Joel Cabrera Professor Robin

Regression and Time Series Analysis (16:954:596) December 5, 2020

NJ Unemployment and Labor Rates: Estimation & Forecasting in Wake of COVID-19

I. Introduction

The recent COVID-19 pandemic has impacted the state of the U.S economy such that all U.S. citizens' livelihoods have been under the constant threat of poverty and joblessness. Indeed, back in April 2020, around 22 million U.S. citizens filed for unemployment **[3]**, indicating that overall U.S. unemployment rates are at an all-time high – not since the Great Depression **[3]**. While some may say that the ongoing pandemic will eventually pass and the U.S. will recover from its economic predicament, U.S. citizens, on average, think otherwise. 48% of Americans feel depressed or hopeless about whether the U.S. economy and their lives can be improved in the aftermath of the pandemic **[7]**. As a data analyst working for the U.S. Bureau of Labor Statistics (BLS), I have been tasked with analyzing the New Jersey (NJ) unemployment (UR) and labor force participation rate (LFPR) rates **[5]**. My tasks are to examine the current state of UR and LFPR rates, determine what their average monthly values are, and provide 12-month forecasts for them.

Such tasks are assigned to me for three reasons. The first is that information generated by my statistical analyses will be used to inform NJ policymakers and businesses of what best courses of action to take when focusing on improving NJ unemployment and, by extension, the NJ economy. The second is that any discoveries made during my analyses can be used for future reference. That way, NJ officials will be prepared to tackle another possible pandemic, avoiding the disastrous world-wide economic effects we are experiencing right now. The third reason is that any such generated information generated can hopefully inspirit U.S. citizens about their future livelihoods in wake of the ongoing pandemic.

II. Data & Description

The datasets that will be used for my data analysis are the NJUR **[6]** and LBSSA34 **[2]** datasets from the Federal Reserve Economic Data (FRED) database, housed and publicly made available by the Federal Reserve Bank of St. Louis. Both datasets are formatted as Excel files. They each span observed time series of monthly NJ UR and monthly NJ LFPR rates (both measured as percentages) from January 1976 to October 2020. Given these facts, these data have frequencies of 12 (monthly data), dimensions of 538 x 2 (sample size T = 538), and 2 variables: monthly dates and NJ UR/NJ LFPR rates. For the sake of simplicity, these datasets will hereafter be referred to as UR and LFPR, respectively. UR and LFPR have means of 6.3 and 65.38, respectively. Both datasets have no missing entries. Finally, both UR and LFPR have already been seasonally adjusted, so seasonality is accounted for. Table 1 provides additional summary statistics for the data. All analyses of these data were conducted using the 3.6.2 version of R.

III. Quantitative Methodologies

Given the properties of the UR and LFPR datasets, my tasks, and my current knowledge of time series methods, I will employ two kinds of statistical models and appropriate statistical tests to satisfy all of the above. For the sake of organization and clarity, this employment follows three guidelines for time series models: specification, estimation, and diagnostics.

For model selection and specification purposes, I will be utilizing a seasonal-means trend model and an ARIMA model for each of the UR and LFPR datasets. First, I will conduct exploratory data analysis (EDA) of both datasets, to observe and determine the current UR & LFPR. Then, I will use seasonal-means trend models to capture the average monthly values of UR and LFPR rates. They will also serve as precursors to the UR's and LFPR's ARIMA models. These ARIMA models will then not only serve as robustness checks on the seasonal-means trend models, but also be used to provide 12-month forecasts for UR and LFPR rates. The orders of these ARIMA models (e.g. their (p, d, q) values) will be determined by the auto.arima() function in R.

For model estimation purposes, the mean values obtained from the UR's and LFPR's seasonal-means trend models and the coefficients in their respective ARIMA models will all be estimated by R. The decision of allowing R to conduct the estimations of the models' values/coefficients is done for the sake for swiftly obtaining accurate and consistent results.

For model diagnostics, the seasonal-means trend and ARIMA models will be checked for their adequacy via rigorous residual analyses and appropriate statistical tests.

For the originally observed time series, I test if the data are stationary via the augmented-Dickey Fuller (ADF) test. EDA and the ADF test will help me in testing the stationarity assumption.

For the seasonal-means trend models, their residuals will be checked via examination of their residual plots/histograms, QQ-plots, and autocorrelation functions (ACFs). Checking the results of the Shapiro-Wilk (SW) and runs tests of their residuals will also be included. This plans helps me in testing the normality assumption (residuals are normally distributed) for these models, deciding whether their residuals are independent, and determining if they are appropriate (e.g. white noise; same for below).

For the ARIMA models, their residuals will be checked via all of the above and also checking the partial autocorrelation functions (PACFs) of their residuals and results of the models' Ljung-Box (LB) tests. This plans helps me in testing the normality assumption for these models, deciding whether their residuals are independent, and determining if they are appropriate.

IV. Results & Discussion

EDA shows that UR and LFPR depict upward trends and have high frequencies for lower and higher values, respectively, as shown in Figure 1. Their time plots could, then, indicate some non-stationarity (unit root), which is confirmed by their ADF tests in Table 3 and their ACFs/PACFs in Figure 4 (autocorrelation/no white noise).

