

Forecasting US Monthly Unemployment Rate

Weilin Yin

14/04/2022

Abstract

Unemployment has been one of the most popular and important concepts in Economics since it reflects the economic growth of a country. Hence, understanding the dynamic of the unemployment rate can provide an alternative way to assess the condition of an economy. In this study, the 1948-2016 US monthly unemployment rate data is utilized to build a seasonal autoregressive intergrated moving average model (SARIMA). Using the final model, 10-step ahead (10 months) prediction is made. The result is that, for the next 10 months after 2016, the monthly unemployment rate remains flat. Also, a spectral analysis is performed to identify the first three predominant periods. Due to large confidence interval range, no significance is established for these periods. The forecasting model also has a few pitfalls which are also discussed.

Keywords

Time Series Analysis, Unemployment, Seasonal Autoregressive Integrated Moving Average (SARIMA)

Introduction

Unemployment has been one of the most popular topics in Economics because it is one of the measurements of a country's economic performance. Thus, it is strongly related to the gross domestic product (GDP). In macroeconomic, Okun (1962) developed the Okun's law states that a 1% reduction in GDP would have 3% higher unemployment rate. Farsio and Quade (2003) proved that the negative relationship exists using the twenty-five years quarterly US unemployment data. Hence, having a model that could forecast the unemployment rate would be beneficial as it also provides a prediction to the GDP and economic performance as well. In the current literature, there have been many researchers using different methods and models to make prediction on US unemployment rate. Kreiner and Duca (2019) applied multiple models (e.g., neural network, Lasso regression, Naïve forecast, and SPF forecast) to make predictions on US unemployment rate, all the prediction indicating a decreasing trend after 2016. In this study, a seasonal autoregressive intergrated moving average model (SARIMA) is utilized to forecast the unemployment rate. The data I used in this study is the US monthly unemployment rate data from 1948 to 2016.

The rest of the paper is organized in the following way: **Statistical Methods** section discusses the exploratory data analysis and the model building process. **Results** section demonstrates the final model used to forecast, the result of prediction, and the spectral analysis. **Discussion** section concludes the study and points out the limitations of this study as well as suggestion on future researches.

Statistical Methods

Before building the model, it is important to examine the data to see if any adjustment is needed to correctly fit a model. From Figure 1, we can see that there are significant fluctuations from year to year. Hence, in order to stabilized the data, the monthly unemployment rate x_t is differenced for one time. Now by doing so, the trend becomes stable and we have

a time series ∇x_t . Moreover, it is noticeable that the differenced data still have seasonal pattern, thus a twelve-order difference on the time series ∇x_t is applied. The results after these two transformations is presented in Figure 2. From Figure 2, it can be confirmed that the time series $\nabla_{12}\nabla x_t$ is stationary and has a constant mean and variance.

