimates of a deterministic linear trend for 2010-2019

Based on the R^2 and adjusted R^2, the model can explain almost 97% of the variation observed in the very promising series. β_1 = -0.7 can be interpreted as for every one unit increase in time, the unemployment rate is expected to decrease by 0.7%. The *p*-values for both parameters β_o and β_1 are less than the predetermined significance level of 0.05. Therefore, these two parameters make significant contributions to the model.

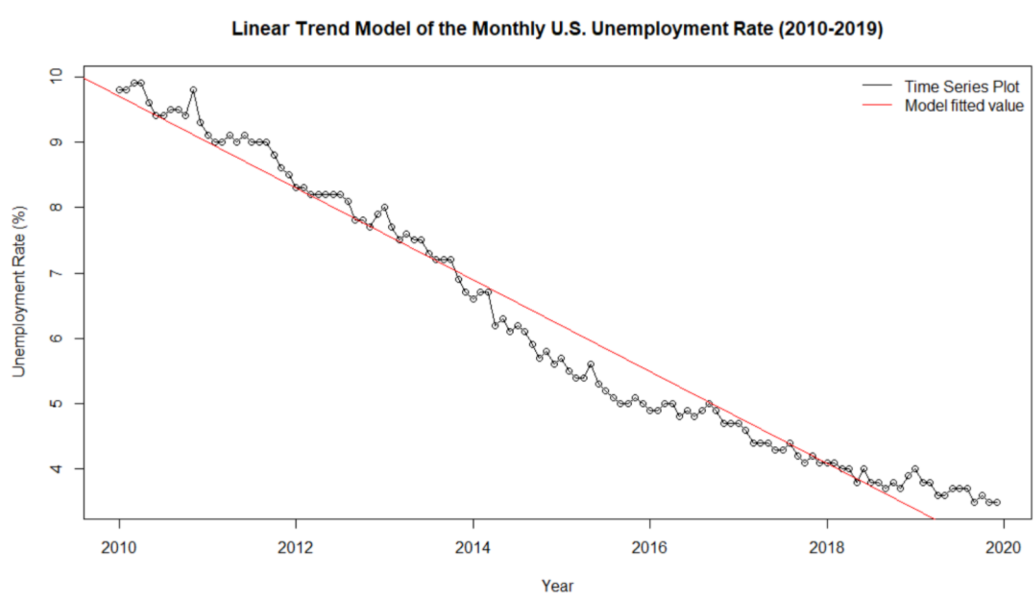

Figure 18. Time series plot with a fitted linear trend model for 2010-2019

Figure 18It is observed that there is a distinct upward u-shaped trend from the plot of standardized residuals (Figure 19). A *p*-value from the non-constant variance score test is 0.005, which is smaller than 0.05, meaning that we reject the null hypothesis that the variance is constant. This pattern suggests that the linear trend does not adequately explain the quadratic relationship.

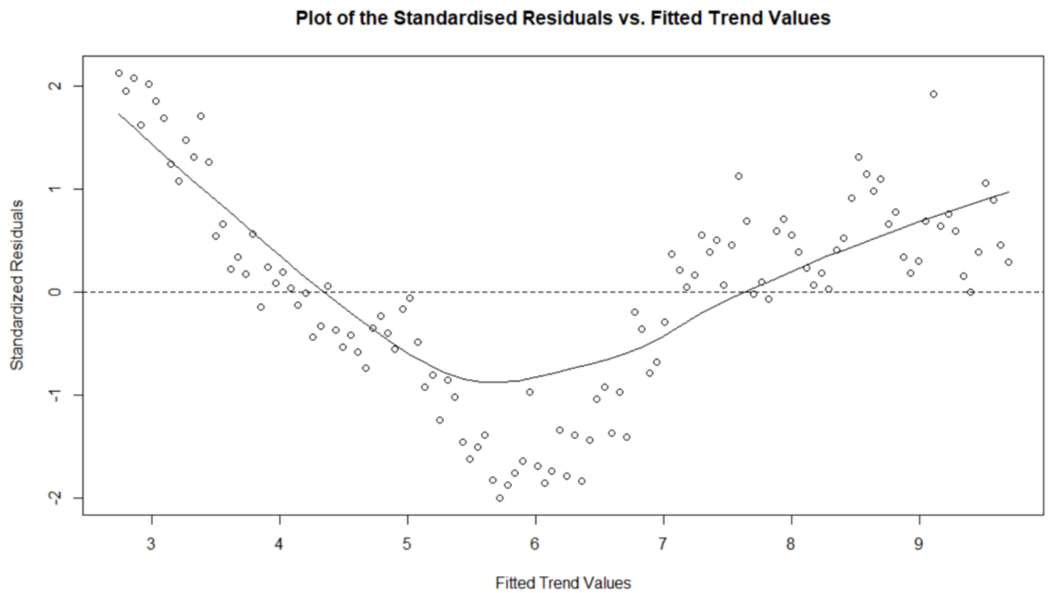

Figure 19. Standardised Residuals versus Fitted Trend Values (2010-2019 Series)

This can be confirmed by the Runs test, where a *p*-value of <0.01 (i.e. 2.19e-20), which is smaller than 0.05, leads us to reject the null hypothesis that the residuals are independent. We can, therefore, say that the residuals are dependent and thus, violating one of the assumptions behind the trend model.

A Shapiro-Wilk test also confirms that with a *p*-value of 0.03, it leads us to reject the null hypothesis that residuals are normally distributed. Since the residuals are not independent and not normally distributed, it is not a valid model for the time series despite earlier results that show that the model can explain almost 97% of the variance in the data set.

2.3. Quadratic Trend Model

A quadratic trend model in the form of $\mu_t = \beta_o + \beta_1 t + \beta_2 t^2$ was also fitted to the data, based on the results from Figure 19, which shows a curvilinear pattern in the standardized residuals. β_o, β_1 and β_2 stands for the intercept, the slope and quadratic trend of the model, respectively.

	Estimate	Std. Error	t-value	p-value
$\hat{\beta}_0$	1.573e+05	1.140e+04	13.80	<0.01
$\hat{\beta}_1$	1.554e+02	1.131e+01	-13.74	<0.01
$\hat{\beta}_2$	3.839e-02	2.807e-03	13.68	<0.01

Quantity	Values
Residual Standard Error	0.2291
R^2	0.9879
Adjusted R^2	0.9877
F-statistic	4770

Table 9. Least squares regression estimates of a deterministic quadratic trend for 2010-2019

Based on the R^2 and the adjusted R^2, the model can explain approximately 98% of the variation observed in the series. The *p*-values for all parameters are less than the predetermined significance level of 0.05. Therefore, these three parameters make a significant contribution to the model.

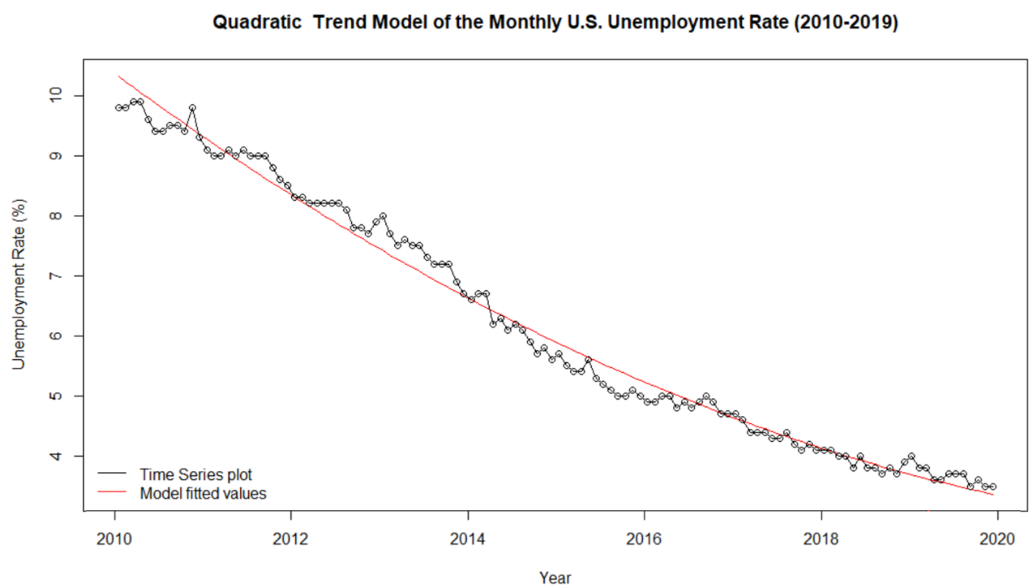

Figure 20. Time series plot with a fitted quadratic trend model for 2010-2019

The residuals were also analyzed to confirm the good fit of the quadratic trend. From Figure 21Figure 20**Error! Reference source not found.Error! Reference source not found.**, it is observed that there is a distinct pattern to how residuals are spread around a horizontal line. A *p*-value from the non-constant variance score test is 0.0003, which is smaller than 0.05, meaning that we reject the null hypothesis that the variance is constant. Similar to the linear trend model, the relationship between standardized residuals and fitted values by the quadratic trend model is curvilinear. This shape suggests that the quadratic trend model still does not capture the relationships in the observed values.