Fitting seasonal-means trend models to UR and LFPR, I obtain the regression coefficients on the months, as Table 2a shows. Interpreting the coefficient on the intercept for each series: Any month is associated with, on average, a 6.25%-point increase in NJ UR rates. Any month is associated with, on average, a 65.35%-point increase in NJ LFPR rates. Table 2b indicates that 0.3% of the variation in UR and 0.04% of the variation in LFPR are explained by their respective models. Such low variation, then, would allow one to entertain other models for better explanation.

Figures 2a and 2b suggest that the residuals of the UR's seasonal-means model do not have a constant mean, have autocorrelation (ACF is decaying/no white noise), and no normal distribution, respectively. Its QQ-plot seems to have questionable quantiles. This observation also seems to be true for the residuals of the LFPR's seasonal-means model, as shown in Figures 3a and 3b (its histogram come close to a normal distribution, however). Table 3 suggests that the runs and SW tests have p-values < 0.05 for both UR's and LFPR's models, which would mean that their residuals are, indeed, not independent or normally distributed (and that the series should be differenced). Perhaps alternative models (e.g. ARIMA) for them would be better, particularly for forecasting.

R fits an ARIMA(0, 1, 0) model and a SARIMA(2, 2, 4)(0, 0, 1)[12] model to UR and LFPR, respectively. Figure 5a suggests that the residuals of UR's model may have constant mean and a normal distribution but have some autocorrelation (note the outliers in the residual plot around 2020). Its QQ-plot seems to have acceptable quantiles. This observation also seems to be true for the residuals of the LFPR's model, as shown in Figure 5b (note, again, the outliers in the residual plot around 2020). Table 3 shows that the runs, SW, and LB tests have p-values < 0.05 for both UR's and LFPR's models, which would mean that their residuals are, indeed, not independent or normally distributed.

Despite the flaws in the residuals of UR's and LFPR's models, I still used the models to provide 12-month forecasts. This is because of the fact that they can provide the most accurate and reliable estimates of all time series methods that I know of (they were chosen by R, after all). Table 4a shows the 12-month forecasts for UR, , which is 8.2, the same for every month. This observation and the model's widening prediction intervals can be found in Figure 6a. Table 4b shows the 12-month forecasts for LFPR, which fluctuate between 61 and 62 every month. Figure 6b depicts these fluctuating forecasts and widening prediction intervals.

V. Conclusion

My analysis of the NJ UR and LFPR datasets has allowed for multiple observations. EDA of UR and LFPR has shown that such rates are disastrous in 2020. Estimates of the mean monthly values of these rates based on trend models, then, might be skewed due to the current unprecedented rates. Such a belief seems to be confirmed by statistical tests - where many statistical assumptions have been violated as a result. Using ARIMA models to provide forecasts for UR and LFPR would mean encountering this belief again, but I believe that they are valid and reliable. This can be shown by checking recent events, where NJ's UR has increased by 8.2% **[4]** and NJ'S LFPR has also been steadily increasing **[1]**. They can be improved upon, however, perhaps by utilizing non-parametric time series methods instead. While the present may be bleak for U.S. citizens (including NJ residents) amidst the pandemic, the future seems promising for all.

References

[1] Jarrett, Kim. “Growth in New Jersey Labor Force Potential Positive Sign amid Increased Unemployment.” *The Center Square*, 20 Nov. 2020, www.thecentersquare.com/new_jersey/growth-in-new-jersey-labor-force-potential-positive-sign-amid-increased-unemployment/article_ad6da902-2b49-11eb-9e66-07daa178ffca.html.

[2] “Labor Force Participation Rate for New Jersey.” *FRED*, 20 Nov. 2020, fred.stlouisfed.org/series/LBSSA34.

[3] Long, Heather. “U.S. Now Has 22 Million Unemployed, Wiping out a Decade of Job Gains.” *The Washington Post*, WP Company, 17 Apr. 2020, www.washingtonpost.com/business/2020/04/16/unemployment-claims-coronavirus/.

[4] “NJ Unemployment Rate Increases to 8.2%.” *You Are Being Redirected...*, 19 Nov. 2020, njbmagazine.com/njb-news-now/nj-unemployment-rate-increases-to-8-.

[5] “Overview of BLS Statistics on Unemployment.” *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, 16 Dec. 2013, www.bls.gov/bls/unemployment.htm.

[6] “Unemployment Rate in New Jersey.” *FRED*, 20 Nov. 2020, fred.stlouisfed.org/series/NJUR.

[7] USAFacts. “48% Of Americans Are Feeling down, Depressed, or Hopeless during the COVID-19 Pandemic.” *USAFacts*, USAFacts, 22 Sept. 2020, usafacts.org/articles/45-americans-are-feeling-down-depressed-or-hopeless-during-covid-19-pandemic/.

