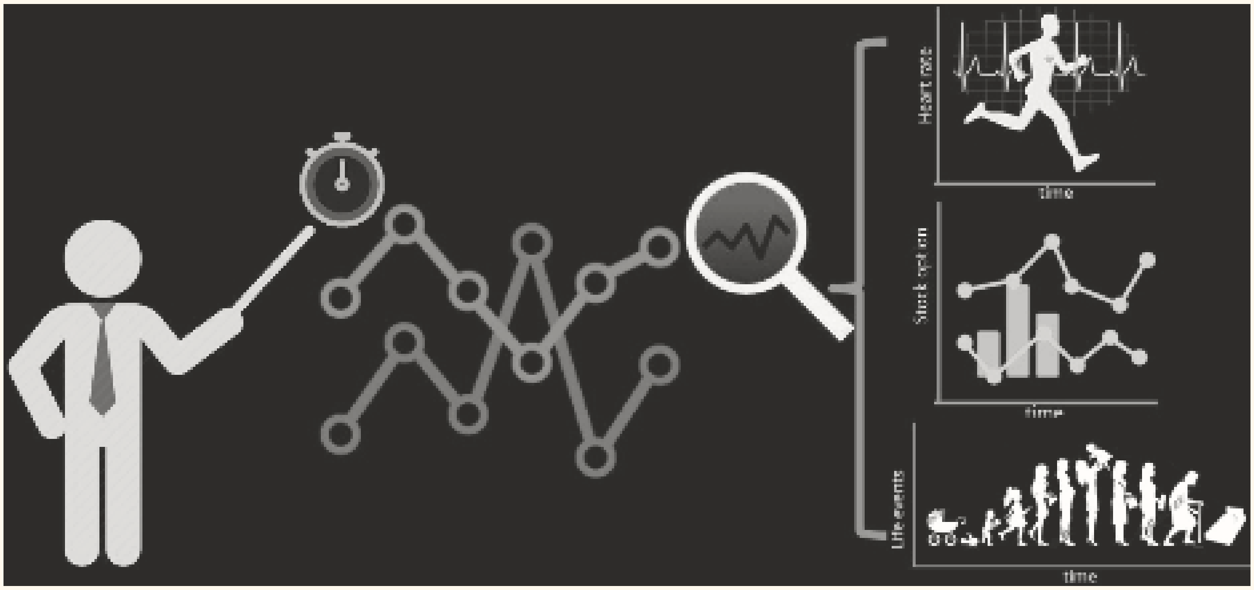
Time Series Analytics and Forecasting
Dr. Ye Cai

US Unemployment Rate Time Series

Team: Apurva Shekhar, Harsh Tandon, Jaskaran Singh Kohli, Vidhi Gandhi



Background

Unemployment is an important yardstick that defines the condition of a country's economy. The unemployment rate has several consequences on the country's economy such as loss of productive forces, loss of income, as well as a burden on the state budget. Continued and persistent employment rate helps bolster the country's social and economic status.

The unemployment rate has been the primary summary statistic for the health of the labor market for quite some time. Recently, however, forecasts of the unemployment rate have come to the forefront, as monetary policy makers are trying to formulate a way of conditioning expectations in the new and extraordinary policy environment.

The unemployment rate has varied from as low as 1% during World War I to as high as 25% during the Great Depression. More recently, it hit 10.8% in November 1982 and 10.0% in October 2009. Unemployment tends to rise during recessions and fall during expansions. From 1948 to 2015, unemployment averaged about 5.8%. The United States has experienced 11 recessions since the end of the postwar period in 1948.

Purpose

The purpose of this report is to identify the most appropriate model to forecast future unemployment rate in the US using the historical data. We present an in-depth study of the forecasts for the monthly U.S. unemployment rate using various time series models and comparing them to further our understanding of the strengths and deficiencies of these methods.

Data

The data used for the purposes of this report represents the US Unemployment Statistics from January 1990 - December 2016, broken down by state and month. The formatted version of the data in CSV format for the purposes of this analysis was obtained from Kaggle. The raw unformatted data is available at the United States Bureau of Labor Statistics Website.