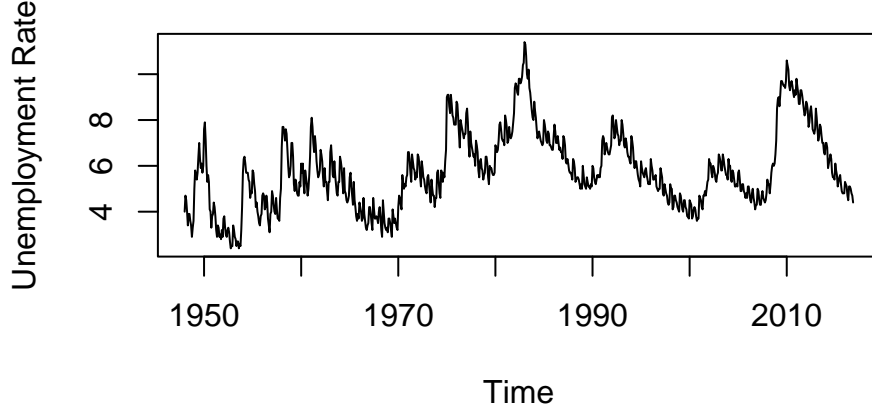


Figure 1: Monthly Unemployment rate in percent from 1948 to 2016

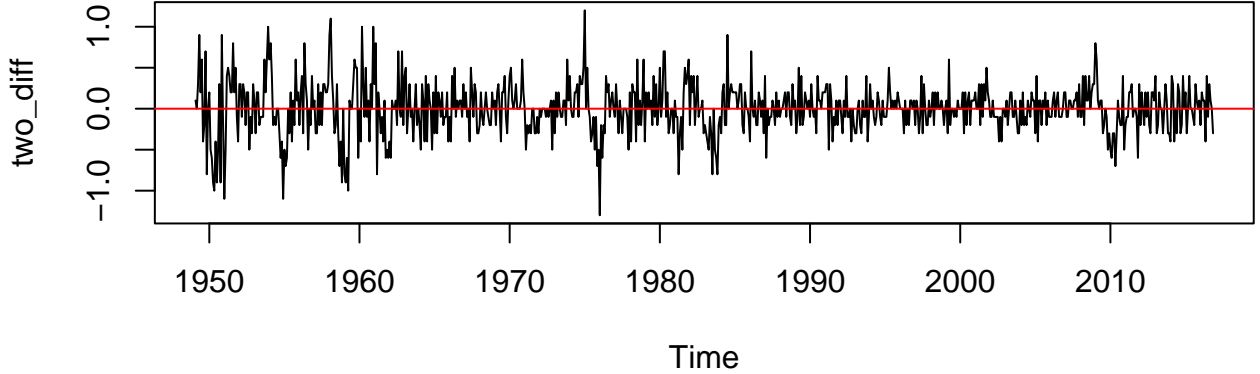


Figure 2: Plot of the differenced data

Since the data was taken seasonal difference, the SARIMA model is strongly preferred. In order to build the model, we need to find out the seasonal elements and non-seasonal elements of the model. This procedure is done by inspecting the ACF and PACF graphs of the time series $\nabla_{12}\nabla x_t$ in Figure 3. For seasonal components, we can see that both ACF and PACF cut off after lag = $1 \times s$ where $s = 12$ since we took the twelve-order difference. Hence, we can set $P = 1$, $Q = 1$, $D = 1$, and $s = 12$. In terms of the non-seasonal components, from Figure 3, the ACF cuts off after lag = 2 or lag = 3 and the PACF cuts off after lag = 2.

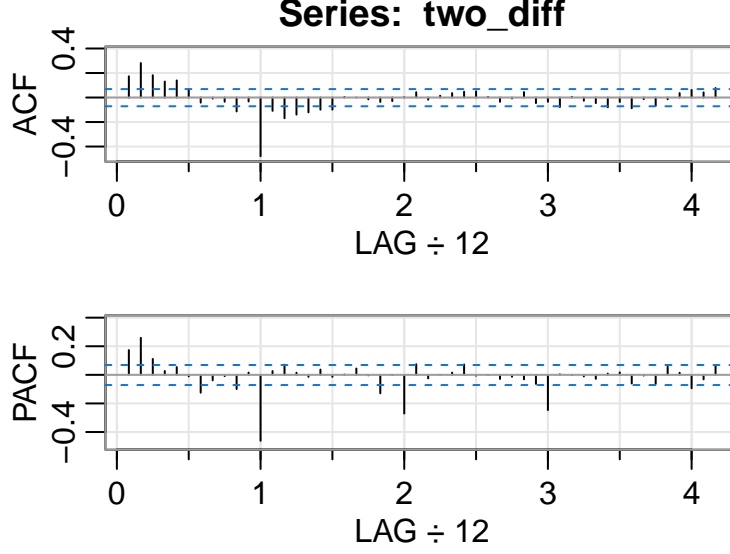


Figure 3: ACF and PACF for differenced data

Hence, we can say that $p = 2$, $q = 2$ or 3 , and $d = 1$. With both of the seasonal and non-seasonal components, two models are proposed.

