1.

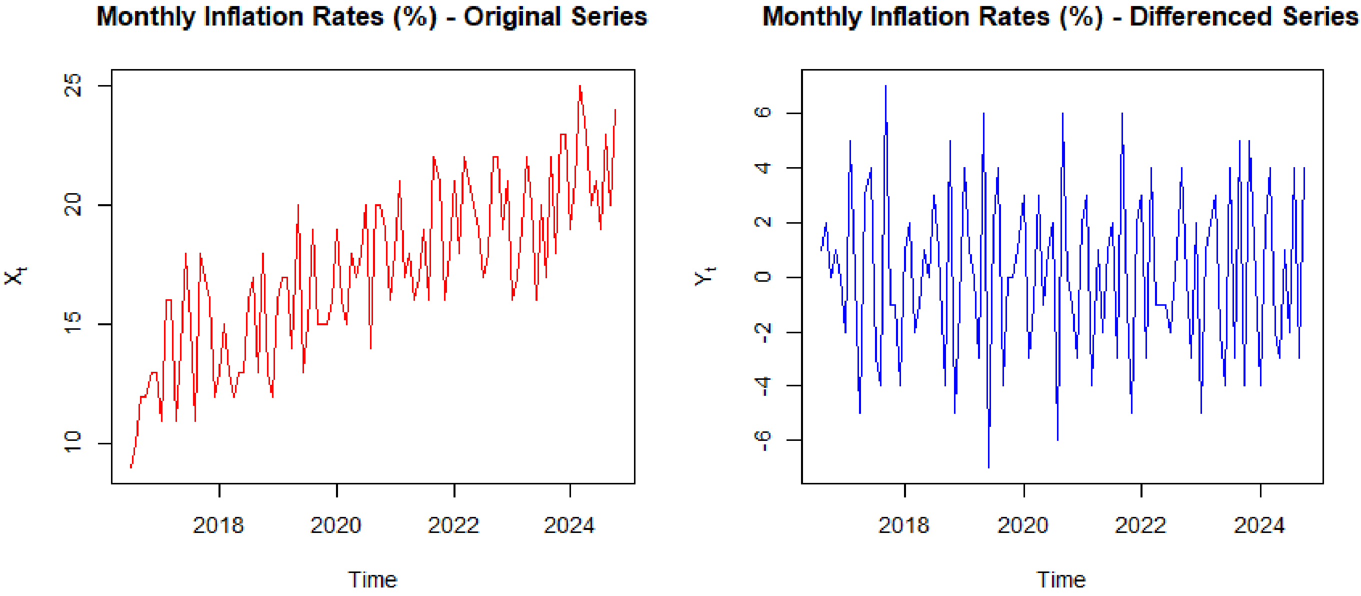

Figure 1: Plot of original inflation rate data (left) and first-differenced data (right)

In Figure 1, we can see that there is an obvious upward trend over time in the original data, indicating non-stationarity. For a time series to be stationary, there should be a constant mean and variance. However, due to the upward trend, it seems that the mean is changing over time. On the other hand, after taking first differences, this is visually much more stationary. The residual differences are centred quite closely around 0, with the range of values being seemingly lying between -7 and 7. This implies a significant reduction in trend in this graph and will allow for more accurate time series analysis.

2.

In the ACF of the original data, as shown in Figure 2, we see that there is a gradual decline. This is indicative of a continued strong autocorrelation across the time lags, which backs up the earlier claim that this is not a stationary time series. It also continues to have lag ACF values outside the significant error threshold, which again is usually an indication of non-stationarity. For the PACF, we see some large spikes at the beginning, which quickly tails off to being consistently within the significance threshold. This is fitting for a standard ACF plot for choosing an AR(p) model, which would likely be an AR(2) model based on this plot with lags starting at 0.