Appendix A – Tables

Table 1: Descriptive Statistics for UR & LFPR Datasets

	Dimensions of Data	Minimum Value	1st Quarter	Median	Mean	3rd Quarter	Maximum Value	Number of Missing Values
UR	538 x 2	3.3	4.6	5.9	6.3	7.6	16.8	0
LFPR	538 x 2	61.10	64.10	65.90	65.38	66.60	67.90	0

Table 2a: Regression Coefficients on Months in Seasonal-Means Trend Models for UR & LFPR

	UR	LFPR
Intercept	6.25*** (0.31)	65.35*** (0.23)
February	-0.02 (0.44)	0.02 (0.33)
March	-0.04 (0.44)	0.04 (0.33)
April	0.24 (0.44)	0.004 (0.33)
May	0.20 (0.44)	0.02 (0.33)
June	0.22 (0.44)	0.02 (0.33)
July	0.15 (0.44)	0.04 (0.33)
August	0.07 (0.44)	0.04 (0.33)
September	-0.04 (0.44)	-0.03 (0.33)
October	-0.02	-0.01

	(0.44)	(0.33)
November	-0.07 (0.44)	0.07 (0.33)
December	-0.09 (0.44)	0.09 (0.33)

Note: Standard errors in parentheses. *** indicates p < 0.01. ** indicates p < 0.05. * indicates p < 0.10. Numbers are rounded to nearest 2 digits.

Table 2b: Summary Statistics of Seasonal-Means Trend Models for UR & LFPR

	Residual Standard Error	(Multiple) R-squared	F-Statistic	P-Value
UR	2.095	0.003091	0.1483	0.9994
LFPR	1.576	0.0004007	0.01917	1

Table 3: Results of Statistical Tests Done on UR & LFPR Models' Residuals

	Runs	**Shapiro-Wilk (SW)**	**Augmented Dickey-Fuller (ADF)**	**Ljung-Box (LB)**
	P-Value			
Seasonal-Means Trend for UR	7.26e-145	8.468e-16	0.1314*	
Seasonal-Means Trend for LFPR	1.01e-143	2.312e-15	0.7435*	
ARIMA(0, 1, 0)	1.09e-38	< 2.2e-16		9.976e-12
SARIMA(2, 2, 4)(0, 0, 1)[12]	2.93e-13	< 2.2e-16		6.945e-06

Note: * indicates p-values for originally observed series.

Table 4a: Forecasts Based on UR ARIMA Model

Month	Year	Point Forecast	Lower 95% Prediction Interval	Upper 95% Prediction Interval
November	2020	8.2	6.98	9.42
December	2020	8.2	6.48	9.92
January	2021	8.2	6.09	10.31
February	2021	8.2	5.76	10.64
March	2021	8.2	5.47	10.93
April	2021	8.2	5.21	11.19
May	2021	8.2	4.97	11.43
June	2021	8.2	4.75	11.65
July	2021	8.2	4.54	11.86
August	2021	8.2	4.34	12.06
September	2021	8.2	4.16	12.24
October	2021	8.2	3.98	12.42

Note: Integers are rounded to the nearest 2 digits.

Table 4b: Forecasts Based on LFPR SARIMA Model

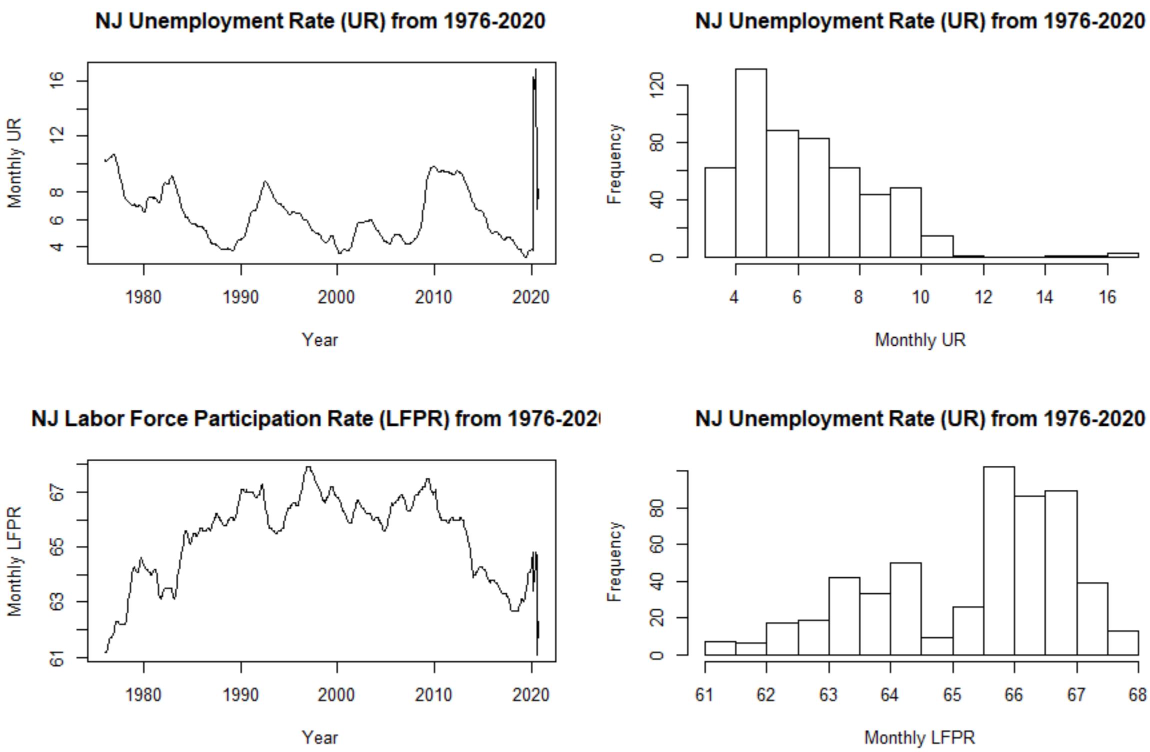
Month	Year	Point Forecast	Lower 95% Prediction Interval	Upper 95% Prediction Interval
November	2020	61.45	61.08	61.82
December	2020	62.68	62.17	63.19
January	2021	62.16	61.46	62.85
February	2021	61.66	60.89	62.44
March	2021	61.83	60.98	62.67
April	2021	62.33	61.39	63.27
May	2021	61.92	60.90	62.95
June	2021	61.79	60.70	62.89
July	2021	61.58	60.42	62.75