$$ARIMA(2, 1, 2) \times (1, 1, 1)_{12}$$

$$ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$$

Results

Figure 4 and Figure 5 demonstrates the diagnostic plots for the two proposed model. For $ARIMA(2, 1, 2) \times (1, 1, 1)_{12}$ and $ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$, the residual plots demonstrates no obvious pattern and the ACF of residual plots shows that there is no spike that is significant enough at 5% significance level. Hence, we can conclude that both model satisfy the randomness assumption. The Normal Quantile-Quantile plots of standard residuals for both models illustrates that the normality assumption is satisfied well enough even though there are a few outliers at the tails. For the plot of p-values for Ljung-Box statistics of two models, most of the points are above the 0.05 line which means we do not reject the null hypothesis that the residuals are independent. So far, both models perform equally well

according to the set of diagnostic plots. Table 1 displays the AIC, BIC, and AICc for each model. By comparing these criteria, we can see that $ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$ has slightly smaller AIC, BIC, and AICc.

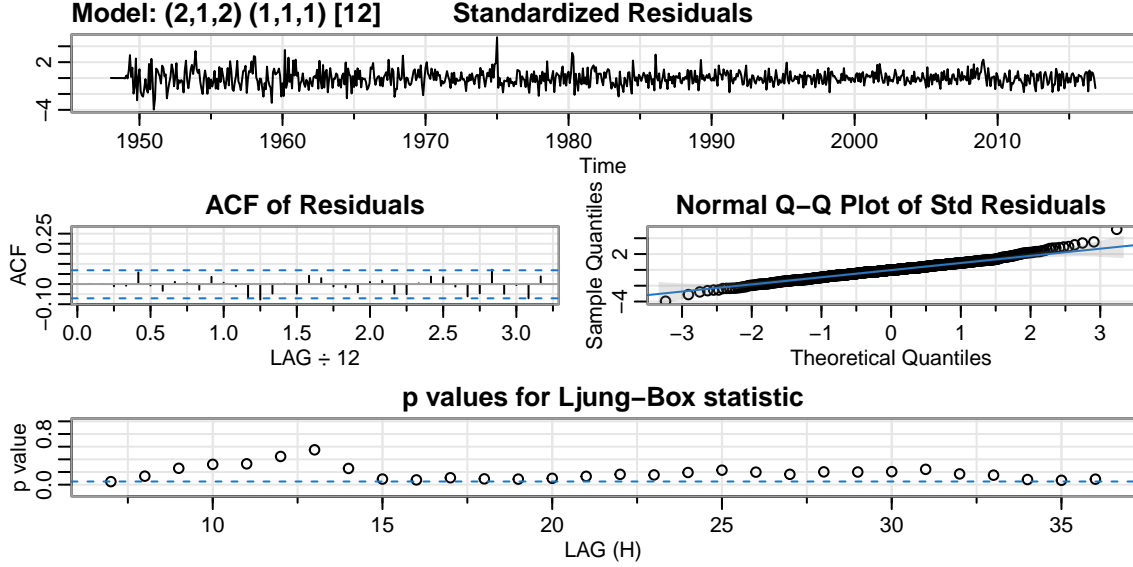


Figure 4: Residual Analysis for $ARIMA(2, 1, 2) \times (1, 1, 1)_{12}$

Table 1: AIC, AICc, BIC for two proposed models

Model	AIC	AICc	BIC
$ARIMA(2, 1, 2) \times (1, 1, 1)_{12}$	-0.0214464	-0.0213186	0.0189881
$ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$	-0.0378803	-0.0377096	0.0083306

After the model diagnostic and comparison process, the final model we use to forecast is displayed below.

$$(1 - \Phi_1 B^{12})(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})(1 - B)x_t = (1 + \Theta_1 B^{12})(1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)w_t$$

where x_t is unemployment time series, w_t is the white noise that $w_t \sim N(0, \sigma_w^2)$, B is for backshift operator. $\Phi_1, \phi_1, \phi_2, \Theta_1, \theta_1, \theta_2, \theta_3$ are all model parameters.