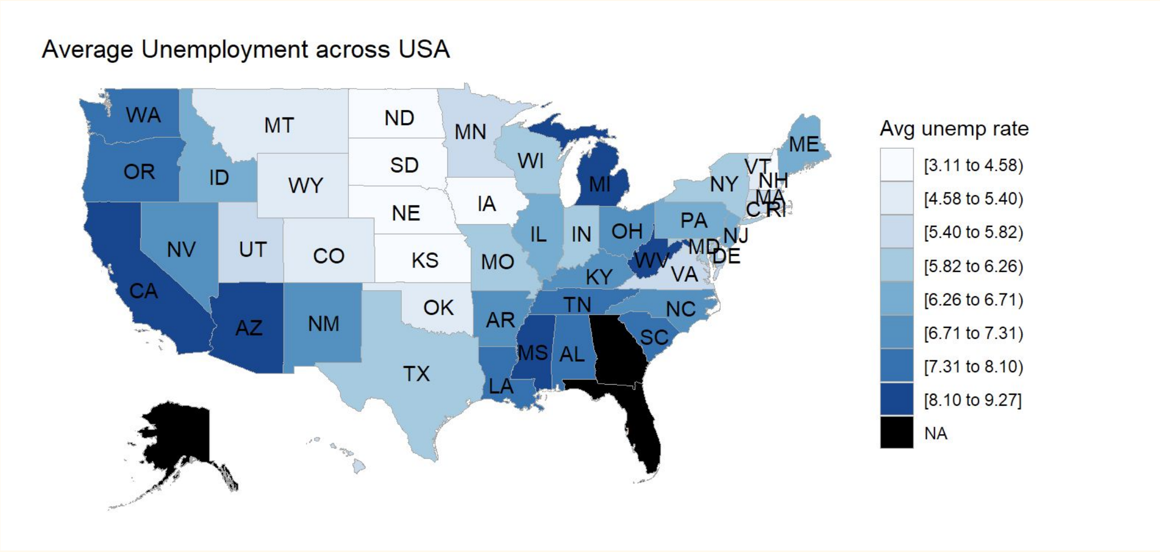
Note: These unemployment rates are monthly U-3 rates, and are **NOT** seasonally adjusted or categorized by age, gender, level of education, etc.

	Year	Month	State	County	Rate
	<int>	<chr>	<chr>	<chr>	<dbl>
1	2015	February	Mississippi	Newton County	6.1
2	2015	February	Mississippi	Panola County	9.4
3	2015	February	Mississippi	Monroe County	7.9
4	2015	February	Mississippi	Hinds County	6.1
5	2015	February	Mississippi	Kemper County	10.6
6	2015	February	Mississippi	Calhoun County	6.9