August	2021	61.52	60.28	62.77
September	2021	62.71	61.40	64.02
October	2021	62.25	60.86	63.63

Note: Integers are rounded to the nearest 2 digits.

Appendix B – Figures

Figure 1: Descriptive Graphs of UR & LFPR Datasets



Note: Title of the bottom-right graph should be "NJ Labor Force Participation Rate (LFPR) from 1976-2020" instead.

Figure 2a: Residual Plot & Histogram of UR Trend Model

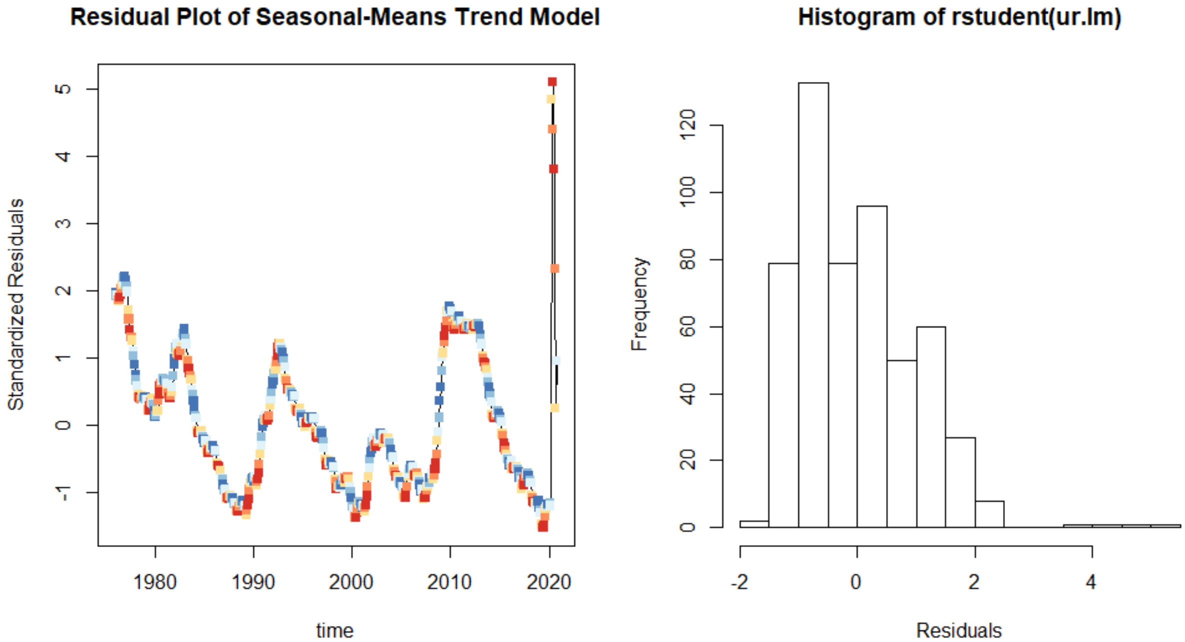


Figure 2b: ACF & QQ-Plot of UR Trend Model's Residuals