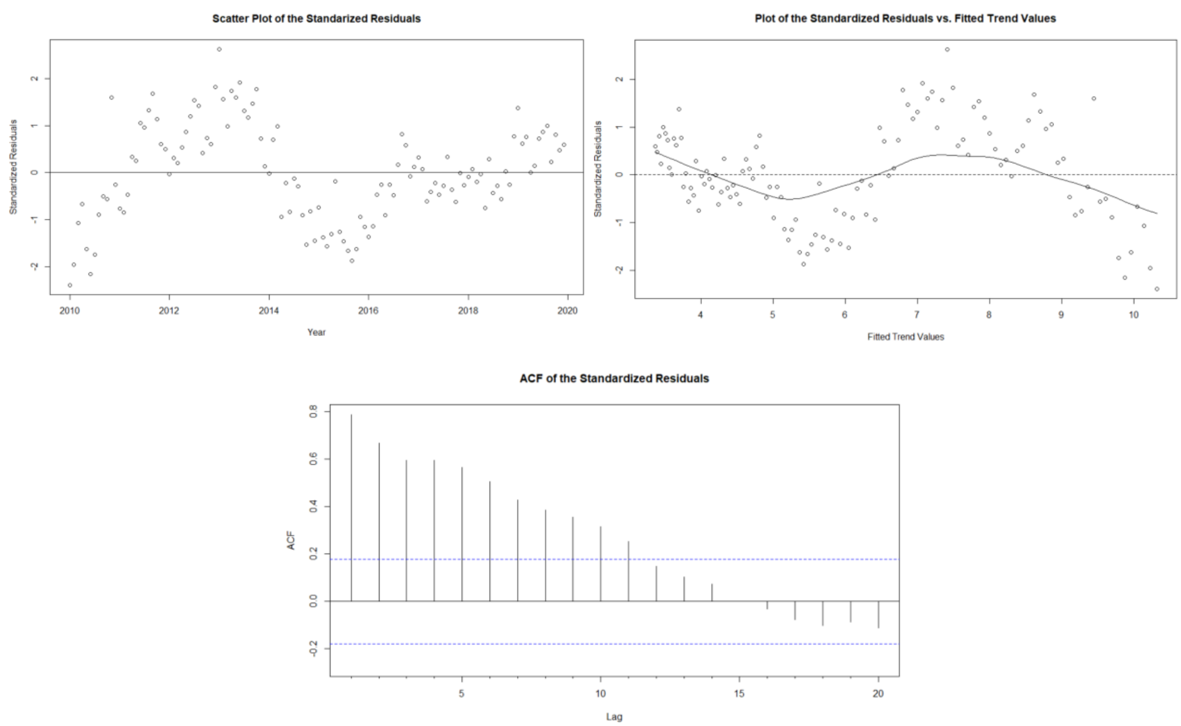

Figure 21. Standardised residuals for the fitted quadratic trend model for 2010-2019

Meanwhile, upon running the Runs test, a *p*-value of 2.15e-13, which is smaller than 0.05, leads us to reject the null hypothesis that the residuals are independent and thus conclude that the residuals are indeed dependent. This violates one of the assumptions behind the deterministic trend model.

Although both models explain variations observed in the series well and a Shapiro-Wilk test, with a *p*-value of 0.859 failing to reject the null hypothesis of normality, none of it met the assumptions of independence and identically distributed residuals. Consequently, stochastic trend models (ARIMA) were fitted in the next section.

2.4. Stochastic Trend Model

2.4.1. Test of Normality and Stationarity

Figure 22 illustrates that the unemployment rate series is non-stationary, based on the many significant correlations that are slowly decaying over the lags. Furthermore, the absence of the wave pattern indicates the non-existence of the cyclical trend. The Augmented Dickey-Fuller Unit-Root Test was used to check the stationarity of the data. The test returned a *p*-value of 0.99, which is higher than 0.05, indicating it fails to reject the null hypothesis stating non-stationarity.

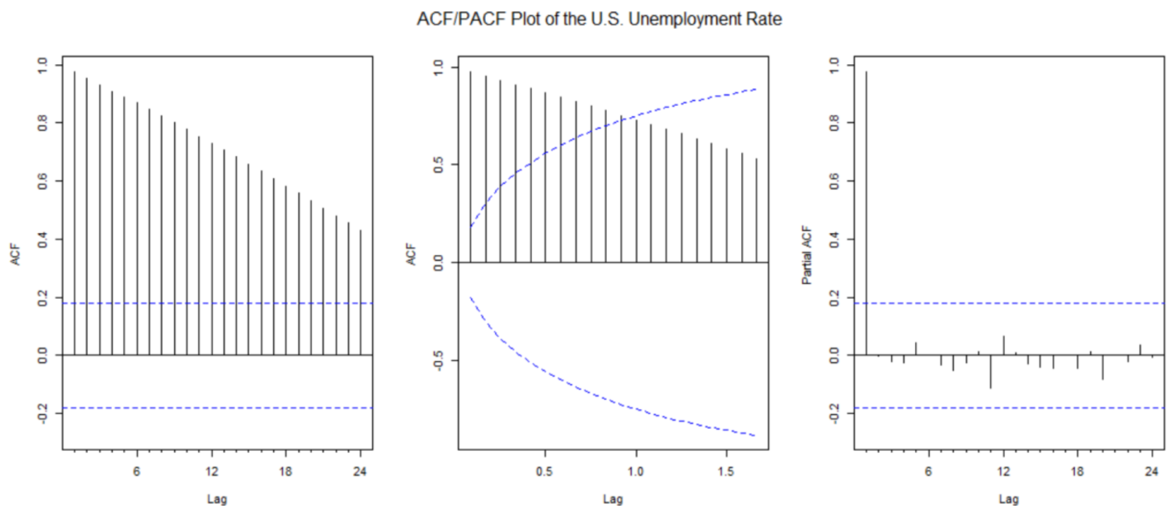

Figure 22. ACF and PACF plots of the U.S. unemployment rate for 2010-2019

A Shapiro-Wilk test was used to check the normality of time series data. The result of the test returned a *p*-value of 4.1e-07, which is lower than 0.05 supporting the non-normality of the series. Therefore, the Box-Cox transformation was applied. To find the optimal value of the lambda (λ) parameter in Box-Cox transformation, a log-likelihood plot was used. From the plot in Figure 23, lambda equals 0.45 was chosen.

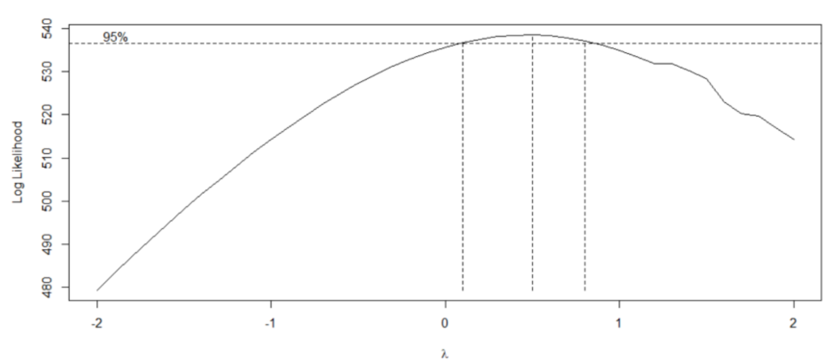

Figure 23. Log-likelihood plot for 2010-2019

Normality was also checked after the transformation. A Shapiro-Wilk test resulted in a *p*-value of 1.478e-06, which is still lower than 0.05. While it did not effectively improve the normality of the series, the slightly improved *p*-value compared to the original time series was enough to proceed with the transformation for the application of time series differencing.

From the result of the ADF test, a *p*-value returns 0.01, which is statistically significant at a 5% significance level. Thus, it can be concluded that the transformed series is stationary after the second-order differencing. As shown in Figure 24, the trend seems to disappear. Since the series has been de-trended, model specification methods can be used to determine the value of parameters p and q.

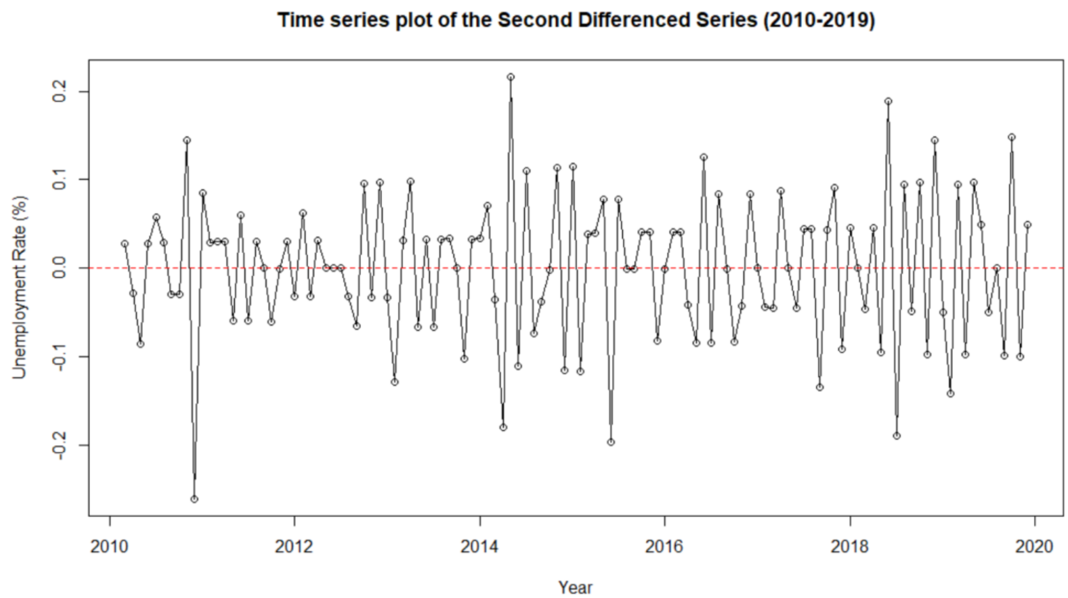

Figure 24. Time series plot of the transformed series (2010-2019) after second differencing

2.4.2. Model Specification

The ACF and PACF plots, EACF and BIC plots were used to investigate the orders of ARIMA (p, q) model. According to the plots in Figure 25, ACF shows a significant correlation at lag one while PACF tails off after lag one. From this result, a candidate ARIMA (0,2,1) can be inferred.

