Written Report – 6.419x Homework Module 4

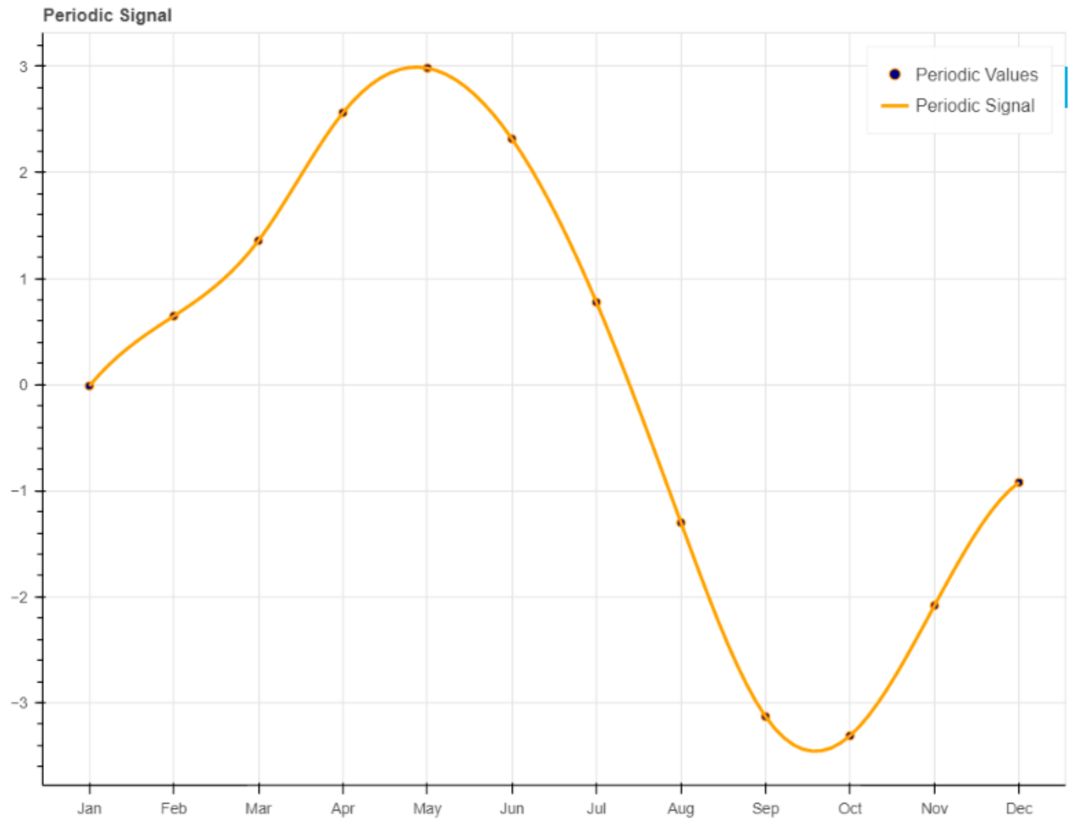
The Mauna Loa CO_2 Concentration

The Final Model

1. *(3 points) Plot the periodic signal P_i. (Your plot should have 1 data point for each month, so 12 in total.) Clearly state the definition of the P_i, and make sure your plot is clearly labeled.*

