Forecast Report

The variable I am forecasting is the Monthly Seasonally Adjusted US Unemployment Rate. This variable measures the unemployment rate of households for working people aged 16 and up.

A graph showing a line

Description automatically generated

Before getting to the methodology, there is something to mention about the dataset I am using. I've made the decision to remove the Covid datapoints from unemployment (From Feb 2020 to March 2022). The data is replaced with data generated from a random normal process with mean and variance calculated from the 12 months before Covid and 13 months after Covid concated together. My logic is that covid was an anomalous event that doesn't represent the trends of the business cycle. It could have, however, affected unemployment after the shock happened but that is hopefully covered by the two years’ worth of datapoints afterwards (2022-2024).

Overview of the methodology:

Step 1. Load in the UNRATE data and remove the covid data points.

Step 2. To know what assumptions and models would work to forecast the data, it is important to understand the structure of the data, specifically whether it is mean and variance stationary. We can do this by looking at the time series plot but to be more thorough, we should run an Augmented Dickey Fuller Test.

Step 3. Next, we use the ACF (Autocorrelation Function), PACF (Partial Autocorrelation Function) and AIC to determine which of the three models, AR (Autoregressive), MA (Moving Average), or ARMA (Autoregressive Moving Average), is the best fit for the data.

Step 4. Use the chosen model to forecast 12 time periods ahead.

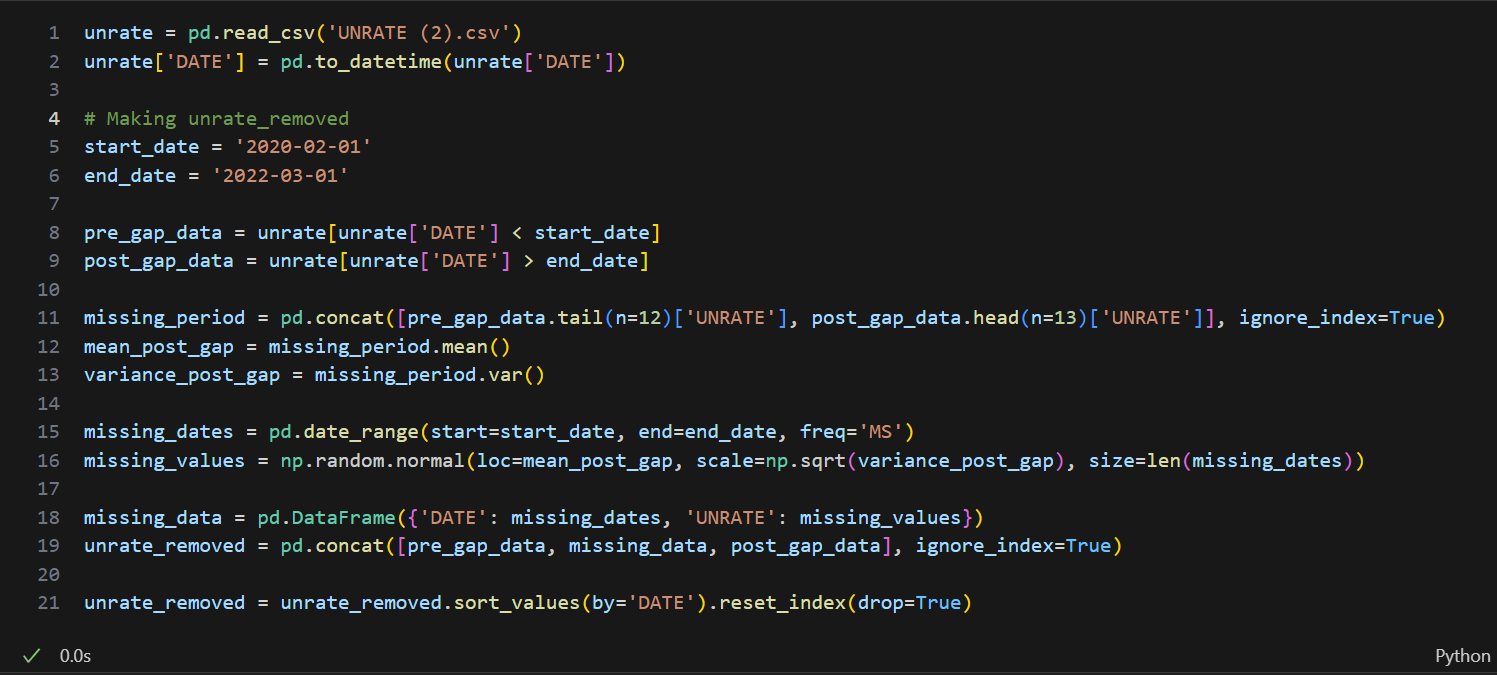
Implementing Methodology Using Code

Importing all the libraries needed (My code was written in python):

A screen shot of a computer code

Description automatically generated

Loading and cleaning the data:



When checking for stationarity I will check both the “Covid Included” and “Excluded” dataset. This is to make sure me removing the covid data didn't affect the data's stationarity.