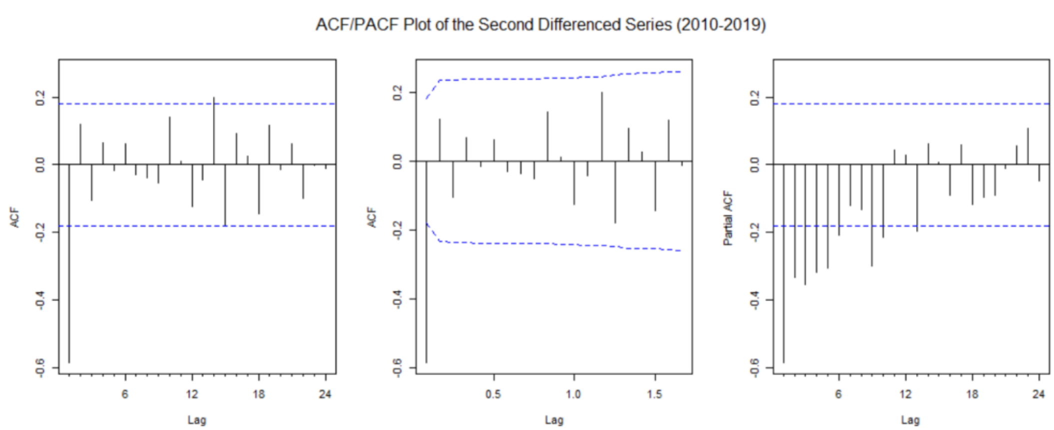

Figure 25. ACF and PACF of the transformed series (2010-2019) after second differencing

Because ACF and PACF plots are not by themselves significantly useful alone in specifying ARIMA models, the EACF table is used to further assist in identifying the orders of p and q.

AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	x	o	o	o	o	o	o	o	o	o	o
1	x	x	o	o	o	o	o	o	o	o	o
2	x	x	x	o	o	o	o	o	o	o	o
3	x	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o
5	x	o	x	o	o	o	o	o	o	o	o
6	x	x	o	x	o	o	o	o	o	o	o
7	x	o	o	x	o	o	o	o	o	o	o
8	x	x	x	x	o	o	o	o	o	o	o
9	x	x	o	o	o	o	o	o	o	o	o
10	x	o	x	o	o	o	o	o	o	o	o

Table 10. EACF table of the transformed series (2010-2019) after second differencing

The EACF table in Table 10Table 10. EACF Table 10 shows a clear vertex at the row corresponding to p = 0 and q = 1. The AR(p) and MA(q) orders corresponding to this vertex are used to determine the orders of ARIMA model. The neighbouring models, where p = 0 and q = 2 as well as p = 1 and q = 2, were also chosen. These are highlighted in blue along with the vertex highlighted in yellow. Therefore, the possible candidates of ARIMA from EACF table include ARIMA (0,2,1), ARIMA (0,2,2) and ARIMA (1,2,2).

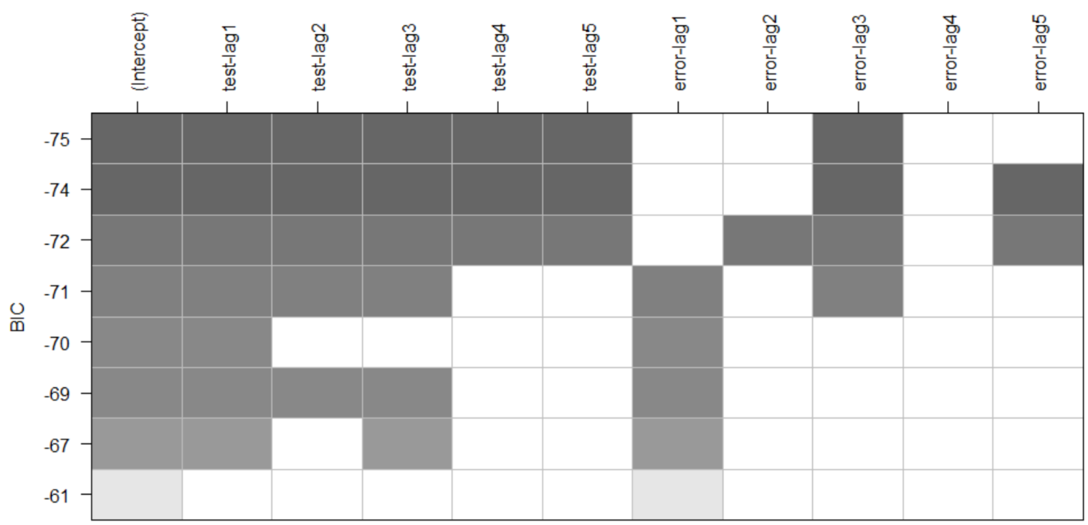

Table 11. BIC table of the transformed series (2010-2019) after second differencing

Table 11 (BIC) represents the most shaded columns at the top row as the best model, while the ones in the second row indicate the second-best model. The shaded columns at the top row correspond to AR (5) and MA (3), while the shaded columns in the second row correspond to AR (5) and MA (5). In addition to this, the most shaded column would indicate the parameters that occur most often in the top models. AR(3) and MA(3) frequently appear in the top models and, therefore, was also included as one other possible candidate model. Thus, ARIMA(3,2,3), ARIMA (5,2,3) and ARIMA (5,2,5) were included in the set of possible candidate models.

In summary, the candidate model inferred from ACF and PACF plot is {ARIMA (0,2,1)}, the set of possible candidate models inferred from EACF table are {ARIMA(0,2,1),ARIMA(0,2,2), ARIMA(1,2,2)} and the set of possible candidate models inferred from BIC table are {ARMA(3,2,3), ARIMA(5,2,3) and ARIMA(5,2,5)}. The whole set of the candidate models summarised in Table 12**Error! Reference source not found.**.

Candidate Models	ACF/PACF Plots	EACF Plot	BIC Table
ARIMA (0,2,1)	√	√	
ARIMA (0,2,2)		√	
ARIMA (1,2,2)		√	
ARIMA (3,2,3)			√
ARIMA (5,2,3)			√
ARIMA (5,2,5)			√

Table 12. Summary of Candidate Models (2010-2019 Series)

2.4.3. Model Fitting and Parameter Estimation

After ensuring the series is stationary and specifying the orders of the autoregressive and moving average elements of ARIMA models in the previous section, the next step is to estimate parameters of the specified tentative candidate models based on maximum likelihood estimates with coefficient significance test. Maximum Likelihood Estimation (MLE) is the best estimator and deemed efficient for a large sample, as is the case for this dataset.

The results of the coefficient test are enumerated below. Full results from R can be found in Appendix B.

- Models retained for further residual analysis:

  ARIMA (0,2,1) – The coefficient of MA (1) returned by MLE is -0.99 with a *p*-value of less than 2.2e-16, which can be concluded as statistically significant.

  ARIMA (0,2,2) – The coefficients of MA (1) and MA (2) returned by MLE are -1.38 (p-value < 2.2e-16) and 0.406 (p-value = 0.0002), respectively. It can thus be inferred that both parameters are statistically significant.

  ARIMA (1,2,2) – Since coefficients of AR (1), MA (1), and MA (2) are all significant.

  ARIMA (5,2,3) – All coefficients are statistically significant (p-values < 0.05).

- Models excluded from the residual analysis:

  ARIMA(3,2,3) – Beside MA (1), all coefficients' p-values are higher than 0.05. Since most coefficients are not significant, this model was eliminated.

  ARIMA (5,2,5) – Except for AR (2), all coefficients' p-values are higher than 0.05. Since most coefficients are not significant, this model was eliminated.

In conclusion, only candidate models ARIMA (0,2,1), ARIMA (0,2,2), ARIMA (1,2,2), ARIMA(5,2,3) were retained for subsequent residual analysis.

2.4.4. Residual Analysis

This section summarises the results of the residual analysis performed for four candidate models identified in the previous section. The same criteria enumerated in section Residual Analysis1.4.4 applies.

ARIMA (0,2,1)

Figure 26 illustrates the plots of the residuals resulting from fitting model ARIMA (0,2,1). The top left plot shows the standardised residuals fluctuates around zero. It also shows some randomness since the residuals appear to spread in a rectangular shape. However, there are a few significant autocorrelations presented in the ACF of standardised residuals, which is confirmed by Ljung-Box Test's p-value of 0.0416 < 0.05. This led us to reject the null hypothesis that residual correlations as whole up equal to 0. These results suggesting that ARIMA (0,2,1) does not have adequate properties of white noise and was eliminated as an appropriate model for the next step.