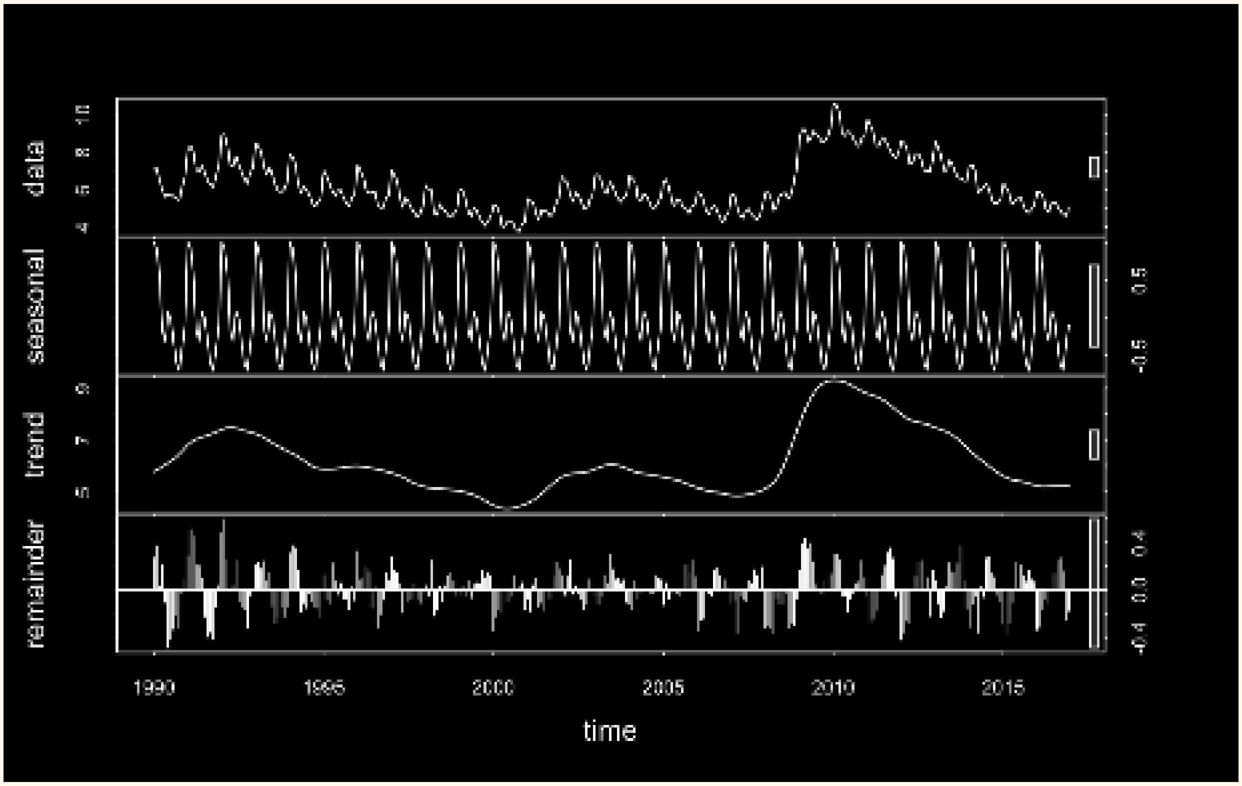
Exploratory Data Analysis

Overview Of Data:

To see how these rates varied from state to state, we plotted a map shown below. We see that states like California, Arizona and Michigan have fairly high average rates of unemployment. We can also see that Central America has a fairly lower unemployment rate than the east & west coast, suggesting that coasts are rather more volatile in jobs.



Before we proceed with Forecasting, let's do some exploratory data analysis. We broke the data down into its seasonality, trend and remainder components. We used the *stl* function.