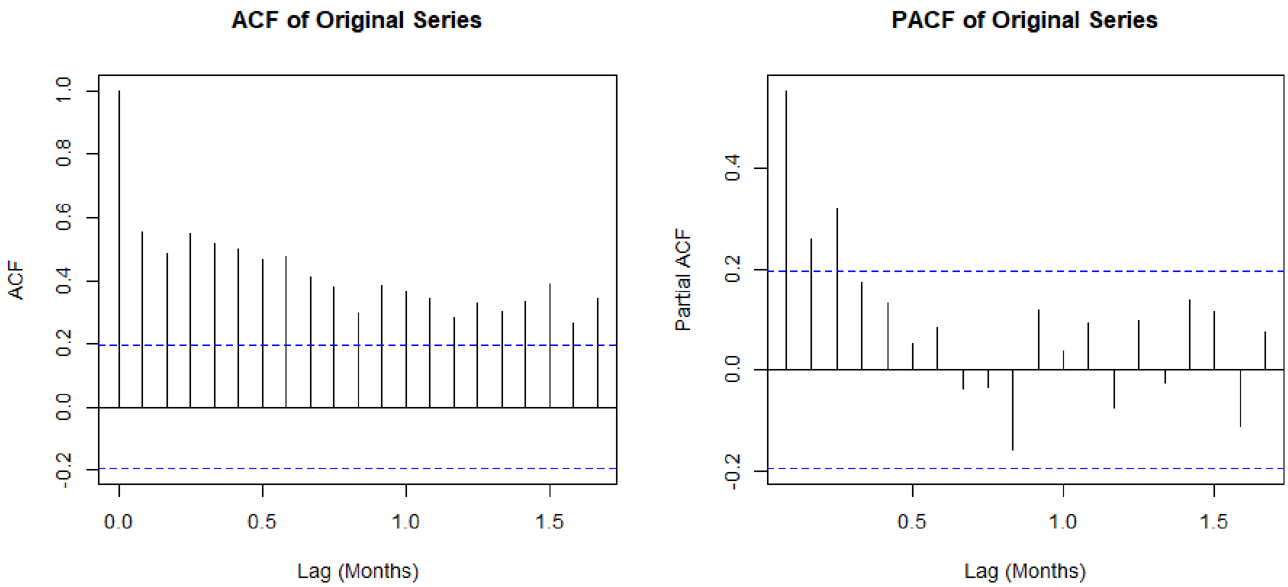

Figure 2: ACF (left) and PACF (right) of original data

The ACF of the differenced data has two of larger spikes at lags 0 and 1 respectively, rapidly centring much closer to 0 after that. This implies stationarity, which corresponds to the understanding gained from the line graph of the differenced data. This would also justify the differencing operation on this time series data. There are, however, some spikes at lags 10, 18, 19 and 20 which are just outside of the error threshold, which could make picking the MA(q) model less trivial than perhaps thought. As for the PACF of the differenced data, after lag 3, all further lag spikes are within the desired threshold, indicating that a AR(3) process could be a justifiable way to model this time series.

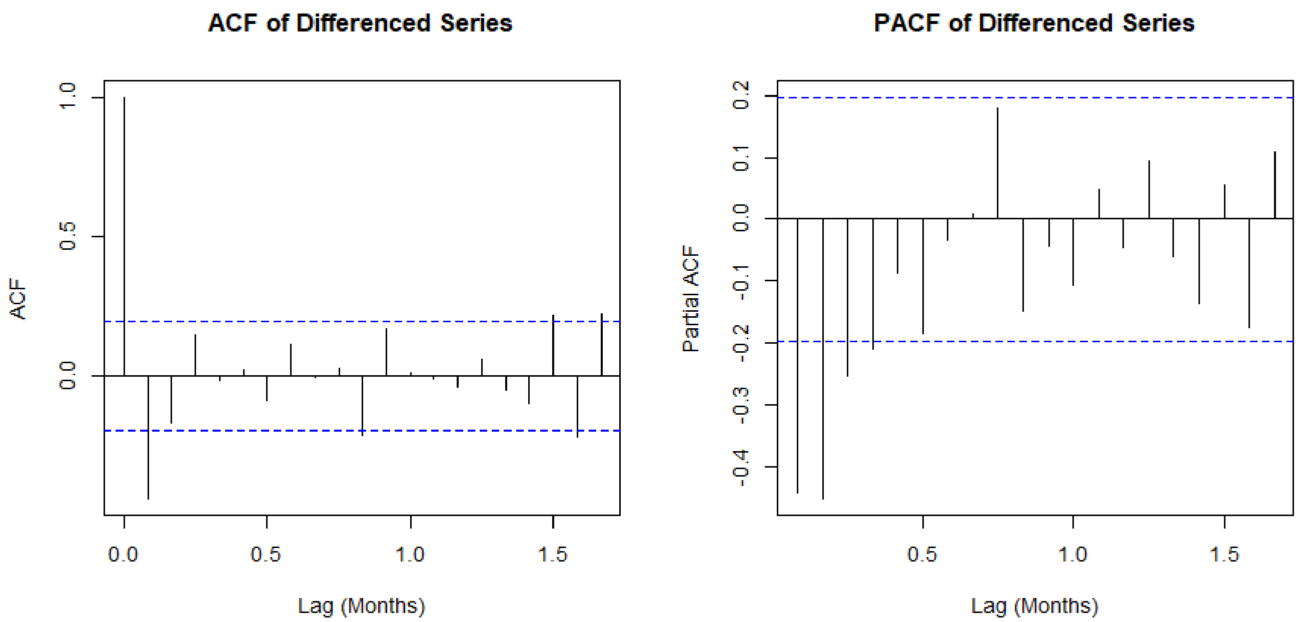

Figure 3: ACF (left) and PACF (right) of differenced data

3.