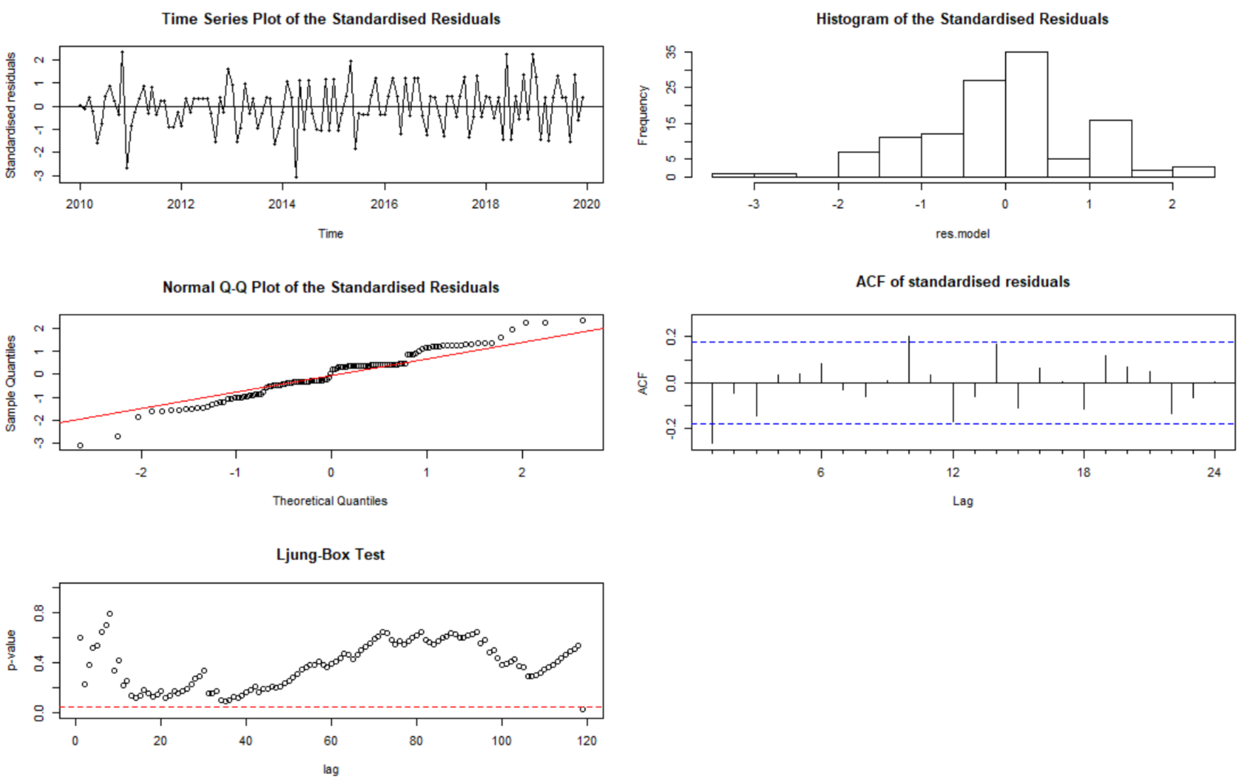

Figure 26. Residual analysis of ARIMA (0,2,1) for 2010-2019

ARIMA (0,2,2)

Similarly, for ARIMA (0,2,2), the residual analysis of this model failed the uncorrelated residuals assumption of white noise. As shown in Figure 27**Error! Reference source not found.**, there is a significant autocorrelation starting at lag ten as visualized in the Ljung-Box test plot, although its p-value of 0.1147 is higher than 0.05. Hence, this was eliminated as an appropriate model.

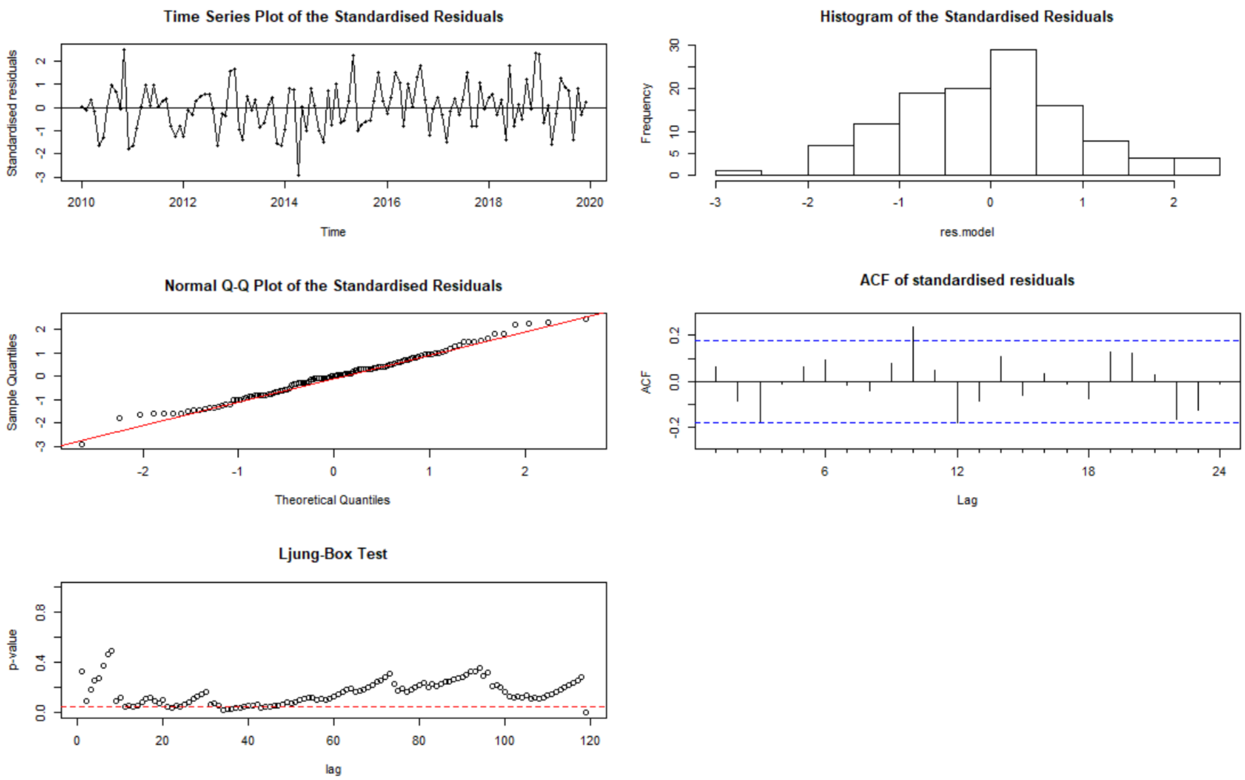

Figure 27. Residual analysis of ARIMA (0,2,2) for 2010-2019

ARIMA (1,2,2) and ARIMA (5,2,3)

In contrast, the residuals of ARIMA (1,2,2) and ARIMA (5,2,3) follow a normal distribution as well as exhibit random and constant variance (). The Ljung-Box tests' p-values are substantially larger than 0.05, suggesting we fail to reject the null hypothesis and conclude that the residuals are uncorrelated. Together, these findings validate the white noise assumptions behind the residuals of these models and thus deemed as adequate.