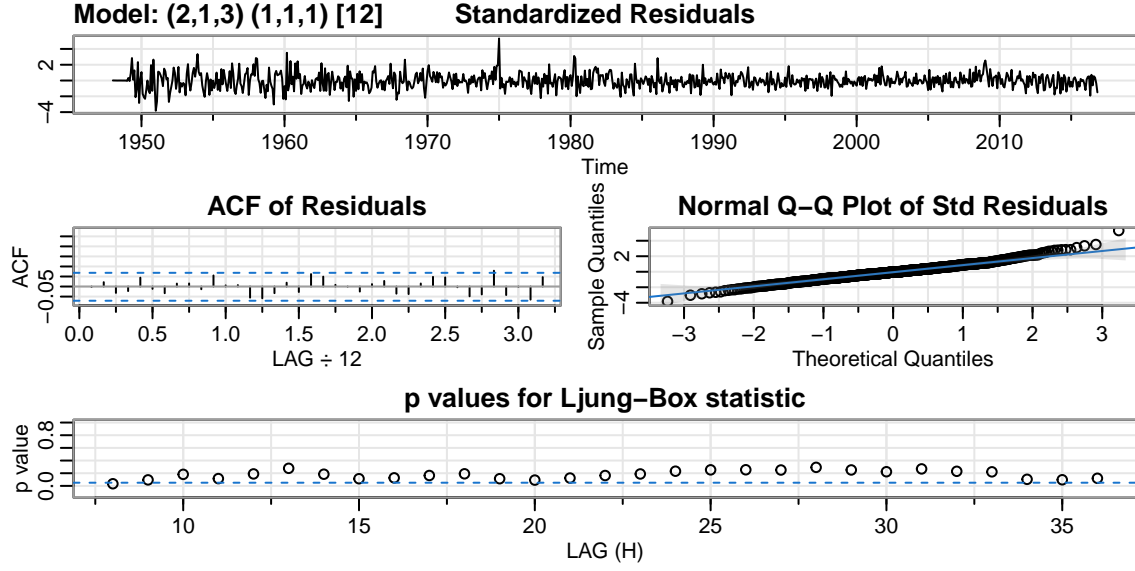


Figure 5: Residual Analysis for $ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$

Table 2: Parameter Estimation for $ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$

	Estimate	Std. Error	t-value	p-value
ϕ_1	1.8265	0.0505	36.1530	0.0000
ϕ_2	-0.8361	0.0495	-16.8781	0.0000
θ_1	-1.7432	0.0610	-28.5747	0.0000
θ_2	0.8323	0.0711	11.7135	0.0000
θ_3	-0.0891	0.0394	-2.2590	0.0241
Φ_1	0.0007	0.0530	0.0129	0.9897
Θ_1	-0.7527	0.0382	-19.7040	0.0000

By fitting the final model we developed in the previous section, we get the estimations of each model parameter and display in Table 2. Among all of the estimates, most of estimation for parameters are statistically significant as their p-values are less than 0.05, except for the

estimate of Φ_1 . Therefore, the fitted model is in the form of the following:

$$(1-1.8265B+0.8361B^2)(1-B^{12})(1-B)\hat{x}_t = (1-0.7527B^{12})(1-1.7432B+0.8323B^2-0.0891B^3)\hat{w}_t$$

The estimates we got for parameters tell us how influential the previous months unemployment rates are on current month unemployment rate. For example, the estimate of ϕ_1 is 1.8265 which means that the current unemployment rate depends on the previous month unemployment rate with the weight of 1.83. Similarly, the estimate of ϕ_2 is -0.8361 means that the current unemployment rate is influenced negatively by the unemployment rate of the month before last month with a weight of 0.84. To be more specific, for the parameters of the model, $d = 1$ indicates there is first-order differencing, $D = 1$ means that there is one seasonal differencing. $p = 2$ and $P = 1$ represents the number of AR terms respectively. $q = 3$ and $Q = 1$ represents the number of MA terms respectively. $s = 12$ means that there are 12 observations in one seasonal cycle.