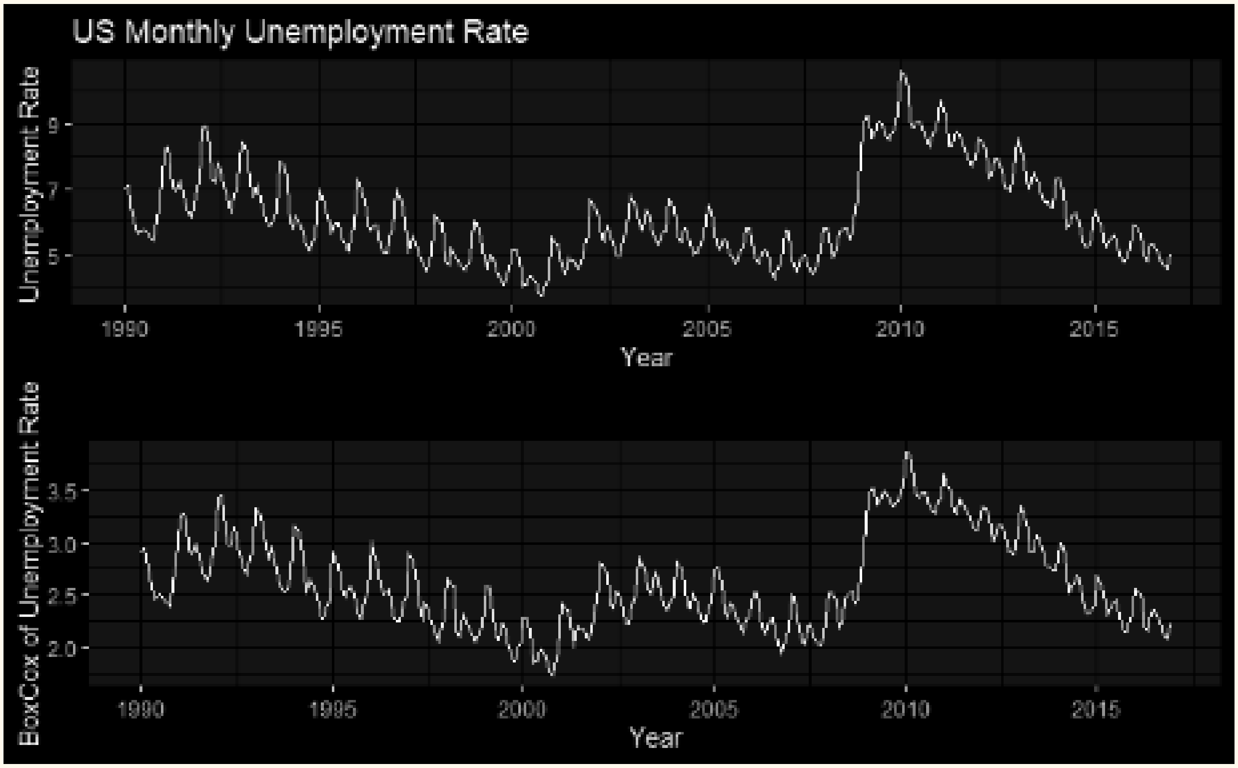


Observations:

- Strong seasonality in the data.
- No strong trend.
- It has a cyclic pattern shown by the rise and fall of unemployment rates during recessions & expansions.
- The cycle lengths (in years) vary over time.

Box-Cox Transformation:

For a good analysis, we want our data to be normally distributed. Box Cox Transformation is a way to transform this non-normal data into a normal shape. The transformation parameter lambda is chosen using *BoxCox.lambda*. We got *lambda* = ***0.3888.***

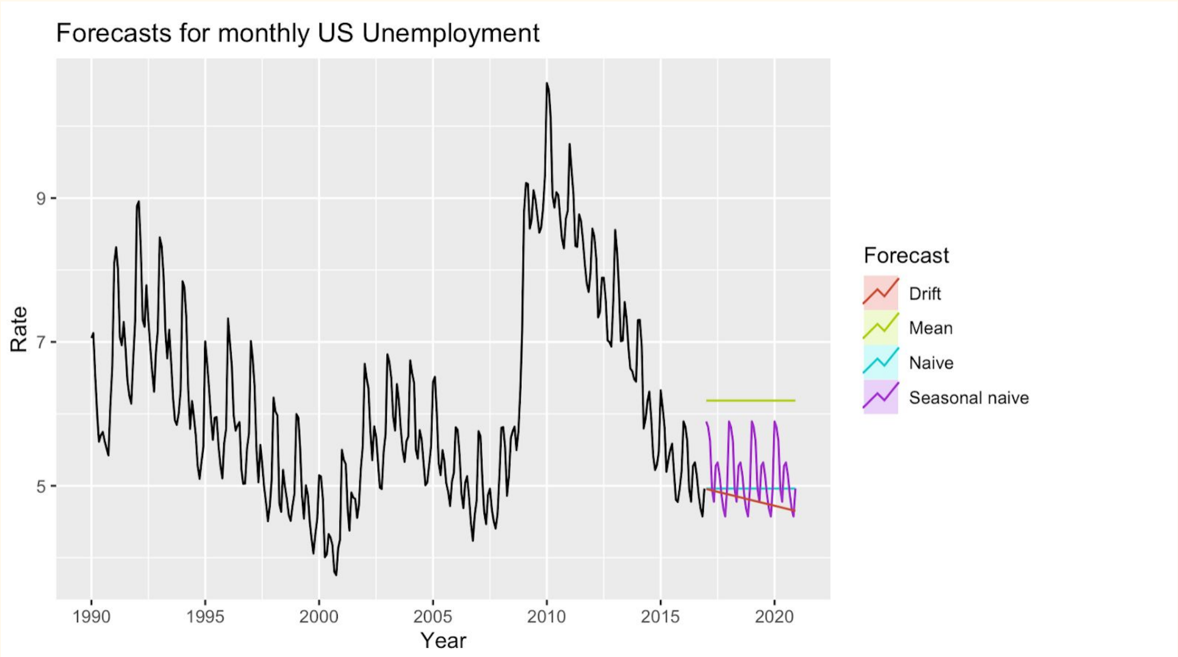


The graph of the Box Cox transformed series appears to have a slightly more even magnitude for the seasonal component than the untransformed one. So we will be using this transformed series for further analysis.