Regarding picking an ARIMA model for the differenced data, there are a couple of things to consider. As stated previously, the PACF indicates that an AR(p) component with p=3 would be reasonable for this as lag three is the last lag on the graph that does not lie within the threshold. The MA(q) component, however, appears to be more nuanced in this case as it appears that there may be a few lags after the initial spikes that marginally violate the error threshold. Since the these violation are very, very small, the best course of action would likely to be to model this time series data with an ARIMA(1, 1, 3), but with the idea that – if these points of violation become an issue further down the line, the AR(p) component may need to be adjusted.

4.

Firstly, after fitting the ARIMA(1, 1, 3) model to the data, we get the coefficients in Figure 4:

```
       ar1        ma1        ma2        ma3
-0.1296751 -0.7468822 -0.2324281  0.2776159
```

Figure 4: Estimated coefficients of ARIMA(1, 1, 3) model

Which, when rounded to three decimal places, gives the coefficients $\phi_1 = -0.130, \theta_1 = -0.747, \theta_2 = -0.232, \theta_3 = 0.278$.

When taking this into account, we obtain the following fitted model equation:

$$(1 + 0.130B)(1 - B)X_t = (1 - 0.747B - 0.232B^2 + 0.278B^3)\epsilon_t$$

Where the estimated mean is 0 for what is now assumed to be a stationary function.

The AIC value for this model is 457.11.

5.

The best way to determine what model is better between the ARIMA(1, 1, 3) and the ARIMA(2, 1, 5) are made by comparing the model complexities, AIC and BIC. As far as complexity goes, the colleague's suggestion is far more complex, potentially leading to overfitting. The (AIC, BIC) results for the ARIMA(1, 3, 3) and ARIMA(2, 1, 5) models respectively are (457.11, 470.09) and (461.48, 482.24). Since lower values for AIC and BIC indicate a better fit as well as being harsh on complexity, we can say that the ARIMA(1, 1, 3) model is better, both in terms of complexity and fir.

6.

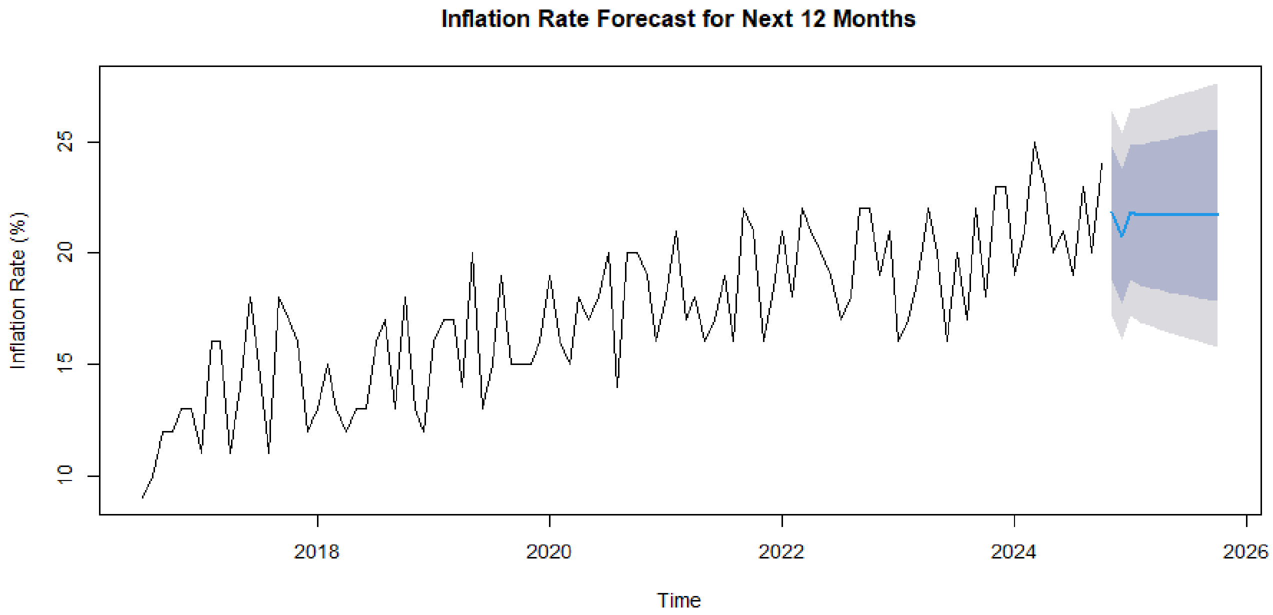

Figure 5: Forecasted inflation over the next 12 months, based on ARIMA(1, 1, 3) model

The forecasted mean in Figure 5 shows somewhat of a levelling of inflation rates over the next 12 months. It can not be concluded that it will not continue in an upward trend in general, despite this, as historically, the data appears to increase even after a period of stability. The 95% interval here appears to be quite wide which means that it is hard to make concrete conclusion regarding future trend based on this model alone. The AIC being rather high suggests that perhaps, with some fine tuning, it may be possible to create a model that has a more concrete conclusion.