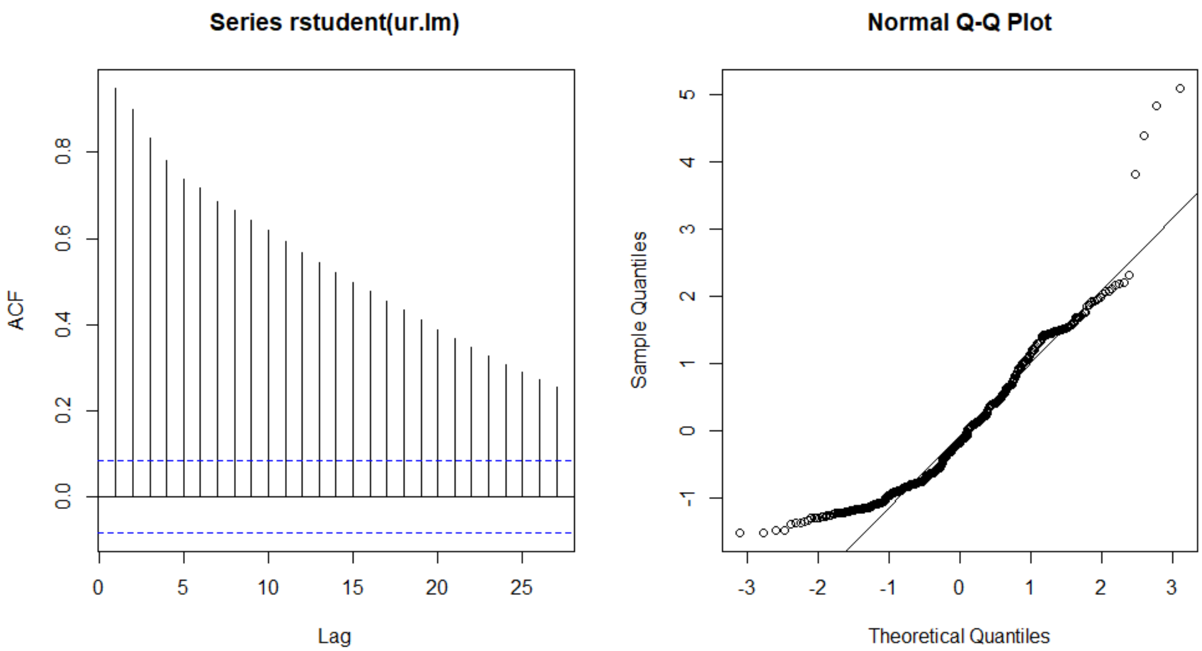


Figure 3a: Residual Plot & Histogram of LFPR Trend Model

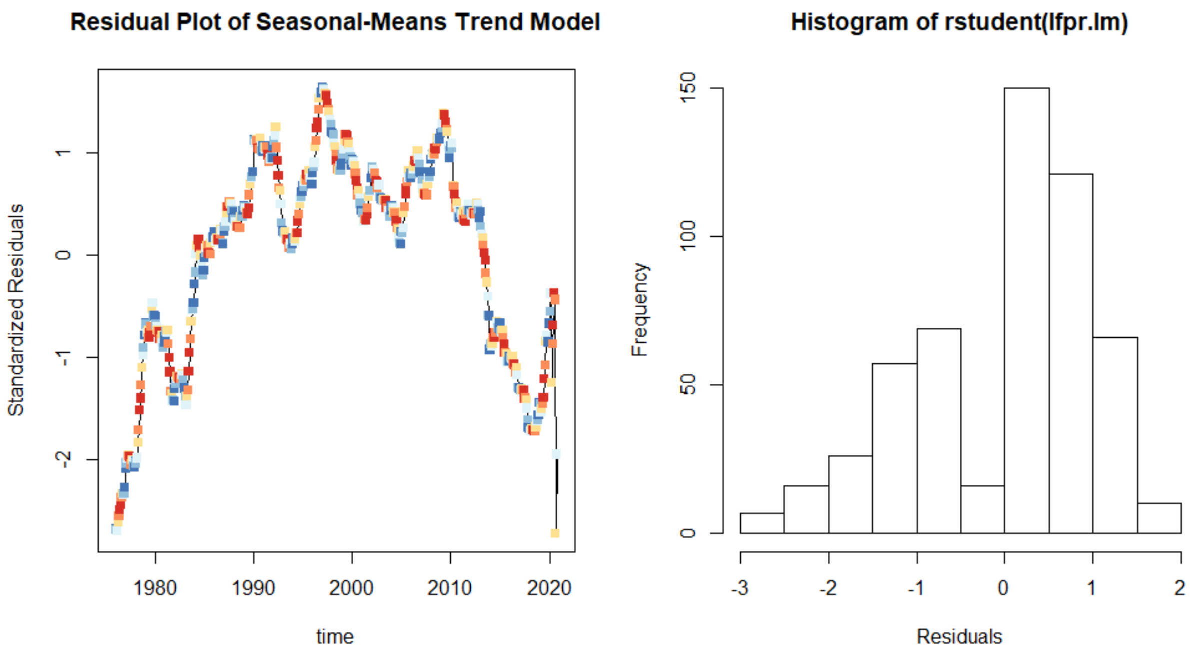


Figure 3b: ACF & QQ-Plot of LFPR Trend Model's Residuals

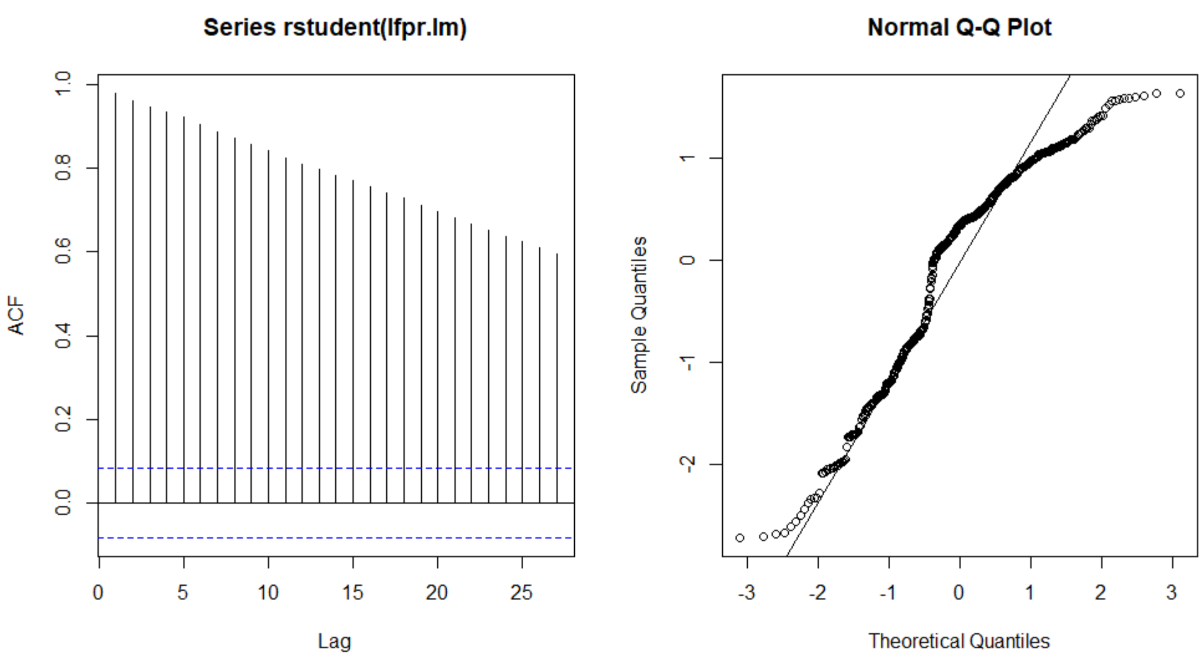