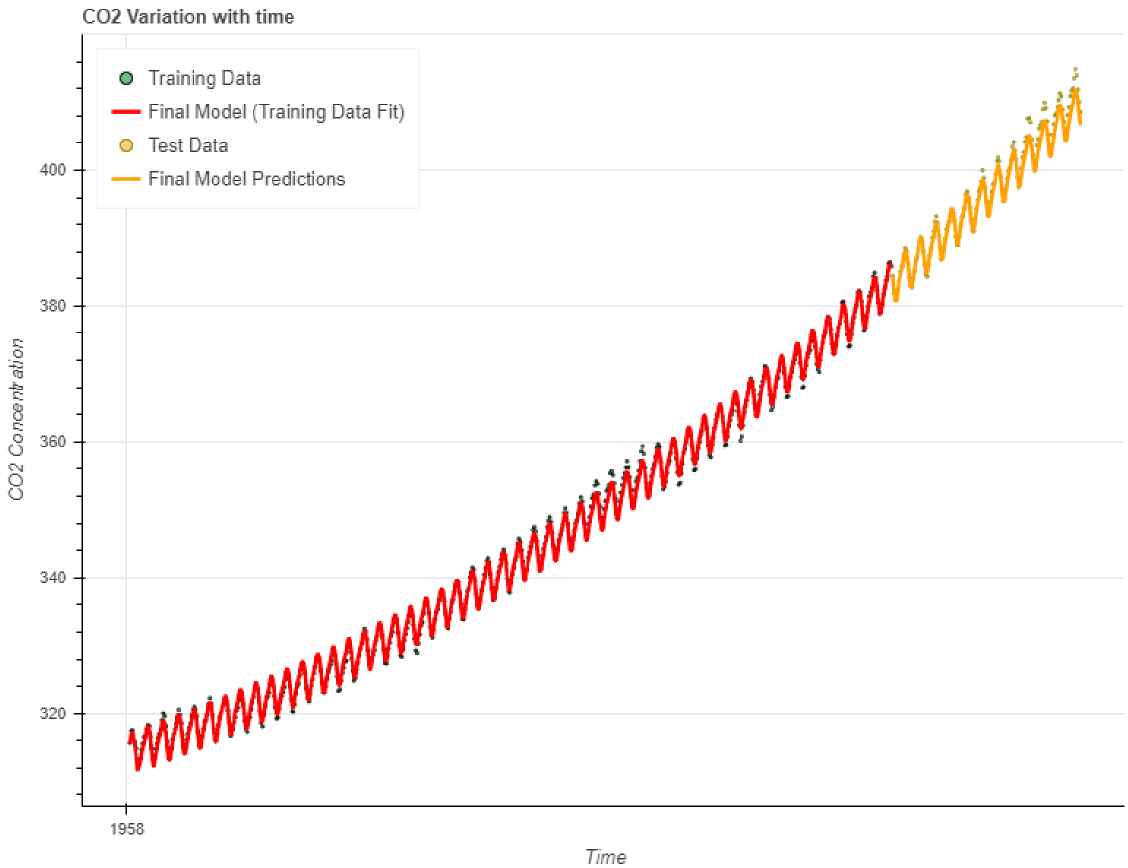
```
# Interpolating the months with the 1D interpolate function to get a continuous signal

periodic_signal = scipy.interpolate.interp1d(monthly_residuals.index, monthly_residuals.values, kind='cubic')
```



The figure above shows the predicted periodic signal based on the training data. The minimum signal was in October and the maximum in May.

2. *(2 points) Plot the final fit $F_n(t_i) + P_i$ Your plot should clearly show the final model on top of the entire time series, while indicating the split between the training and testing data.*



The figure shows the history of the increase of CO2 concentrations measured at Mauna Loa, Hawaii since 1958, and the predicted trajectory afterwards.

3. *(4 points) Report the root mean squared prediction error RMSE and the mean absolute percentage error MAPE with respect to the test set for this final model. Is this an improvement over the previous model $F_n(t_i)$ without the periodic signal? (Maximum 200 words.)*

```
The mean squared prediction error of the final model is 1.1493602690795508
The mean squared prediction error of the quadratic model is 2.5013322194898326
The mean absolute percentage error of the quadratic model is 0.5320319129740952
The mean absolute percentage error of the final model is 0.20859165947993008
```

From the results above, the prediction errors have decreased by adding the periodic signal component to the final model showing significant improvement over the previous quadratic model alone.