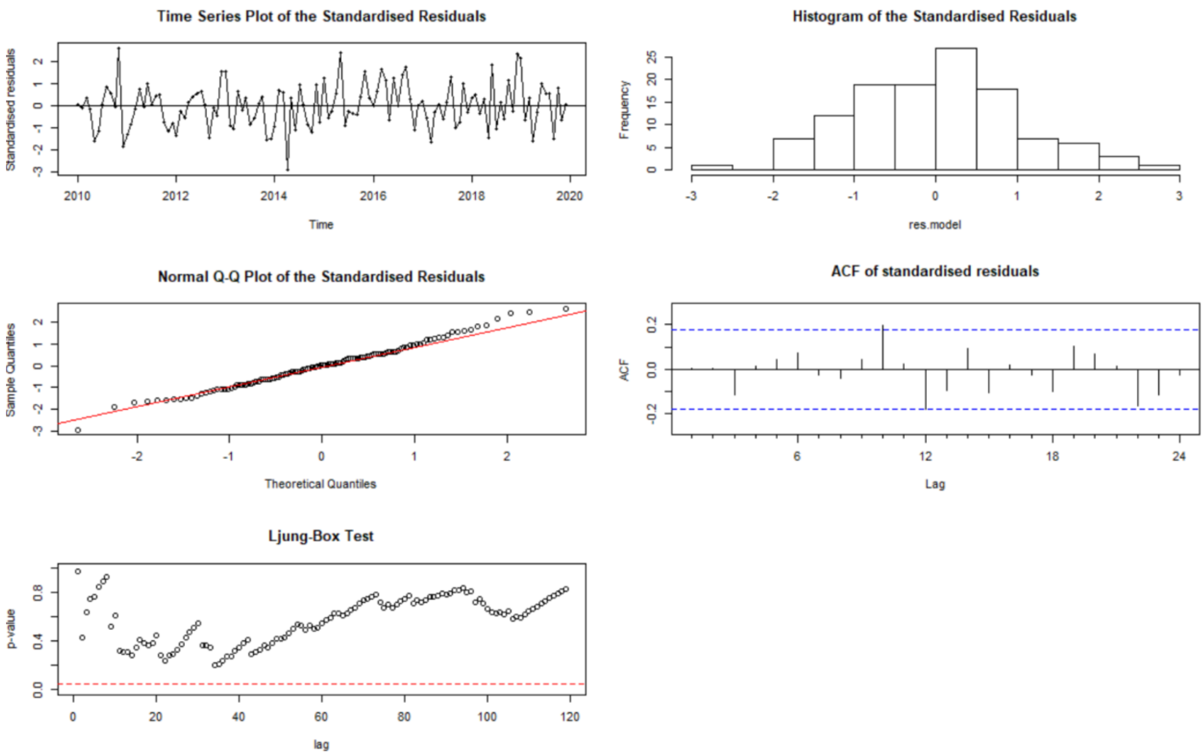

Figure 28. Residual analysis of ARIMA (1,2,2) for 2010-2019

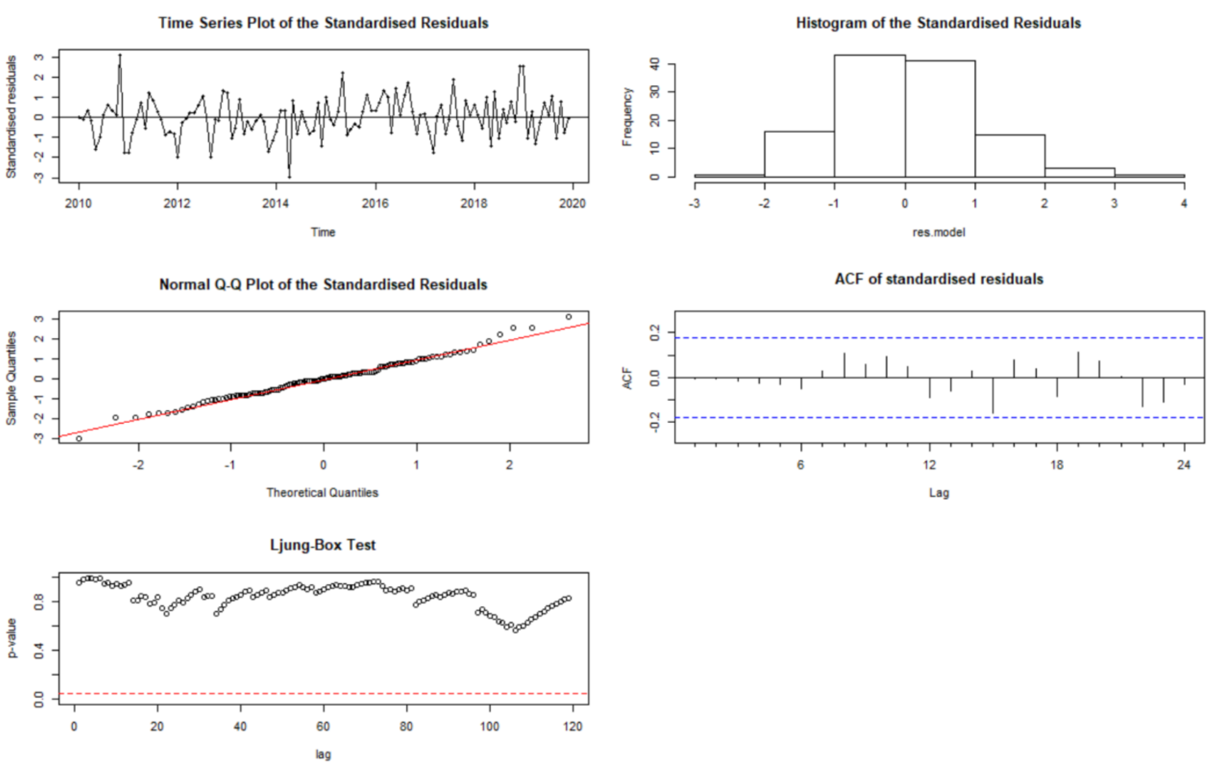

Figure 29. Residual analysis of ARIMA (5,2,3) for 2010-2019

2.4.5. Model Selection

After residual diagnostic check confirmed independence and normality of the models' standardized residuals, the values of Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were computed and ranked to select the best model. The Corrected Akaike Information Criterion (AICc) was also referred to as similar to Chapter 1.

Model	Df	AIC	AICc	BIC
model.122	4	-367.3105	-368.9566	-356.2278
model.523	9	-369.1023	-369.4357	-344.1662

Table 13. Model comparison for the 2010-2019 series

Results in Table 13 highlights that ARIMA (5,2,3) has smaller AIC and AICc values, whereas ARIMA (1,2,2) has lower BIC value. Nevertheless, ARIMA(1,2,2) was chosen as the best model based on the principle of parsimony as it has far fewer parameters to estimate, indicated by fewer degrees of freedom.

2.4.6. Overfitting

After choosing an adequate model, slightly general models or "close by" models that contain the original model was fitted to detect anomalies in terms of goodness of fit. Thus, ARIMA (2,2,2) and ARIMA (1,2,3) are overfitting of ARIMA (1,2,2). The coefficients of AR (2) of ARIMA (2,2,2) and MA (3) of ARIMA (1,2,3) were found to be insignificant. These findings further validate ARIMA (1,2,2) as the best model for this Box-cox transformed series. For the results of the coefficient tests, refer to Appendix B.

2.5. Ten-month Forecast

ARIMA (1,2,2) was used to forecast the monthly U.S. unemployment rate from January of 2020 to October of 2020. Figure 30Figure 30**Error! Reference source not found.** illustrates that the unemployment rates would continue with its downward trend in the first ten months of 2020.

Figure 30. Ten-month forecast for the 2010-2019 time series

As compared to the forecast in chapter 1, the predicted model looks like it is continuing the downtrend, which shows that removing the intervention points for the modeling process has improved the ten-month forecast. The prediction values in percentage with 95% confidence intervals are displayed in Table 14Table 14.. The actual published numbers for January and February are within the 95% confidence intervals bounds, indicating the accuracy of our model.

However, on March 13, President Trump announced a travel ban, and lockdown was imposed shortly after. The initial impacts of his policies in response to the pandemic are reflected immediately in the published unemployment rates in March and April 2020, as it jumps to 4.4% and 14.7%, respectively.

Year	Forecast	Low 95%	High 95%	Actual Data	Difference
January 2020	3.478	3.290	3.673	3.6	0.12
February 2020	3.454	3.230	3.687	3.5	0.05
March 2020	3.430	3.183	3.687	4.4	0.97
April 2020	3.405	3.137	3.685	14.7	11.29
May 2020	3.380	3.092	3.682		
June 2020	3.356	3.047	3.681		
July 2020	3.331	3.001	3.680		
August 2020	3.307	2.955	3.680		
September 2020	3.283	2.909	3.682		
October 2020	3.258	2.862	3.683		

Table 14. Ten-month forecast for the 2010-2019 time series and actual 2020 U.S. unemployment rate figures

Discussion

This analysis has modeled and conducted a ten-month forecast on the unemployment rates in the U.S. twice. The first of which was from January 2000 to December 2019, and the second was from January 2010 to December 2019 (post the Great Recession, pre-COVID-19 pandemic).

The deterministic models were fitted first but were found to be invalid due to the residuals being correlated and non-normal. Hence, stochastic trend models were fitted in both parts of the analysis. In the first instance, ARIMA(1,1,5) was chosen as the best model for the ten-month forecast. However, two economic events, the dot-com bubble and 2008 Global Financial Crisis, are included in the data set and these events seem to have had impacted the ten-month forecast. The forecasted unemployment rates defied the general downtrend from 2010 to 2019.

To overcome the effects of these intervention points, the second part of the analysis focused on modeling and forecasting the data. Since the latest actual unemployment rate in April 2020 is the highest thus far and even surpassed that of the Great Recession's, predicting the time-series based on the early years before this is irrelevant and no longer comparable. Henceforth, Chapter 2 focused mainly on January 2010 to December 2019, after the peak unemployment rate from the 2008 Global Financial Crisis passed.

The ten-month forecast, based on ARIMA (1,2,2), followed the general downtrend of the time series data indicating that the second model is better than the first model ARIMA(1,1,5). In both instances, model diagnostics were performed to ensure that the residuals were independent and normally distributed.

The forecasted unemployment rates were then compared to the published unemployment rates in 2020 to illustrate the accuracy of the model. For the first two months, the actual values were within the 95% confidence interval bounds of the forecast, further confirming the efficiency and accuracy of ARIMA(1,2,2).

As of March 2020, the published unemployment rates in the U.S. have started to go outside of the 95% confidence interval bounds, as we see the initial impacts of the COVID-19 pandemic where President Trump announced a travel ban. The unemployment rates in April have since increased dramatically, likely due to the lockdown issued by the President.

Based on this analysis, it is safe to conclude that the COVID-19 has had a massive impact on the unemployment rates in the U.S. as unemployment rates in April 2020 has surpassed the unemployment rates at the height of the GFC.

Conclusion and Recommendations

Before the coronavirus pandemic, unemployment in the U.S. has been declining steadily. The forecast based on 2010 to 2019 historical figures have shown a highly optimistic scenario following this trajectory. The first two months of 2020 even provided an accurate reading in which the predicted values were revealed to be close to the latest published data.

However, the trend was reversed upon the abrupt implementation of the new policies, particularly the imposed travel ban and lockdown in the majority of the states. As of the moment, the U.S. is not past the worst of the pandemic as the number of infections continues to rise. This means that the true effect on unemployment is yet to be determined.

In this regard, the next step is to improve the initial model, as more data is collected. Additionally, it is imperative to perform an intervention analysis to assess the full impact of the pandemic. The trend in the unemployment rate displays a variety of patterns, which can be split into components. A time series decomposition approach, which will analyze the trend-cycle vs. seasonal components, may be more appropriate in this type of situation. Understanding the underlying patterns in more detail can help increase the model accuracy.