Figure 4: ACF & PACF of Differenced Observed UR & LFPR Series

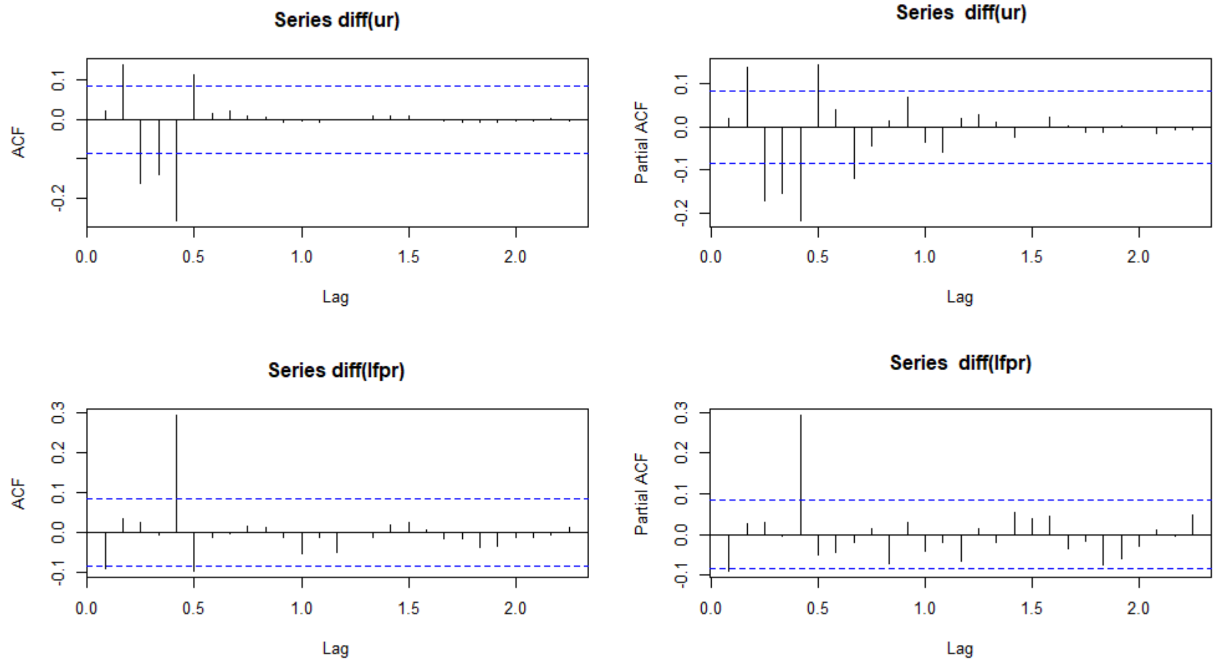


Figure 5a: Residual Analysis of ARIMA(0, 1, 0) Model

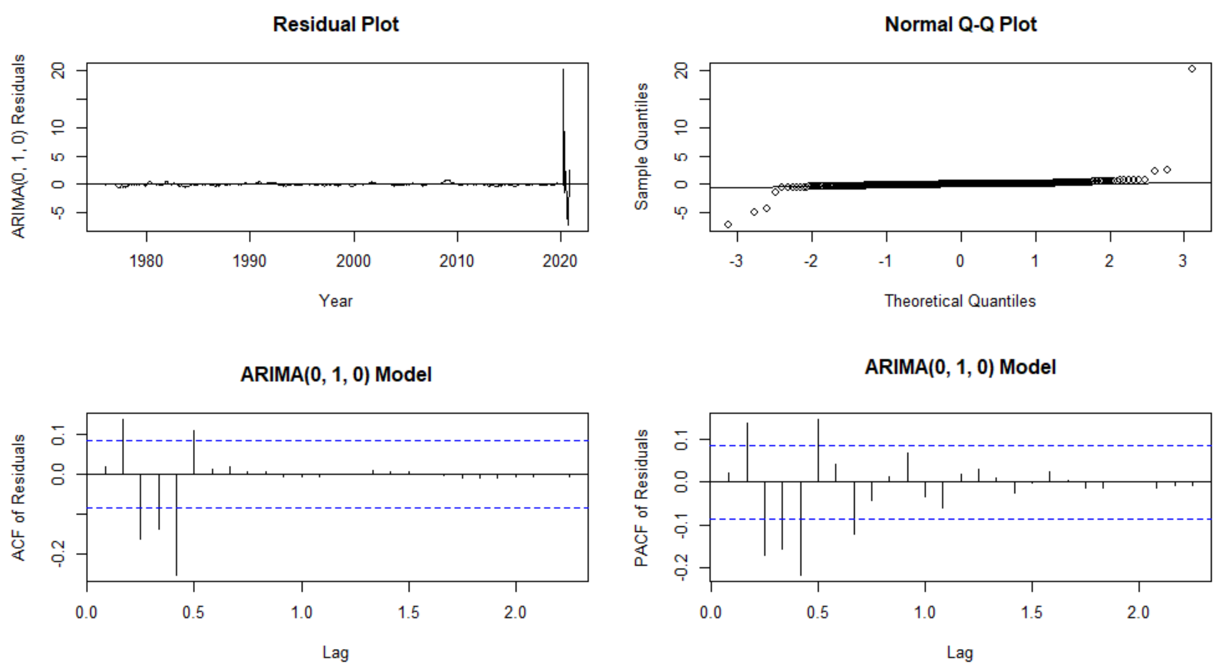
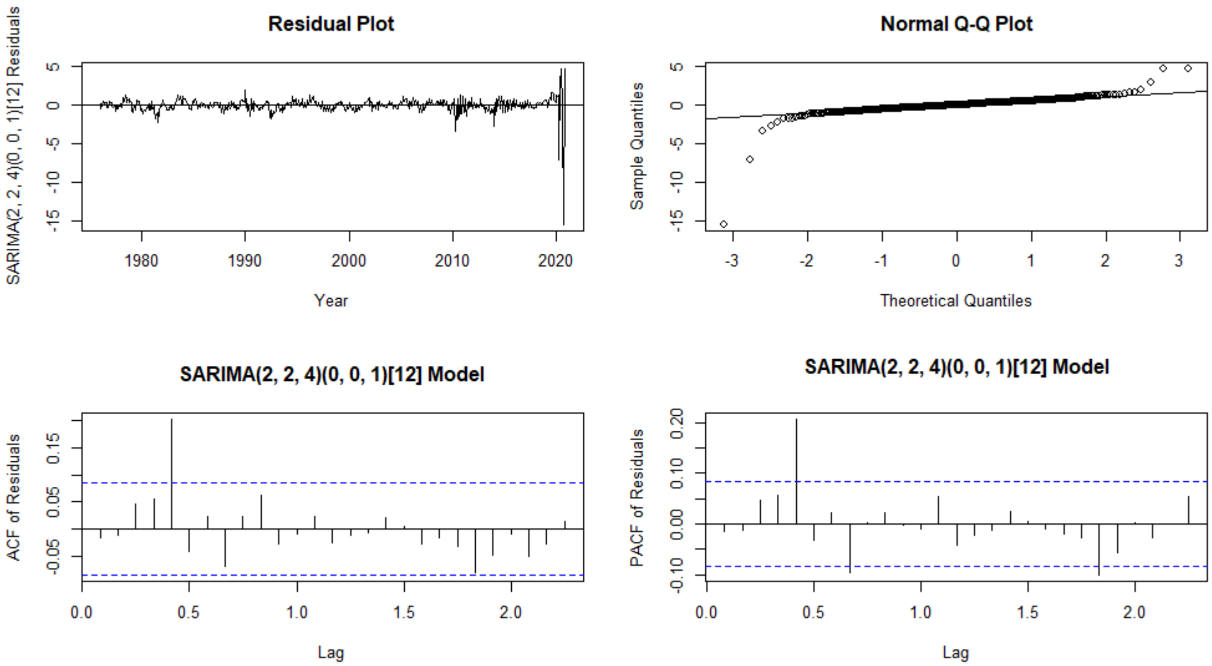
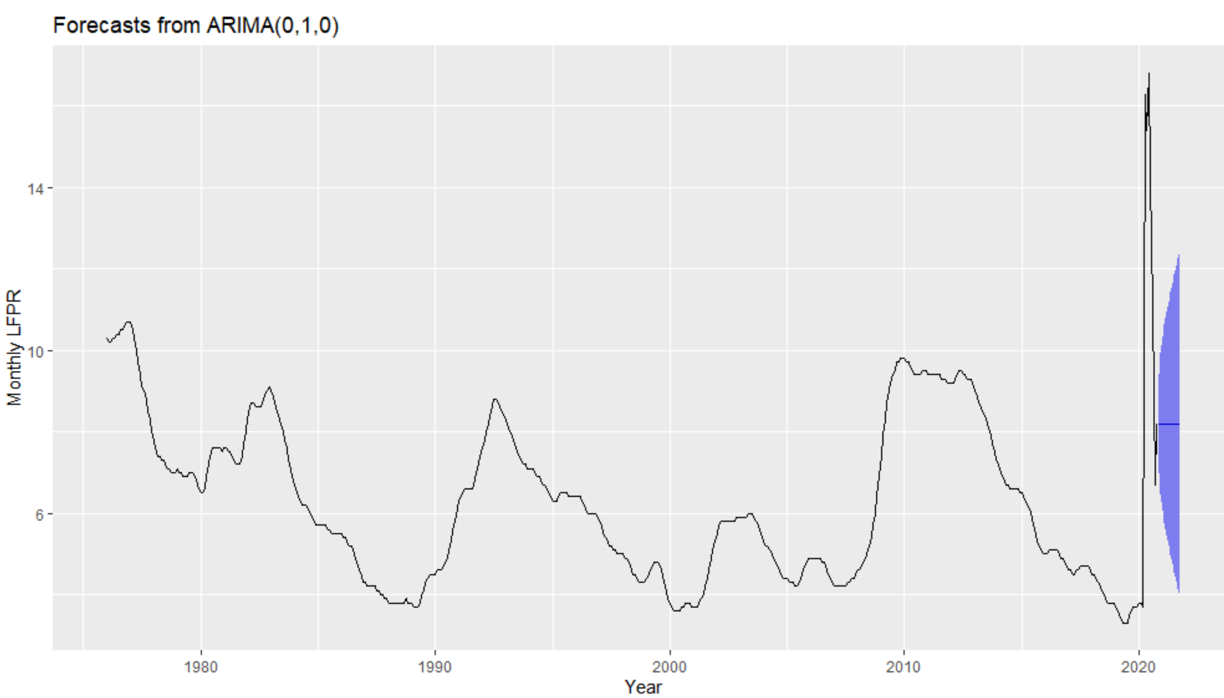