Using the fitted model, we can continue to forecast the monthly unemployment rate for the next 10 months after 2016. Figure 6 shows the predications for the next 10 months in red points with their 95% confidence interval as grey. Also, the values of predictions and the values of 95% confidence interval for each prediction are listed in Table 3.

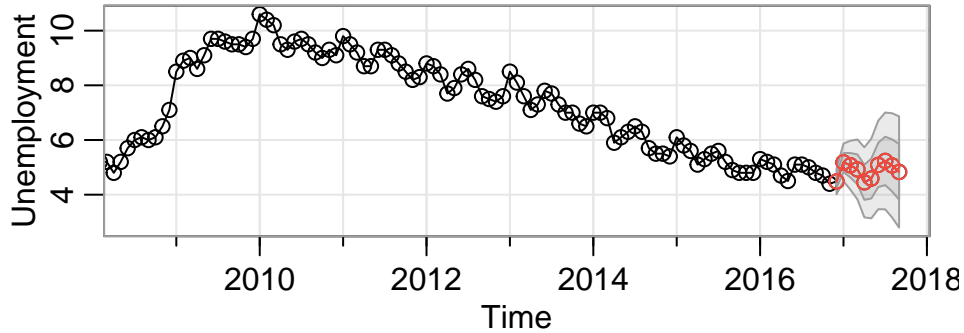


Figure 6: Forecasting unemployment rate for next 10 months using $ARIMA(2,1,3) \times (1,1,1)_{12}$

Table 3: Forecasting unemployment rate for next 10 months using $ARIMA(2, 1, 3) \times (1, 1, 1)_{12}$ with 95 percent Confidence Interval

Future Month	Predicted Value	Lower Bound of 95% CI	Upper Bound of 95% CI
1	4.482261	4.025475	4.939048
2	5.181088	4.507236	5.854941
3	5.075584	4.196999	5.954168
4	4.925324	3.852473	5.998176
5	4.451201	3.195027	5.707374
6	4.590319	3.162437	6.018201
7	5.088450	3.500919	6.675980
8	5.234989	3.500033	6.969946
9	5.069969	3.199720	6.940217
10	4.830603	2.836908	6.824299

From Figure 6 and Table 3, we can conclude several findings. First of all, the trend of unemployment rate for the next 10 months is slightly upward. This means that, for the next 10 months after 2016, unemployment would remain at the same level. This finding is consistent to what Kreiner and Duca (2019) found using several models. Secondly, by looking at the 95% confidence interval, we can see that the magnitude of the confidence interval is relatively small compared to the magnitude of the estimation. Thus, we can say that the forecasting is relatively precise in terms of the economic significance.

Next, we perform a spectral analysis to identify the first three predominant periods. In Figure 7, we can observe that the first peak occurs when frequency is 0.03. In theory, $\text{frequency} = \frac{1}{\text{cycle}} = \frac{1}{12}$. What we observed is close to what the theory states.