Time Series Analysis

Simple Forecast

We started the analysis with a simple forecast model using basic techniques such as Mean, Naive, Seasonal Naive and Drift. The drift model best fits the data because it has the lowest RMSE. However, the model is primitive as it does not include the seasonality factor.

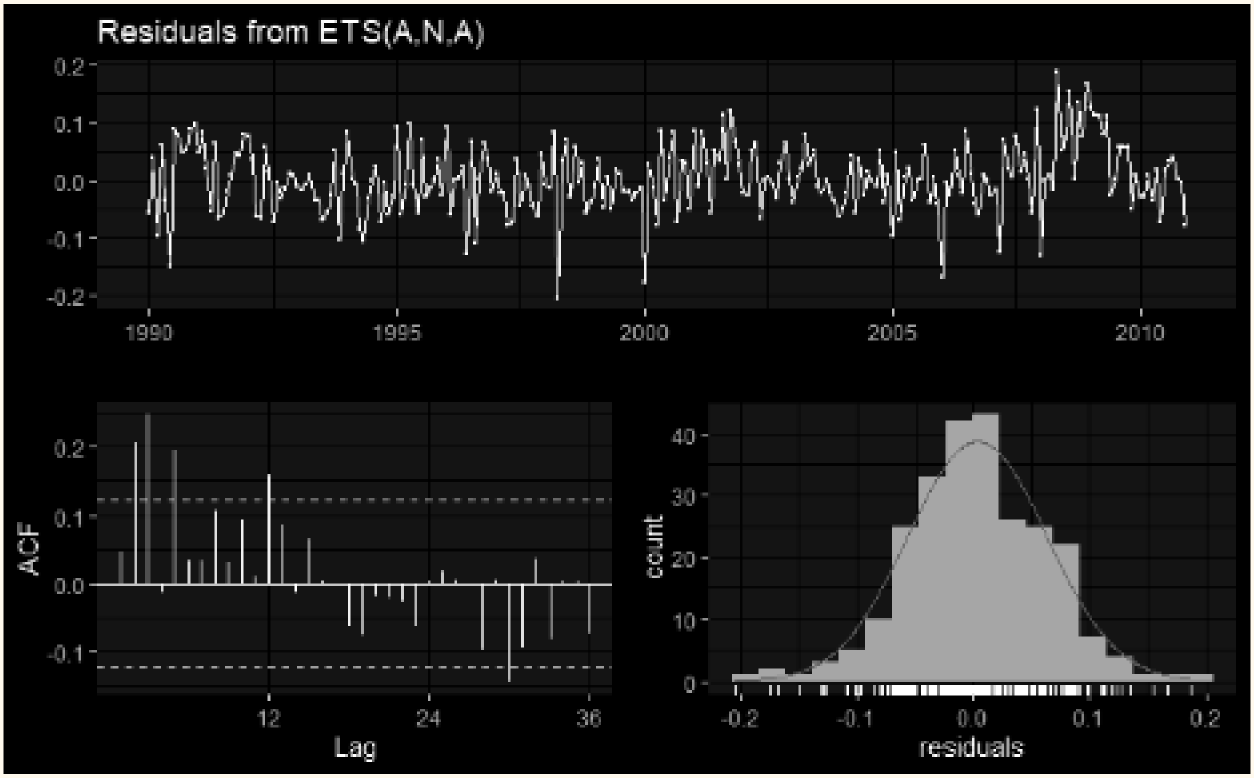


Method	RMSE
Mean	1.395
Naive	0.485
Seasonal Naive	0.884
Drift	0.484

Next we will try both ETS and Arima models. To compare these models, we'll divide our Box-Cox Transformed data into train and test sets. For training, we're using data from January 1990 to December 2010, while for testing, we are using data from January 2011 to December 2016.

ETS Forecast

We used ets() on the training dataset as ets() can handle any combination of trend, seasonality and damping. The ets() automatically chose the **ANA** model i.e Seasonal Exponential model with additive errors for the dataset. Below are the plots generated by the ETS model.