A black screen with text

Description automatically generated

The results are as follows:

P value for the Covid INCLUDED Dataset: 0.002, and EXCLUDED Dataset: 0.038

The Null Hypothesis: The time series has a unit root (the series is non-stationary).

We can reject the null hypothesis since both p-values are below 0.05. These results mean that the series is both mean and variance stationary.

A screenshot of a graph

Description automatically generatedThe next step is to decide which model to choose for forecasting. Since we already showed the data is stationary, using ARIMA is the same as using ARMA, meaning the three models to choose from are AR, MA and ARMA. To choose which model is best we can go about it in two ways. First is checking the ACF and PACF graphs and deciding based on their shapes. The other is the more cumbersome way of checking which is done by calculating the AIC for a range of all three models. (For both the ACF and PACF, I used 24 lags to try and capture any yearly and semi long run trends).

A graph with blue dots and numbers

Description automatically generatedA graph of a graph with blue dots

Description automatically generated with medium confidence

Given the structure of the ACF, an AR(6) model seems appropriate. This is because for an AR Model, the PACF should show significant lags up to a certain point (in this case 6 lags) and then cut off into insignificance and the ACF should taper off gradually (Slow geometric decay), which is what is observed in the graphs above.

For the MA and ARMA Models:

MA Model: Used when the ACF shows significant lags up to a certain point (and then cuts off) but the PACF tapers off gradually.

ARMA Model: Used when both ACF and PACF taper off without a clear cutoff.

Neither of these is what is seen in the graphs so I don't think they are the correct models, even so we can double check by testing them using the AIC.

AR Model AIC:

A screen shot of a computer

Description automatically generated

The Model with the minimum AIC SSR is an AR(6) model: -435.042

MA Model AIC:

A screen shot of a computer program

Description automatically generated

The Model with the minimum AIC SSR is an AR(0) MA(11) model: -151.400

ARMA Model AIC:

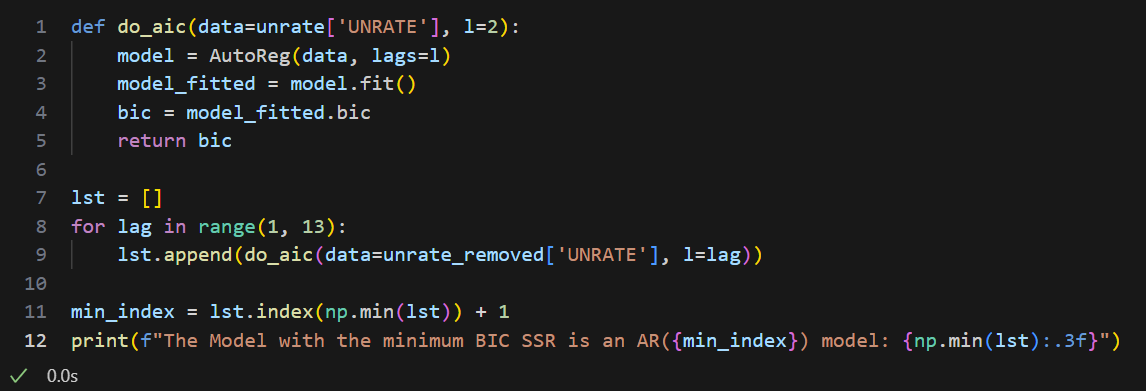
A computer screen with text on it

Description automatically generated

The Model with the minimum AIC SSR is an AR(7) MA(7) model: -449.677

It seems that an ARMA(7,7) is the better model to use here. It’s strange, especially considering the shape of the ACF and PACF indicates an AR(6). What I will do to make sure of this is calculate the BIC since this might be a case of overfitting (Notice how q, p can range from 1 to 7 in the code and its choosing ARMA(7,7), the largest values for q and p. It may be choosing as many parameters as possible to try and perfectly fit the data.). I’m using BIC since it has a stronger penalty if there are many parameters. It favors simpler models unless the improvement in fit is substantial.

AR Model BIC:



The Model with the minimum BIC SSR is an AR(6) model: -396.482

ARMA Model BIC:

A computer screen shot of a program

Description automatically generated

The Model with the minimum BIC SSR is an AR(1) MA(1) model: -151.400

Here is the full data:

|  |  |  |
| --- | --- | --- |
| Model | AIC | BIC |
| AR | -435.042 | -396.482 |
| ARMA | -449.677 | -151.400 |

With respect to AIC, the ARMA model seems to barely edge out the AR model. However, with respect to the BIC, The AR model is absolutely the better option and given how much worse ARMA got when we considered BIC, I think its plausible to call this a case of overfitting.