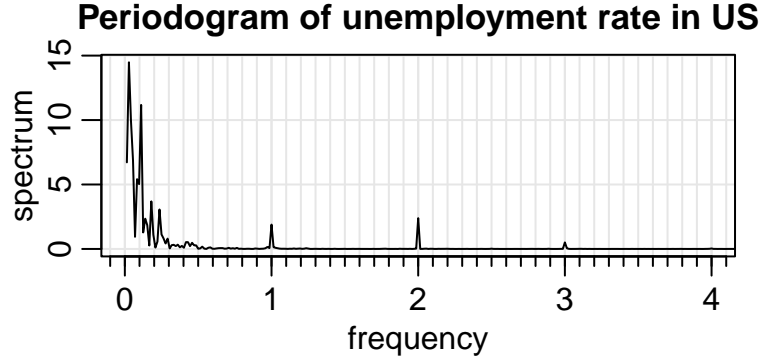


Figure 7: Periodogram of unemployment rate in US

Table 4: First three predominant periods and their 95 percent confidence interval

Frequency	Spectrum	Period	Lower bound of 95% CI	Upper bound of 95% CI
0.0278	14.4829	35.97122	3.926097	572.0440
0.1111	11.1720	9.00090	3.028562	441.2704
0.0417	9.9625	23.98082	2.700685	393.4977

Table 4 above shows the first three predominant periods of US monthly unemployment rate and their 95% confidence interval. Each of the confidence interval has a large range which means that it is hard to retrieve any useful information of significance. For the first period, the periodogram ordinate is 14.48 which is also within the 95% confidence intervals for period 2 and period 3, thus we cannot establish significance for the first peak. For the second period, the periodogram ordinate is 11.17 which is also with in the 95% confidence intervals of period 1 and period 3, hence we cannot say that a the significance for the second peak is established. For the third period, the periodogram ordinate is 9.96 which is also within the 95% confidence intervals of period 1 and period 2. Therefore, no significance for the third peak can be established. In short, since the confidence interval is too large, we can extract little information in terms of the significance.

Discussion

According to the result of forecasting, we can conclude that the monthly unemployment rate would remain at a constant level for the next 10 months after 2016. One important factor that could explain the changes of the downward trending pattern to a flat pattern is that Donald Trump became the President of USA in 2017. Thus, at the beginning of 2017, most people were still observing and waiting to see how Trump would do to the market. Thus, there was no significant change in labour market. If we look at the actual unemployment rate after October in 2017, we would see a huge drop in the rate. This is because Trump introduced the tax cut Act which reduces the corporate tax (Silva, 2021). This Act would promote lower unemployment rate in labor market as firms had more budget to hire more workers. Hence, our prediction indeed is consistent with the actual data. Despite the reasonable forecasting, the model we built still suffers from a few limitations. First of all, the data we used to build the forecasting model is before the COVID-19 pandemic outbreak. The global pandemic has changed the labor market entirely. Firms have tighten their budget to survive from the economic recession caused by the pandemic. Numbers of workers became unemployed due to the layoffs of their companies. The model does not account for the exogenous factors like COVID-19 pandemic. Secondly, there are outliers in the data that we did not remove. From the Normal Q-Q plot of the final model, we can see that the normality assumption is not completely satisfied as outliers exist at the tail. However, we considered that the assumption is satisfied well enough so that we proceeded. This may potentially introduce bias to the model. For future researches, we can collect more latest data to incorporate the effect of global pandemic to the labor market, and find a data that has better quality and less outliers so that assumptions are well satisfied.

Reference

- [1] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- [2] Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- [3] David Stoffer (2021). *astsa: Applied Statistical Time Series Analysis*. R package version 1.14. <https://CRAN.R-project.org/package=astsa>
- [4] Yihui Xie (2021). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.34.
- [5] Farsio, F., & Quade, S. (2003). An empirical analysis of the relationship between GDP and unemployment. *Humanomics*, 19(3), 1–6. <https://doi.org/10.1108/eb018884>
- [6] Kreiner, A., & Duca, J. (2019). Can machine learning on economic data better forecast the unemployment rate? *Applied Economics Letters*, 27(17), 1434–1437. <https://doi.org/10.1080/13504851.2019.1688237>
- [7] Silva. (2021). Who benefited most from the trump tax cuts? Policygenius. Retrieved April 17, 2022, from <https://www.policygenius.com/taxes/who-benefited-most-from-the-tax-cuts-and-jobs-act/>