4. *(3 points) What is the ratio of the range of values of F to the amplitude of P_i and the ratio of the amplitude of P_i to the range of the residual R_i (from removing both the trend and the periodic signal)? Is this decomposition of the variation of the CO_2 concentration meaningful? (Maximum 200 words.)*

```
Amplitude of Trend :
69.14369234539129

Amplitude of Periodic Signal :
6.2924106671596185

Amplitude of Residuals :
3.836399237579493

Ratio of Amplitudes of Trend(F) to Periodic Signal(Pi) :
10.98842653519984

Ratio of amplitudes of Periodic Signal(Pi) to Residuals(Ri) :
1.6401866118422288
```

The decomposition of the variation of the CO2 concentration is meaningful given that the range of the trend is much larger compared to that of the periodic signal and the amplitude of the periodic signal is also larger compared to the range of the residuals.

Autocovariance Functions

5. *(4 points) Consider the MA (1) model, $X_t = W_t + \theta W_{t-1}$, where W_t ~W~ $N(0, \sigma^2)$. Find the autocovariance function of X_t. Include all important steps of your computations in your report.*

$\gamma(1) = E(W_t W_{t-1}) + \theta E(W_t W_{t-2}) + \theta E(W_{t-1}^2) + \theta^2 E(W_{t-1} W_{t-2})$

$= 0 + 0 + \theta var(W_{t-1}) + \theta E(W_{t-1}^2) + 0$

$= \sigma^2 \theta$

Since $E[W_{t-i} W_{t-j}] = E[W_{t-i}] E[W_{t-j}] = 0$ where $i \neq j$

6. *(4 points) Consider the AR (1) model, $X_t = \phi X_{t-1} + W_t$, where $\{W_t\}$ ~W~ $N(0, \sigma^2)$.*

$\gamma(1) = E[W_t + \phi X_{t-1})(X_{t-1})]$

$= E[W_t X_{t-1} + \phi X^2_{t-1}]$

$= \phi \sigma^2_{AR}$

$= \dfrac{\phi\sigma^2}{(1 - \phi^2)}$

Since $E[W_t W_s] = \dfrac{\sigma^2}{(1 - \phi^2)}$

when s = t and 0 otherwise for white noise.

Converting to Inflation Rates

1. *Repeat the model fitting and evaluation procedure from the previous page for the monthly inflation rate computed from CPI.*