A graph showing the number of coronavirus

Description automatically generatedA computer screen shot of text

Description automatically generatedMoving forward with the AR(6) Model, we code up and forecast 12 time periods to the future:

The forecasts the AR(6) model produced:

|  |  |  |  |
| --- | --- | --- | --- |
| **DATE** | **UNRATE** | **CI\_lower** | **CI\_upper** |
| 2024-11-01 | 4.118563 | 3.747599 | 4.489528 |
| 2024-12-01 | 4.139862 | 3.615239 | 4.664484 |
| 2025-01-01 | 4.144824 | 3.502295 | 4.787352 |
| 2025-02-01 | 4.163935 | 3.422006 | 4.905863 |
| 2025-03-01 | 4.193444 | 3.363943 | 5.022945 |
| 2025-04-01 | 4.225394 | 3.316721 | 5.134066 |
| 2025-05-01 | 4.260069 | 3.278590 | 5.241548 |
| 2025-06-01 | 4.295893 | 3.246648 | 5.345138 |
| 2025-07-01 | 4.334941 | 3.222049 | 5.447834 |
| 2025-08-01 | 4.375552 | 3.202460 | 5.548643 |
| 2025-09-01 | 4.416989 | 3.186640 | 5.647337 |
| 2025-10-01 | 4.459086 | 3.174029 | 5.744143 |

So, the calculated forecast for UNRATE in November 2024 is ~ **4.1**%. Comparing this forecast to news releases discussing unemployment rate there’s Yahoo Finance: *“November jobs report expected to show hiring rebound, unemployment hold steady at* ***4.1%****”*, Morning Star: *“unemployment rate is expected to hold steady at* ***4.1%****”*, CNN: “The unemployment rate — which has served as an unofficial polestar amid the wild distortions — is expected to remain at **4.1%**” and Peterson institute for international economics: *“The US unemployment rate is projected to hover* ***just above 4*** *percent through the end of next year.”*. It seems that the forecast agrees with the general consensus which is great.

One last thing I’d like to do is to run an ADL forecast with Initial Unemployment Claims. The variable has a strong correlation with unemployment, acting like a predictor, meaning if it goes up so will the unemployment rate after a slight lag. Running this ADL can help reaffirm the forecast that was achieved from the AR(6) Model.

A screen shot of a computer program

Description automatically generatedFirst, I loaded in the data and cleaned it up. I removed the COVID data points and made it monthly by considering the first value in the month as its value.

Then I found the ADL model with the lowest AIC and BIC (I prioritized a low BIC so that overfitting isn’t an issue). I got these parameters, AIC and BIC:

Best model combination: Lag=7, Order=4, Lowest BIC: -637.367, Lowest AIC: -700.820

I then used it to forecast the data:

A screen shot of a computer code

Description automatically generated

A screen shot of a computer program

Description automatically generatedHere is the forecast code and results:

A computer code on a black background

Description automatically generated

A graph showing the growth of the stock market

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| DATE | UNRATE | CI\_lower | CI\_upper |
| 2024-11-01 | 4.125685 | 3.846450 | 4.404920 |
| 2024-12-01 | 4.074182 | 3.679284 | 4.469079 |
| 2025-01-01 | 3.976416 | 3.492767 | 4.460064 |
| 2025-02-01 | 3.904840 | 3.346371 | 4.463310 |
| 2025-03-01 | 3.889468 | 3.265080 | 4.513856 |
| 2025-04-01 | 3.874645 | 3.190662 | 4.558628 |
| 2025-05-01 | 3.894763 | 3.155977 | 4.633549 |
| 2025-06-01 | 3.890447 | 3.100651 | 4.680242 |
| 2025-07-01 | 3.874526 | 3.036822 | 4.712230 |
| 2025-08-01 | 3.879950 | 2.996932 | 4.762968 |
| 2025-09-01 | 3.879065 | 2.952948 | 4.805182 |
| 2025-10-01 | 3.942086 | 2.974789 | 4.909384 |

Looking at the results of the ADL forecast, something interesting happened. The point forecast for November 2024 is similar to the AR(6) model’s and the general consensus at 4.1%. However, the drift of the forecasts are completely different. The AR(6) model is pessimistic and says that UNRATE will increase next year but the ADL says it will decrease next year.

A graph showing the number of coronavirus

Description automatically generatedA graph showing the growth of the stock market

Description automatically generated