Auto Regressive Model (AR)

The most basic of all the time series models we start with is the Auto Regressive Model (AR), a great way to predict stationary datasets. As the name suggests, the model will attempt to do a regression based on past observations from a time series. The difference here is that in a multivariate regression model we use a linear combination of predictors to forecast, while in AR we use a linear combination of past values of the variable we are trying to predict – hence the term “auto” to indicate we are doing a regression on the variable itself. The model for AR is expressed as [1] where p is the order.

.

The Partial Auto Correlation Function (PACF) helps us determine what lags have a significant positive or negative correlation impact to the model and therefore should be included in the model build [2].

Below are the final AR expressions created:

|  |  |
| --- | --- |
| Feature | Final Equation |
| positive |  |
|  |  |

Below are the final order and accuracy of the AR models:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Model | MAPE | RMSE |
| positive | AR (2,0) | 0.01863278742046370 | 9329.592789722843 |
|  |  |  |  |

Moving Average Model (MA)

The MA model is another simple time series model that predicts a pattern for a variable centralized on its mean [3]. The model for MA is therefore expressed as [1] where q is the order.

The Auto Correlation Function (ACF) is helpful in determining if a dataset is a good candidate for the moving average model if it is determined that after some lag (q) the correlation is determined to be 0 [4].

Below are the final MA expressions created:

|  |  |
| --- | --- |
| Feature | Final Equation |
| positive |  |
|  |  |

Below are the final order and accuracy of the MA models:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Model | MAPE | RMSE |
| positive | MA (0,1) | 0.4265513028681226 | 167536.96708900554 |
|  |  |  |  |

Auto Regressive Moving Average Model (ARMA)

The ARMA model is another simple time series model that is combination of the AR and MA model, and best suited for stationary data [5]. The model for ARMA is expressed as [1], where is a white noise process [1].

As in the stationarity and invertibility conditions, the ACF and PACF of an ARMA process are determined by the AR and MA components, respectively [1].

In ARMA we consider the following:

* p: Trend autoregression order (this is the AR order)
* q: Trend moving average order (this is the MA order)

For positive feature, data is stationary (d = 0). ARMA and ARIMA produces same result.

Auto Regressive Integrated Moving Average Model (ARIMA)

The ARIMA model is a good choice for data that has a moving mean or in other words, where data is non-stationary [6]. We say it is integrated because we are predicting the differences in one time stamp to a previous one. We do this in hopes of identifying a constant, which will be used as a mean to make the data stationary. Once an appropriate time series model has been fit, it may be used to generate forecasts of future observations. The expression for ARIMA is as follows where is the lag operator.

In ARIMA we consider the following:

* p: Trend autoregression order (this is the AR order)
* d: Trend difference order (this is the Integrated order)
* q: Trend moving average order (this is the MA order)

As in the stationarity, the ACF and PACF of an ARIMA process are determined by the AR and MA components, respectively [1].

Below are the final ARIMA expressions created:

|  |  |
| --- | --- |
| Feature | Final Equation |
| positive |  |
|  |  |

Below are the final order and accuracy of the ARIMA models:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Model | MAPE | RMSE |
| positive | ARIMA (1,0,2) | 0.013209238012813353 | 6039.018610732801 |
|  |  |  |  |

Seasonal Autoregressive Integrated Moving Average Model (SARIMA)

The SARIMA model is a variation of the ARIMA model that considers a timeseries that has seasonality. In ARIMA we consider the following:

However, with SARIMA the updated expression becomes:

In the equation above we add to the ARIMA model a seasonality component. The first 3 arguments of this component represent the same arguments from the first tuple - to recap these values are:

1. The number of auto-regressive lags
2. The number of time steps to take the difference over.
3. The number of moving average lagged errors.

But to add seasonality to these values we have added “m” to our equation, this represents the length of time for the observed seasonality. For example, if we were to see a spike in our data over a 12-month period, we would set the value to 12 since each time step in our data is 1 month long. Without differencing operations, the equation for SARIMA can be expressed as