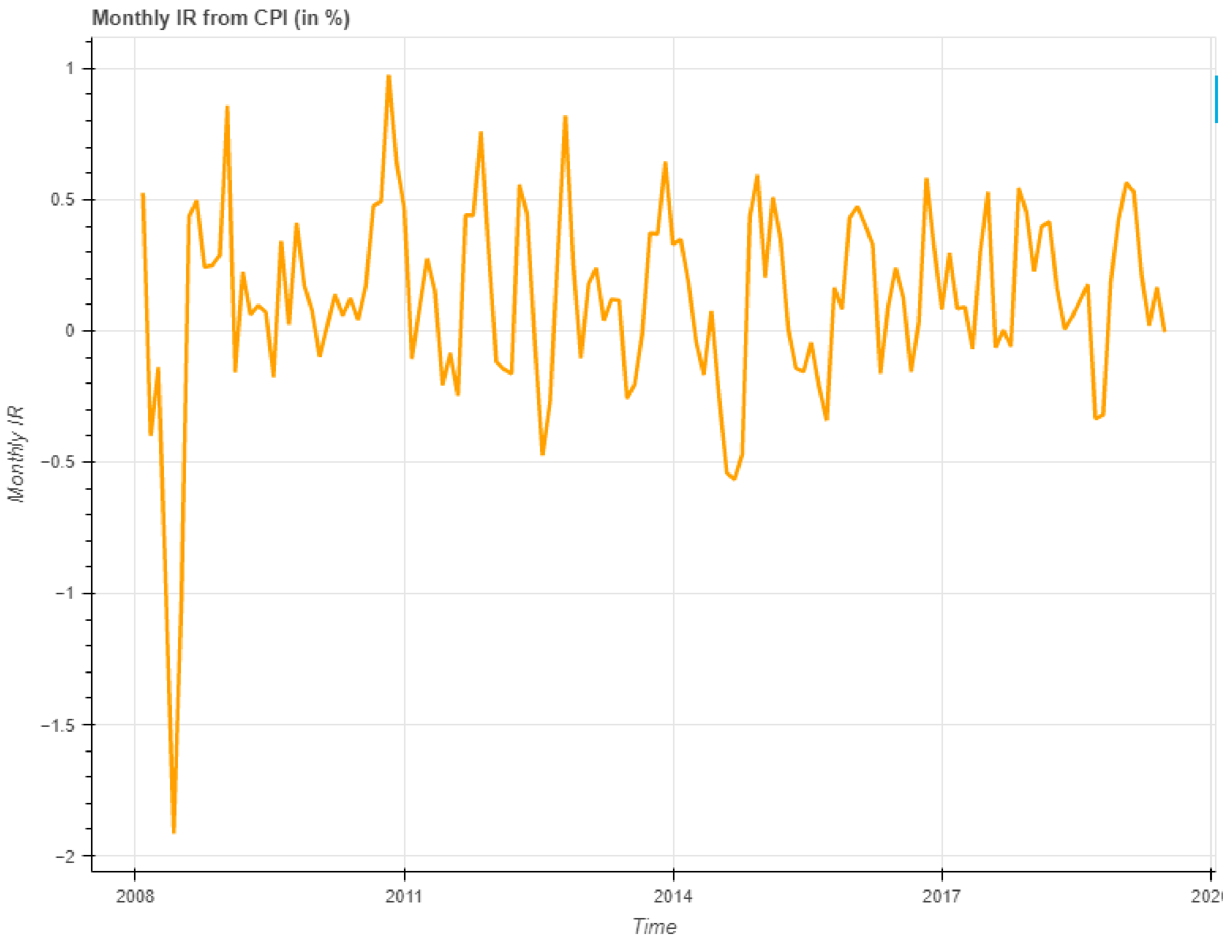
   Your response should include:

   (1 point) Description of how you compute the monthly inflation rate from CPI and a plot of the monthly inflation rate. (You may choose to work with log of the CPI.)

   The monthly inflation rate (IR) was calculated from the CPI column and added as a separate column. The formula below was used –

   Inflation Rate (IR_t) = [(CPI_t – CPI_{t-1}) / CPI_{t-1}]* 100

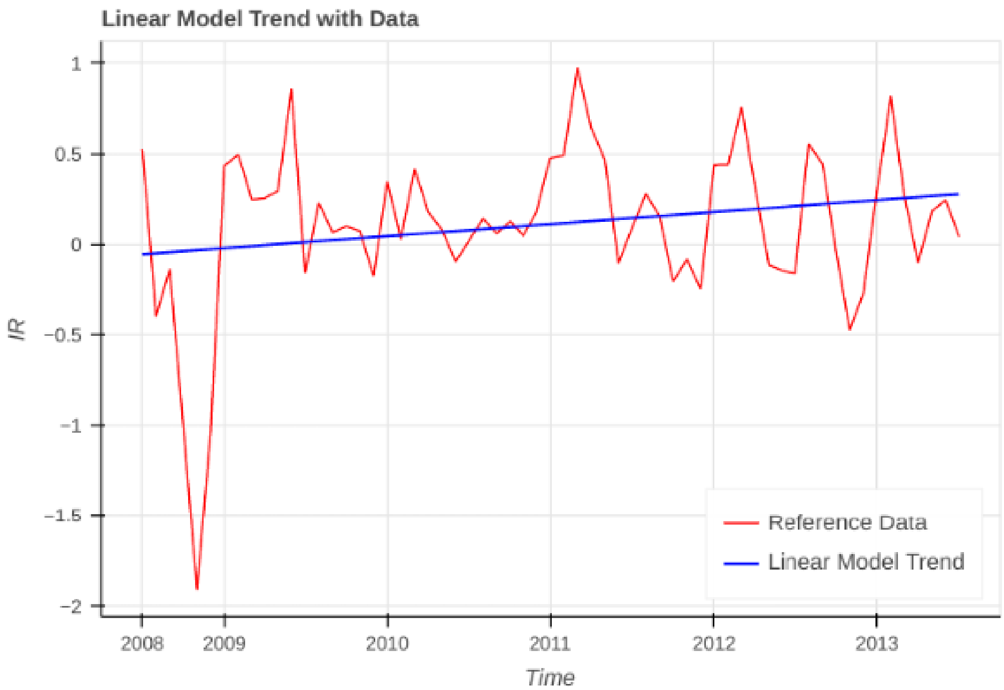
   The monthly data has been represented in the chart below