Appendices

Appendix A: R Codes

```
# INSTALLATION AND LOADING OF THE TIME SERIES PACKAGES
ts_packages <- c('TSA', 'fUnitRoots', 'forecast', 'lmtest', 'FitAR', 'tseries')
for (pkg in ts_packages) {
  if (pkg %in% rownames(installed.packages()) == FALSE)
  {install.packages(pkg)}
  if (pkg %in% rownames(.packages()) == FALSE)
  {library(pkg, character.only = TRUE)}
}

# FUNCTION-SETTING

# Descriptive Analysis Function
descriptive_analysis <- function(ts, ylabel="ylabel", xlabel="xlabel", title="Name of
the Time Series") {
  # Time Series Plot
  plot(ts, ylab=ylabel, xlab=xlabel, type="o", main=paste("Time Series Plot of",
title))
  # Correlation Analysis
  plot(y=ts,x=zlag(ts), ylab=ylabel, xlab = "Previous Month Change", main =
paste("Scatter Plot of ", title))
  y = ts
  x = zlag(ts)
  index = 2:length(x)
  cor(y[index],x[index])
}

# Normality Test function
normality_test <- function(x, title = "Series Name") {
  par(mfrow=c(1,2))
  qqnorm(x, main="")
  qqline(x, col = "blue")
  data_sd <- sd (x)
  data_mean <- mean(x)
  data_median <- median(x)
  hist (x, main = "", probability = TRUE, xlab = "Time Series Data")
  xm <- seq(min(x),max(x),length = 100)
  ym <- dnorm (x= xm, mean = data_mean, sd = data_sd)
  lines(x = xm, y= ym, col = "blue", lwd = 1.5)
  abline (v = data_mean, col = "red", lwd = 3, lty = 2)
  abline (v = data_median, col = "dark green", lwd = 3, lty = 2)
  legend("topright", legend = c("Mean","Median"), col = c("red", "dark green"), pch =
15, bty = "n")
  par(mfrow=c(1,1))
  mtext(paste("Normal Q-Q Plot and Histogram of", title), side = 3, line = -2, outer
= TRUE)
  print(shapiro.test(x))
}

# Deterministic trend function
```

```
deterministic_model <- function(x, model = c("Linear", "Quadratic"), title="Series
Name", ylabel="Label", xlabel="Label") {
  t = time(x)
  t2 = t^2

  if (model == "Linear") {
    model_x = lm(x~t)
    summary_model <- summary(model_x)
    plot_model <- { plot(x, type='o', ylab=ylabel, xlab=xlabel, main = paste(model,
"Trend Model of", title))
      abline(model_x, col ="red")
      legend("topright", lty =1, bty = "n", col = c("black", "red"), c("Time Series
Plot", "Model fitted value"))}
    res_analysis <- {
      # Standardized Residuals vs. Time
      res.model_x = rstudent(model_x)
      plot(y = res.model_x, x = as.vector(time(x)),xlab=xlabel, ylab='Standardized
Residuals', type='p', main="Scatter Plot of the Standarized Residuals")
      abline(h=0)
      # Standardised residuals vs. fitted trend values
      scatter.smooth(y=res.model_x, x=fitted(model_x), xlab='Fitted Trend Values',
ylab='Standardized Residuals', main = "Plot of the Standardised Residuals vs. Fitted
Trend Values")
      abline(h = 0, lty = 2)
      Acf(res.model_x, main="ACF of the Standardized Residuals")
      # Test constant variances
      print(ncvTest(model_x))
      # Test residual independence
      print(runs(res.model_x))
      # Normality Test
      normality_test(res.model_x, title) }
  } else if (model == "Quadratic") {
    model_x = lm(x~ t + t2)
    summary_model <- summary(model_x)
    plot_model <- { plot(ts(fitted(model_x)), ylim =
c(min(c(fitted(model_x),as.vector(x))), max(c(fitted(model_x),as.vector(x)))),
col="red", ylab=ylabel, xlab=xlabel, main = paste(model, " Trend Model of", title))
      lines(as.vector(x), type="o")
      legend("bottomleft", lty =1, bty = "n", text.width = 12, col = c("black",
"red"), c("Time Series plot", "Model fitted values"))
    res_analysis <- {
      # Standardized Residuals vs. Time
      res.model_x = rstudent(model_x)
      plot(y = res.model_x, x = as.vector(time(x)),xlab =xlabel, ylab='Standardized
Residuals', type='p', main="Scatter Plot of the Standarized Residuals")
      abline(h=0)
      # Standardised residuals vs. fitted trend values
      scatter.smooth(y=res.model_x,x=fitted(model_x), xlab='Fitted Trend Values',
ylab='Standardized Residuals', main = "Plot of the Standardized Residuals vs. Fitted
Trend Values")
      abline(h = 0, lty = 2)
      Acf(res.model_x, main="ACF of the Standardized Residuals")
      # Test constant variances
      print(ncvTest(model_x))
```

```
      # Test residual independence
      print(runs(res.model_x))
      # Normality Test
      normality_test(res.model_x, title) }
    }
  } else { stop("Choose between Linear or Quadratic only")
  }
  return(summary_model)
  return(plot_model)
  return(res_analysis)
  par(mfrow=c(1,1))
}

# ACF and PACF plot function
correlograms <- function(x, title = "Series Name") {
  par(mfrow=c(1,3))
  Acf(x, main = "")
  acf(x, ci.type='ma', main= "")
  Pacf(x, main = "")
  mtext(paste("ACF/PACF Plot of", title), side = 3, line = -2, outer = TRUE)
  par(mfrow=c(1,1))
}

# Differencing function
differencing <- function(x){
  # First Differencing
  data <- diff(x, diff = 1)
  order <-  ar(diff(data))$order
  p <- adfTest(data, lags = order,  title = NULL,description = NULL)@test$p.value

  # Second Differencing
  if (p >= 0.05) {
    data_2df <- diff(x, diff = 2)
    order <- ar(diff(data_2df))$order
    p2 <- adfTest(data_2df, lags = order,  title = NULL,description =
NULL)@test$p.value
    p2
    if (p2 >= 0.05) {print("Model may result to overdifferencing after Second
Differencing. Check again.")
    }else {
      print(shapiro.test(data_2df))
      print("Achieved stationarity after Second Differencing")
      print(paste("p-value from the adfTest", p2))
    }
  }else if (p < 0.05) {
    print(shapiro.test(data))
    print("Achieved stationarity after First Differencing")
    print(paste("p-value from the adfTest", p))
  }else {stop("Check error")
  }
}

# Sort score function
sort.score <- function(x, score = c("bic", "aic")){
```

```
  if (score == "aic"){
    x[with(x, order(AIC)),]
  } else if (score == "bic") {
    x[with(x, order(BIC)),]
  } else {
    warning('score = "x" only accepts valid arguments ("aic","bic")')
  }
}

# Residual analysis function
residual.analysis <- function(model, std = TRUE){
  if(std == TRUE){
    res.model = rstandard(model)
  }else{
    res.model = residuals(model)
  }
  par(mfrow=c(3,2))
  plot(res.model, type = 'o', ylab = 'Standardised residuals', main = "Time Series
Plot of the Standardised Residuals")
  abline(h = 0)
  hist(res.model, main = "Histogram of the Standardised Residuals")
  qqnorm(res.model, main = "Normal Q-Q Plot of the Standardised Residuals")
  qqline(res.model, col = 2)
  Acf(res.model, main = "ACF of standardised residuals")
  print(shapiro.test(res.model))
  print(runs(res.model))
  k=0
  LBQPlot(res.model, lag.max = length(model$residuals)-1, StartLag = k+1, k = 0,
SquaredQ = FALSE)
  par(mfrow=c(1,1))
  boxtest <- Box.test(res.model, lag = 12, type = "Ljung-Box", fitdf = 0) # Lags
limited to 12 as per the highest df in the sort score
  print(boxtest)
}

# AICc function
AICc = function(model){
  n = model$nobs
  k = length(model$coef)
  aicc = model$aic + 2*(k+1)*(k+2)/(n-k-2)
  return(aicc)
}

# READ DATA AND CONVERT TO TIME SERIES DATA
library(readr)
library(tidyr)
library(car)
USunemployment <- read_csv("USUnemployment.csv")
head(USunemployment)
unemp_filter <- subset(USunemployment, Year>=2000)