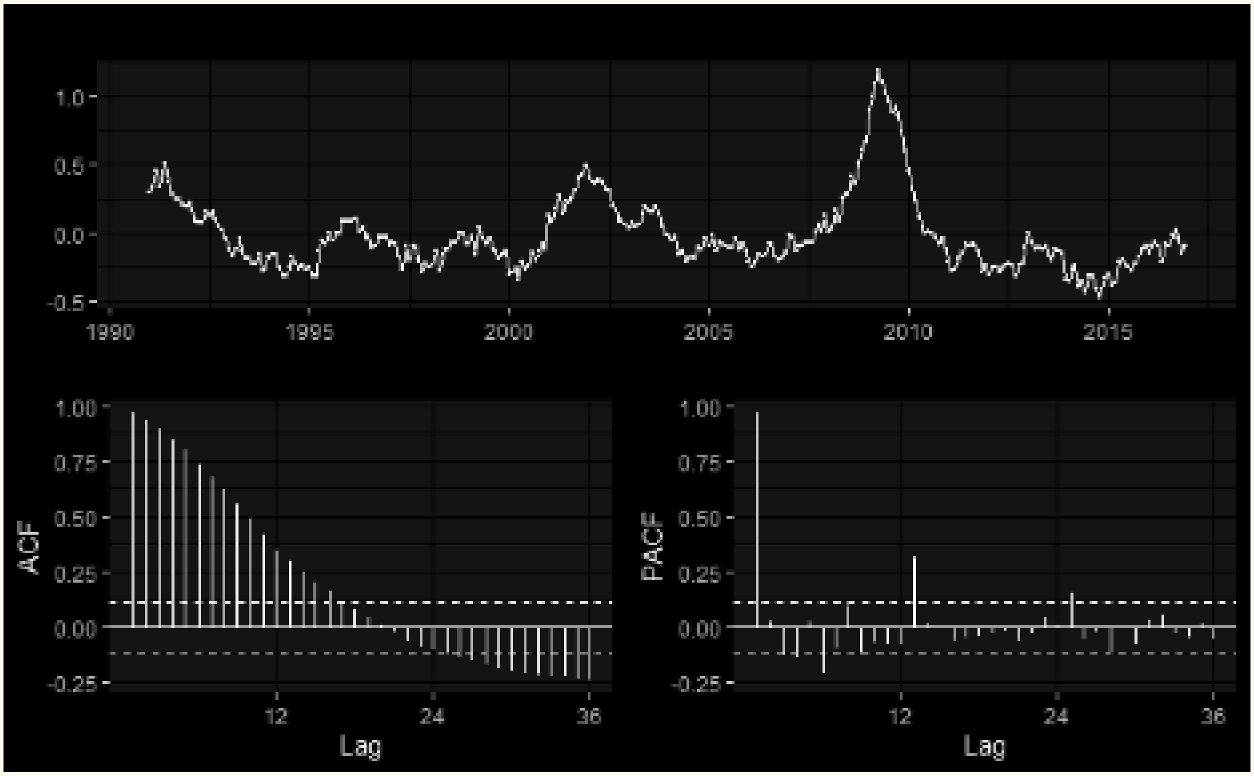
In the ACF plot, we see that the residuals do not fall within the band and there are significant spikes in the plot. Also, the residuals distribution plot seems to be right skewed. This suggests that residuals are correlated and thus forecasts from the ETS model will be biased.

SARIMA Forecast

Next, we will use ARIMA models to forecast the US Unemployment rate. ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data. We will compare the results of the ETS model from the ARIMA model to decide which method is the best fit for the forecasting.

To apply the ARIMA model, the data needs to be stationary. To make data stationary, we introduce seasonal differencing as suggested by *nsdiffs* and *KPSS Test*. Below is the time series plot after applying the lag of 12 to the original data and results of *KPSS Test.*

	t-stat	10pct	5pct	2.5pct	1pct
Critical values	0.2269	0.347	0.463	0.574	0.739



ETS Vs ARIMA On Seasonal Training and Test Data

		RMSE
ETS	**Train**	**0.05960925**
	Test	**0.78705251**
ARIMA	**Train**	**0.05381369**
	Test	**0.28026849**