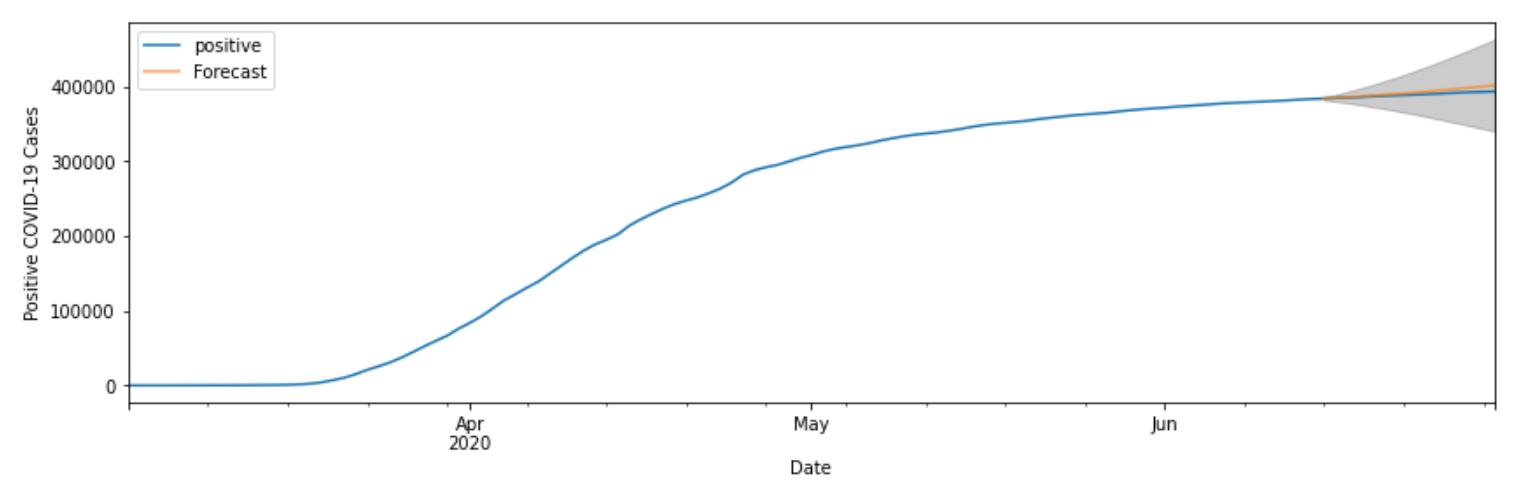
Below are the final SARIMA expressions created:

|  |  |
| --- | --- |
| Feature | Final Equation |
| positive |  |
|  |  |

Below are the final order and accuracy of the SARIMA models:

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Model | MAPE | RMSE |
| positive | SARIMA (1,1,1) (1,1,1,3)3 | 0.00042583177811003625 | 212.37937759826298 |
|  |  |  |  |

The SARIMA model had one of the best prediction accuracies showing superior error values among the rest. Below is a graph of its forecast for positive COVID-19 cases.



Simple Exponential Smoothing (SES) model

Single Exponential Smoothing, SES , also known as Simple Exponential Smoothing, is a time series forecasting method for data without a trend or seasonality. It requires a single parameter, called alpha, also called the smoothing factor. This parameter controls the rate at which the influence of the observations at prior time steps decay exponentially. Alpha is often set to a value between 0 and 1. Large values mean that the model is influenced mostly by the most recent past observations, whereas smaller values mean more of the history is considered when making a prediction.

=

|  |  |
| --- | --- |
| Alpha(α) | RMSE |
| 0.2 | 5548.2775 |
| 0.6 | 4608.6878 |
| 0.8 | 4328.8105 |

Holt Winters Forecasting Model

The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level, one for the trend and one for the seasonal component. There are two variations to this method that differ in the nature of the seasonal component. Since we are considering data for initial 120 days, we do not have a seasonal factor at play here. Therefore, we will be considering only smoothing factor as well as trend.

Where α(0≤α≤1) is the data smoothing factor and β(0≤β≤1) is the trend smoothing factor

|  |  |  |  |
| --- | --- | --- | --- |
| Alpha | Beta | RMSE | MAPE |
| 0.2 | 0.2 | 288.1078 | 0.3798 |
| 0.6 | 0.2 | 472.62 | 0.6687 |
| 0.8 | 0.2 | 328.20 | 0.4324 |

**References**

[1] Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci, “Introduction to Time Series Analysis and Forecasting”, New Jersey: John Wiley & Sons, 2008.

[2] ritvikmath, “Time Series Talk : Autoregressive Model”, *YouTube*, Apr. 11, 2019 [Video file]. Available: https://www.youtube.com/watch?v=5-2C4eO4cPQ. [Accessed: Mar. 20, 2021]

[3] ritvikmath, “Time Series Talk : Moving Average Model”, *YouTube*, Apr. 22, 2019 [Video file]. Available: https://www.youtube.com/watch?v=voryLhxiPzE. [Accessed: Mar. 20, 2021]

[4] ritvikmath, “Time Series Talk : Moving Average and ACF”, *YouTube*, Apr. 22, 2019 [Video file]. Available: https://www.youtube.com/watch?v=\_tgB-ri9-8c [Accessed: Mar. 20, 2021]

[5] ritvikmath, “Time Series Talk : Autoregressive Moving average Model”, *YouTube*, Jun. 25, 2019 [Video file]. Available: https://www.youtube.com/watch?v=HhvTlaN06AM [Accessed: Mar. 20, 2021]

[6] ritvikmath, “Time Series Talk : ARIMA Model”, *YouTube*, Jul. 11, 2019 [Video file]. Available: https://www.youtube.com/watch?v=3UmyHed0iYE [Accessed: Mar. 20, 2021]

[7] https://en.wikipedia.org/wiki/Exponential\_smoothing