USunem <- gather(unemp_filter, "month", "unemployment rate", na.rm = TRUE, 2:13)
head(USunem)
USunem_sorted <- USunem[order(USunem$Year),]
head(USunem_sorted)
```

```
unem <- ts(USunem_sorted$`unemployment rate`, start=c(2000,1), end=c(2019,12),
frequency = 12)

# CHAPTER 1: MONTHLY U.S. UNEMPLOYMENT RATE FROM 2000-2019

# 1.1. DESCRIPTIVE ANALYSIS
descriptive_analysis(unem, "Unemployment Rate (%)", "Year", "the Monthly U.S.
Unemployment Rate for 2000-2019")
Acf(unem, main = "ACF of the Monthly U.S. Unemployment Rate from 2000 to 2019")

# 1.2. LINEAR TREND MODEL
deterministic_model(unem, model="Linear", title="the Monthly U.S. Unemployment Rate
(2000-2019)", ylabel="Unemployment Rate (%)", xlabel="Year")

# 1.3. QUADRATIC TREND MODEL
deterministic_model(unem, model="Quadratic", title="the Monthly U.S. Unemployment
Rate (2000-2019)", ylabel="Unemployment Rate (%)", xlabel="Year")

# 1.4. STOCHASTIC TREND MODEL

# 1.4.1 Test of Normality and Stationarity
normality_test(unem, "the U.S. Unemployment Rate")
correlograms(unem, "the U.S. Unemployment Rate")
adf.test(unem)
BoxCox(unem, interval = c(-1, 1)) # Data Transformation
lambda = -0.709
unem.BC = (unem^lambda-1)/lambda
normality_test(unem.BC,  "the Time Series (2000-2019) after Box-Cox Transformation")
order = ar(diff(unem.BC))$order
adfTest(unem.BC, lags = order, type = 'nc',  title = NULL, description = NULL)
plot(unem.BC, ylab='Unemployment Rate (%)',xlab='Year',type='o', main = "Time Series
Plot of the Monthly U.S. Unemployment Rate after Box-Cox Transformation")

# 1.4.2. Model Specification
differencing(unem.BC)
diff.unem.BC = diff(unem.BC, diff = 1) # Set data to first differencing
plot(diff.unem.BC ,ylab='Unemployment Rate (%)',xlab='Year',type='o', main = "Time
Series Plot of the First Differenced Series (2000-2019)")
abline(h=0, col = 'red', lty = 2)
correlograms(diff.unem.BC, "the First Differenced Series (2000-2019)")
eacf(diff.unem.BC, ar.max = 10, ma.max= 10)
bic = armasubsets(y=diff.unem.BC,nar=10,nma=10,y.name='test',ar.method='ols')
plot(bic)

# 1.4.3. Model Fitting and Parameter Estimation
model.115 <-  arima(unem.BC, order=c(1,1,5), method='ML')
coeftest(model.115)
model.516 <-  arima(unem.BC, order=c(5,1,6), method='ML')
coeftest(model.516)
model.315 <- arima(unem.BC, order=c(3,1,5), method='ML')
coeftest(model.315)
model.316 <- arima(unem.BC, order=c(3,1,6), method='ML')
coeftest(model.316)
```

```r
model.415 <- arima(unem.BC, order=c(4,1,5), method='ML')
coeftest(model.415)
model.710 <- arima(unem.BC, order=c(7,1,0), method='ML')
coeftest(model.710)
model.711 <- arima(unem.BC, order=c(7,1,1), method='ML')
coeftest(model.711)

# 1.4.4. Residual Analysis
residual.analysis(model.115)
residual.analysis(model.516)
residual.analysis(model.315)
residual.analysis(model.316)
residual.analysis(model.415)
residual.analysis(model.710)
residual.analysis(model.711)

# 1.4.5. Model Selection (of the remaining candidate models)
sort.score(AIC(model.115, model.415), score = "aic")
sort.score(BIC(model.115, model.415), score = "bic")
AICc(model.115)
AICc(model.415)

# 1.4.6. Overfitting (of the best model)
model.215 = arima(unem.BC, order=c(2,1,5), method='ML')
coeftest(model.215)
model.116 = arima(unem.BC, order=c(1,1,6), method='ML')
coeftest(model.116)

# 1.5. FORECAST
fit_115 = Arima(unem,c(1,1,5), lambda = -0.709)
predict_115 = forecast(fit_115, h=10)
predict_115
plot(predict_115,  type = 'o', colour = "blue", xlab = "Year", ylab = "Unemployment
Rate (%)", main = "Time Series Plot of the U.S. Unemployment Rate with a Ten-month
Forecast")

# CHAPTER 2: MONTHLY U.S. UNEMPLOYMENT RATE FROM 2010-2019

# DATA PREPROCESSING
unemp2 <- subset(USunemployment, Year>=2010)
unem2_gather <- gather(unemp2, "month", "unemployment rate", na.rm = TRUE, 2:13)
unemp2_sorted <- unem2_gather[order(unem2_gather$Year),]
head(unemp2_sorted)
unem2 <- ts(unemp2_sorted$`unemployment rate`, start=c(2010,1), end=c(2019,12),
frequency = 12)

# 2.1. DESCRIPTIVE ANALYSIS
descriptive_analysis(unem2, "Unemployment Rate (%)", "Year", "the Monthly U.S.
Unemployment Rate for 2010-2019")
Acf(unem2, main = "ACF of the Monthly U.S. Unemployment Rate from 2010 to 2019")
mean(unem2)
sd(unem2)
```

```
# 2.2. LINEAR TREND MODEL
deterministic_model(unem2, model="Linear", title="the Monthly U.S. Unemployment Rate
(2010-2019)", ylabel="Unemployment Rate (%)", xlabel="Year")

# 2.3. QUADRATIC TREND MODEL
deterministic_model(unem2, model="Quadratic", title="the Monthly U.S. Unemployment
Rate (2010-2019)", ylabel="Unemployment Rate (%)", xlabel="Year")

# 2.4. STOCHASTIC TREND MODEL

# 2.4.1 Test of Normality and Stationarity
normality_test(unem2, "the U.S. Unemployment Rate")
correlograms(unem2, "the U.S. Unemployment Rate")
adf.test(unem2)
unem2.transform = BoxCox.ar(unem2) # Data Transformation
unem2.transform$ci # Lambda from 0.1 to 0.8
lambda = 0.45
unem2.BC = (unem2^lambda-1)/lambda
normality_test(unem2.BC, title = "the Time Series after Box-Cox Transformation")
plot(unem2.BC, ylab='Unemployment Rate (%)',xlab='Year',type='o', main = "Time Series
Plot of the Monthly U.S. Unemployment Rate after Box-Cox Transformation")

# 2.4.2. Model Specification
differencing(unem2.BC)
diff2.unem2.BC = diff(unem2.BC,differences = 2) # Set data at Second differencing
plot(diff2.unem2.BC ,ylab='Unemployment Rate (%)',xlab='Year',type='o', main = "Time
series plot of the Second Differenced Series (2010-2019)")
abline(h=0, col = 'red', lty = 2)
correlograms(diff2.unem2.BC, "the Second Differenced Series (2010-2019)")
eacf(diff2.unem2.BC,ar.max = 10, ma.max= 10)
bic_ms2 = armasubsets(y=diff2.unem2.BC,nar=5,nma=5,y.name='test',ar.method='ols')
plot(bic_ms2)

# 2.4.3. Model Fitting and Parameter Estimation
model2.021 <-  arima(unem2.BC, order=c(0,2,1), method='ML')
coeftest(model2.021)
model2.022 <-  arima(unem2.BC, order=c(0,2,2), method='ML')
coeftest(model2.022)
model2.122 <- arima(unem2.BC, order=c(1,2,2), method='ML')
coeftest(model2.122)
model2.323 <- arima(unem2.BC, order=c(3,2,3), method='ML')
coeftest(model2.323)
model2.523 <- arima(unem2.BC, order=c(5,2,3), method='ML')
coeftest(model2.523)
model2.525 <- arima(unem2.BC, order=c(5,2,5), method='ML')
coeftest(model2.525)

# 2.4.4. Residual Analysis
residual.analysis(model2.021)
residual.analysis(model2.022)
residual.analysis(model2.122)
residual.analysis(model2.323)
residual.analysis(model2.523)
residual.analysis(model2.525)
```

```
# 2.4.5. Model Selection (of the remaining candidate models)
sort.score(AIC(model2.122, model2.323, model2.523, model2.525), score = "aic")
sort.score(BIC(model2.122, model2.323, model2.523, model2.525), score = "bic"
AICc(model2.122)
AICc(model2.523)
AICc(model2.323)
AICc(model2.525)

# 2.4.6. Overfitting (of the best model)
model2.222 = arima(unem2.BC, order=c(2,2,2), method='ML')
coeftest(model2.222)
model2.123 = arima(unem2.BC, order=c(1,2,3), method='ML')
coeftest(model2.123)

# 2.5. FORECAST
fit_122 = Arima(unem2,c(1,2,2), lambda = 0.45)
predict_122 = forecast(fit_122, h=10)
predict_122
plot(predict_122, type = 'o', colour = "blue", xlab = "Year", ylab = "Unemployment
Rate (%)", main = "Time Series Plot of the U.S. Unemployment Rate with a Ten-month
Forecast")
```