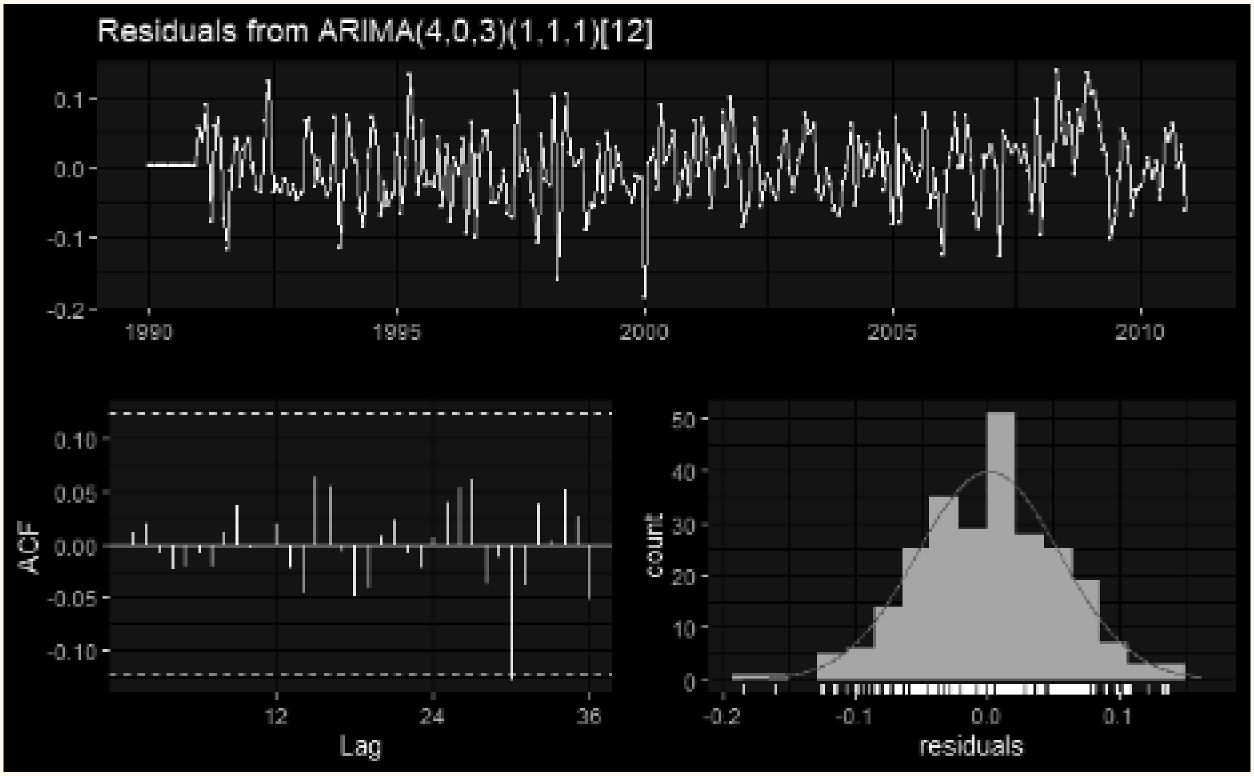
Next, we used the stationary data set with auto.arima() to find the appropriate p,d,q values.

Comparison of ARIMA Models

We constructed three different ARIMA models to find the best model to forecast future unemployment rate. In the first model, we used the p,d,q prescribed by auto.arima() and the other two models' p,d,q was decided by the ACF and PACF graph above.

After comparing the AICC of the three models, the best model is the SARIMA model with parameters **(4,0,3)(1,1,1)**.