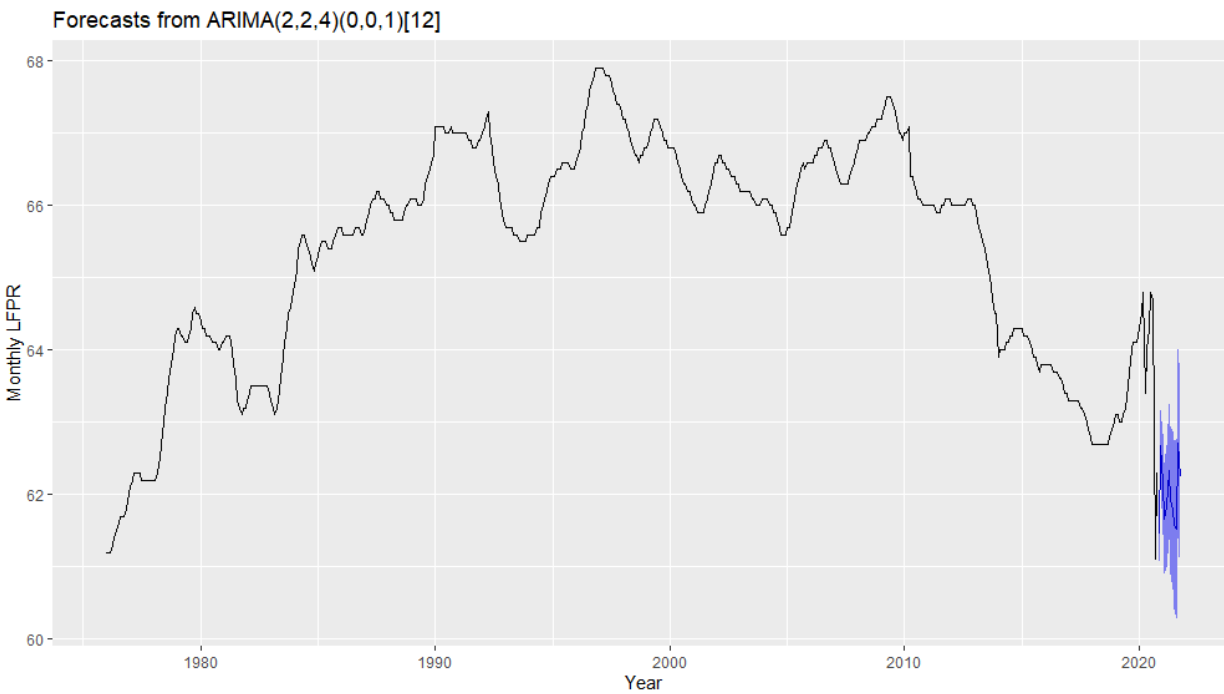
Figure 5b: Residual Analysis of SARIMA(2, 2, 4)(0, 0, 1)[12] Model



<u>Figure 6a: Forecasts Based on ARIMA(0, 1, 0) Model</u>



<u>Figure 6b: Forecasts Based on SARIMA(2, 2, 4)(0, 0, 1)[12] Model</u>


Forecasts from ARIMA(2,2,4)(0,0,1)[12]
Monthly LFPR
68
66
64
62
60
1980
1990
2000
2010
2020
Year