Appendix B: Results of the Coefficient Tests

```
# CHAPTER 1: MONTHLY U.S. UNEMPLOYMENT RATE FROM 2000-2019
model.115 <-  arima(unem.BC, order=c(1,1,5), method='ML')
coeftest(model.115)

##
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar1  0.658543   0.061255  10.7509 < 2.2e-16 ***
## ma1 -0.915808   0.067418 -13.5841 < 2.2e-16 ***
## ma2  0.351743   0.087590   4.0158 5.925e-05 ***
## ma3 -0.010385   0.078013  -0.1331    0.8941
## ma4 -0.120984   0.087331  -1.3853    0.1659
## ma5  0.470767   0.079869   5.8943 3.763e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model.516 <-  arima(unem.BC, order=c(5,1,6), method='ML')
coeftest(model.516)

## Warning in sqrt(diag(se)): NaNs produced

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.401755   0.181989  2.2076  0.027274 *
```

```
## ar2  0.157263         NA       NA         NA
## ar3 -0.224556         NA       NA         NA
## ar4  0.376222   0.124651  3.0182  0.002543 **
## ar5 -0.292877   0.120376 -2.4330  0.014973 *
## ma1 -0.633124   0.198348 -3.1920  0.001413 **
## ma2  0.103100   0.019345  5.3295 9.847e-08 ***
## ma3  0.255758         NA       NA         NA
## ma4 -0.479538         NA       NA         NA
## ma5  0.726024   0.152226  4.7694 1.848e-06 ***
## ma6  0.110340   0.164260  0.6717  0.501750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.315 <- arima(unem.BC, order=c(3,1,5), method='ML')
coeftest(model.315)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ar1  0.8864892  0.1923432  4.6089 4.048e-06 ***
## ar2 -0.2234859  0.2984218 -0.7489   0.45392
## ar3 -0.0072147  0.1844463 -0.0391   0.96880
## ma1 -1.1106511  0.1812544 -6.1276 8.923e-10 ***
## ma2  0.5900124  0.3281587  1.7979   0.07219 .
## ma3 -0.1191044  0.2904066 -0.4101   0.68171
## ma4 -0.0673627  0.1411165 -0.4774   0.63311
## ma5  0.4092670  0.0996060  4.1089 3.976e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model.316 <- arima(unem.BC, order=c(3,1,6), method='ML')
coeftest(model.316)
```

```
## Warning in sqrt(diag(se)): NaNs produced
```

```
##
## z test of coefficients:
##
##     Estimate Std. Error z value Pr(>|z|)
## ar1  0.20830         NA      NA       NA
## ar2  0.37024         NA      NA       NA
## ar3 -0.15623         NA      NA       NA
## ma1 -0.43220         NA      NA       NA
## ma2 -0.15627         NA      NA       NA
## ma3  0.27650         NA      NA       NA
## ma4 -0.15067         NA      NA       NA
## ma5  0.36442         NA      NA       NA
## ma6  0.28072         NA      NA       NA
```

```
model.415 <- arima(unem.BC, order=c(4,1,5), method='ML')
coeftest(model.415)
```

```
##
## z test of coefficients:
```

```
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  1.29663    0.23600  5.4941 3.927e-08 ***
## ar2 -1.03018    0.37941 -2.7152 0.0066227 **
## ar3  0.73692    0.32082  2.2970 0.0216194 *
## ar4 -0.31307    0.14600 -2.1443 0.0320099 *
## ma1 -1.52840    0.24883 -6.1424 8.129e-10 ***
## ma2  1.50109    0.45405  3.3060 0.0009465 ***
## ma3 -1.14127    0.48168 -2.3693 0.0178201 *
## ma4  0.59821    0.34753  1.7213 0.0851960 .
## ma5  0.14806    0.17218  0.8599 0.3898430
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model.710 <- arima(unem.BC, order=c(7,1,0), method='ML')
coeftest(model.710)

##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.174043   0.064126 -2.7141  0.006646 **
## ar2  0.131530   0.063660  2.0661  0.038817 *
## ar3  0.109926   0.062507  1.7586  0.078645 .
## ar4 -0.028578   0.064686 -0.4418  0.658633
## ar5  0.243673   0.064219  3.7944  0.000148 ***
## ar6  0.254723   0.065232  3.9049 9.426e-05 ***
## ar7  0.131616   0.067360  1.9539  0.050710 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model.711 <- arima(unem.BC, order=c(7,1,1), method='ML')
coeftest(model.711)

##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ar1 -0.126911   0.390732 -0.3248 0.7453299
## ar2  0.138429   0.085377  1.6214 0.1049339
## ar3  0.102107   0.089229  1.1443 0.2524902
## ar4 -0.033808   0.077292 -0.4374 0.6618135
## ar5  0.244337   0.064466  3.7902 0.0001505 ***
## ar6  0.242061   0.122517  1.9757 0.0481850 *
## ar7  0.120387   0.116567  1.0328 0.3017125
## ma1 -0.047975   0.391495 -0.1225 0.9024689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model.215 = arima(unem.BC, order=c(2,1,5), method='ML')
coeftest(model.215)

##
## z test of coefficients:
##
```

```
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.891331   0.152155  5.8581 4.683e-09 ***
## ar2 -0.234142   0.129665 -1.8057 0.0709582 .
## ma1 -1.115134   0.144742 -7.7043 1.316e-14 ***
## ma2  0.601037   0.171428  3.5061 0.0004548 ***
## ma3 -0.129300   0.122346 -1.0568 0.2905841
## ma4 -0.063872   0.108391 -0.5893 0.5556749
## ma5  0.408689   0.099162  4.1214 3.765e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model.116 = arima(unem.BC, order=c(1,1,6), method='ML')
coeftest(model.116)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ar1  0.539778   0.105221  5.1300 2.898e-07 ***
## ma1 -0.764280   0.116077 -6.5843 4.571e-11 ***
## ma2  0.286595   0.089832  3.1904  0.001421 **
## ma3 -0.012873   0.081296 -0.1584  0.874179
## ma4 -0.077825   0.087318 -0.8913  0.372781
## ma5  0.382144   0.085366  4.4765 7.587e-06 ***
## ma6  0.161927   0.090069  1.7978  0.072208 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# CHAPTER 2: MONTHLY U.S. UNEMPLOYMENT RATE FROM 2010-2019
model2.021 <-  arima(unem2.BC, order=c(0,2,1), method='ML')
coeftest(model2.021)

##
## z test of coefficients:
##
##      Estimate Std. Error z value  Pr(>|z|)
## ma1 -0.999997   0.026209 -38.155 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2.022 <-  arima(unem2.BC, order=c(0,2,2), method='ML')
coeftest(model2.022)

##
## z test of coefficients:
##
##     Estimate Std. Error  z value  Pr(>|z|)
## ma1 -1.39220    0.10805 -12.8851 < 2.2e-16 ***
## ma2  0.40554    0.11861   3.4191 0.0006282 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2.122 <- arima(unem2.BC, order=c(1,2,2), method='ML')
```

```
## Warning in log(s2): NaNs produced

coeftest(model2.122)

##
## z test of coefficients:
##
##      Estimate Std. Error  z value  Pr(>|z|)
## ar1  0.31688    0.15026   2.1089   0.03495 *
## ma1 -1.65912    0.10764 -15.4138 < 2.2e-16 ***
## ma2  0.68445    0.10878   6.2922  3.13e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2.323 <- arima(unem2.BC, order=c(3,2,3), method='ML')
coeftest(model2.323)

##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.22416    0.49235  0.4553 0.648903
## ar2 -0.17658    0.23950 -0.7373 0.460930
## ar3 -0.19817    0.12110 -1.6365 0.101745
## ma1 -1.56306    0.49814 -3.1378 0.001702 **
## ma2  0.73341    0.80026  0.9165 0.359425
## ma3 -0.13382    0.35775 -0.3740 0.708369
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2.523 <- arima(unem2.BC, order=c(5,2,3), method='ML')
coeftest(model2.523)

##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1 -0.388593   0.098340  -3.9515 7.765e-05 ***
## ar2 -1.135256   0.097350 -11.6616 < 2.2e-16 ***
## ar3 -0.601740   0.137625  -4.3723 1.229e-05 ***
## ar4 -0.327482   0.097071  -3.3736 0.0007418 ***
## ar5 -0.246016   0.098347  -2.5015 0.0123667 *
## ma1 -0.944545   0.057108 -16.5397 < 2.2e-16 ***
## ma2  0.992311   0.070876  14.0006 < 2.2e-16 ***
## ma3 -0.952612   0.076680 -12.4232 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

model2.525 <- arima(unem2.BC, order=c(5,2,5), method='ML')
coeftest(model2.525)

##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
```

```
## ar1 -0.40809    0.49048 -0.8320 0.405393
## ar2 -0.99155    0.34325 -2.8887 0.003868 **
## ar3 -0.58929    0.46532 -1.2664 0.205358
## ar4 -0.20753    0.28901 -0.7181 0.472703
## ar5 -0.21229    0.13568 -1.5647 0.117655
## ma1 -0.93051    0.49508 -1.8795 0.060174 .
## ma2  0.81606    0.81832  0.9972 0.318648
## ma3 -0.76762    0.84328 -0.9103 0.362679
## ma4 -0.17669    0.82647 -0.2138 0.830707
## ma5  0.17046    0.44142  0.3862 0.699382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model2.222 = arima(unem2.BC, order=c(2,2,2), method='ML')
```

```
## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced

## Warning in log(s2): NaNs produced
```

```
coeftest(model2.222)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1  0.296068   0.162873   1.8178    0.0691 .
## ar2 -0.047571   0.111326  -0.4273    0.6691
## ma1 -1.628738   0.135909 -11.9840 < 2.2e-16 ***
## ma2  0.656398   0.135109   4.8583 1.184e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model2.123 = arima(unem2.BC, order=c(1,2,3), method='ML')
coeftest(model2.123)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error z value  Pr(>|z|)
## ar1  0.236430   0.320806  0.7370    0.4611
## ma1 -1.573307   0.314566 -5.0015 5.688e-07 ***
## ma2  0.551902   0.463402  1.1910    0.2337
## ma3  0.049744   0.166083  0.2995    0.7645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

References

Hayes, A., 2019. *What Ever Happened To The Dotcom Bubble?*. [online] Investopedia. Available at: <https://www.investopedia.com/terms/d/dotcom-bubble.asp> [Accessed 2 June 2020].

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 9 May 2020.

Tunguz, B., 2020. *U.S. Monthly Unemployment Rate 1948 - Present*. [online] Kaggle.com. Available at: <https://www.kaggle.com/tunguz/us-monthly-unemployment-rate-1948-present> [Accessed 20 May 2020].

Uchitelle, L., 2009. *U.S. Lost 2.6 Million Jobs In 2008*. [online] Nytimes.com. Available at: <https://www.nytimes.com/2009/01/09/business/worldbusiness/09iht-jobs.4.19232394.html#:~:text=NEW%20YORK%20%E2%80%94%20U.S.%20employers%20shed,promises%20more%20significant%20losses%20ahead.&text=More%20than%2011%20million%20Americans%20are%20now%20unemployed.> [Accessed 3 June 2020].