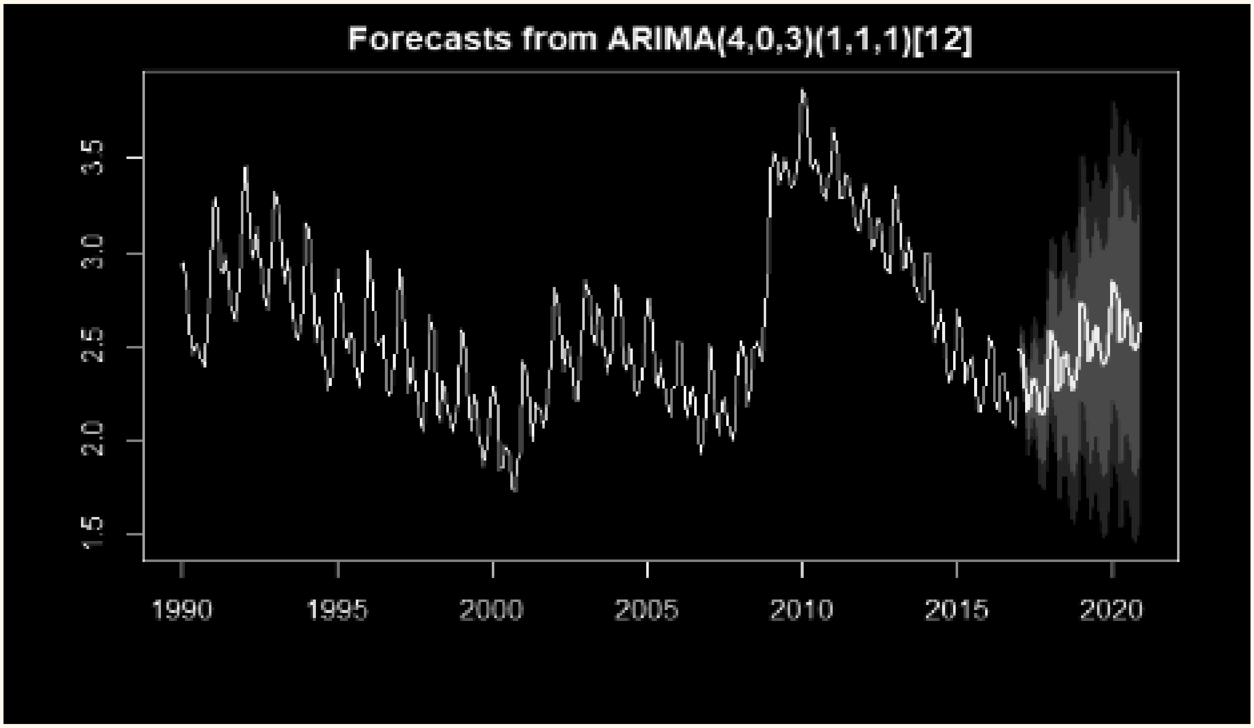
Below is the plot of the residual of the best SARIMA model -



ARIMA Model Parameters	AICc
(4,0,3)(1,1,1)	**-674.4417**
(4,0,2)(1,1,1)	**-672.0945**
(4,0,3)(0,1,1)	**-674.429**

Forecasting for next four years

We use the best ARIMA model (4,0,3)(1,1,1) for forecasting for next four years:



Forecasted values:

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Jan 2017	2.485121	2.412398	2.557845	2.373900	2.596343
Feb 2017	2.468600	2.369736	2.567465	2.317400	2.619801
Mar 2017	2.397897	2.273993	2.521801	2.208402	2.587392
Apr 2017	2.154973	2.005560	2.304385	1.926466	2.383479
May 2017	2.173243	2.003853	2.342632	1.914184	2.432302
Jun 2017	2.314399	2.123141	2.505658	2.021895	2.606904

Date	**Forecasted (%)**	**Actual (%)**
Jan 2017	**5.69**	**4.7**
Jan 2018	**5.97**	**4.1**
Jan 2019	**6.42**	**4**
Dec 2019	**5.92**	**3.5**

Given the historical data, we have found the best model that fits our data to forecast the future unemployment rate. The model does a good job of forecasting and is pretty close to the actual rate.

Limitation

In actuality, there are several models we could explore and include several other factors or features to model our unemployment rate and help us make better forecasts. We can use models like neural networks and also take into account the economic, political or other external factors into consideration.

Conclusion

- The time series is not stationary.
- It has a seasonal component and a cyclic component, but no major trend.
- We used nsdiffs and lagged transformation to remove the trend and seasonal components from the series and found that an ARIMA(4,0,3)(1,1,1) model is a good fit for the remainder of the series.
- We used the auto.arima function from the forecast package to find a best fit ARIMA model and used it to forecast unemployment rate for next four years.
- We made a comparison of the forecasted value and actual values obtained from the BLS website.

References

- Text book: https://otexts.com/fpp2/

- Data Source: https://www.kaggle.com/jayrav13/unemployment-by-county-us