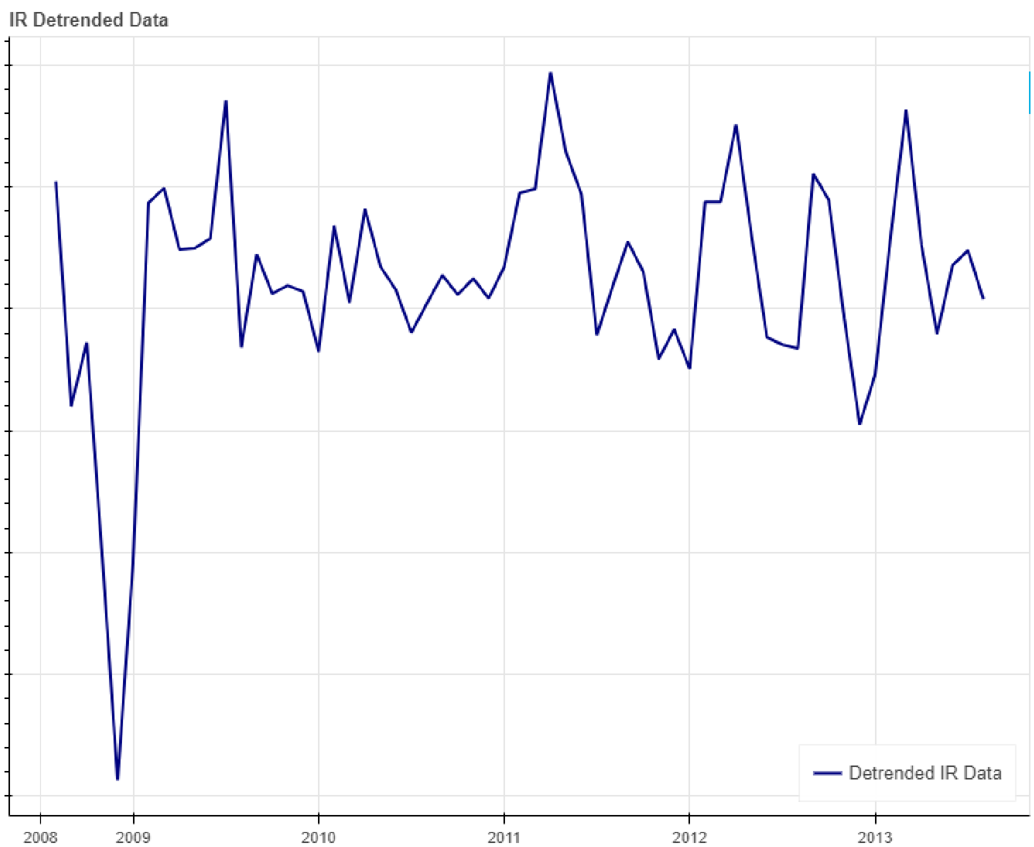


   (2 points) Description of how the data has been detrended and a plot of the detrended data.

   To detrend the data, the dataset was split into training and testing as at September 2013. The training data was then used to train a linear regression model which resulted in a coefficient of 0.0054937024691356084 and an intercept of – 0.0565621354823323.

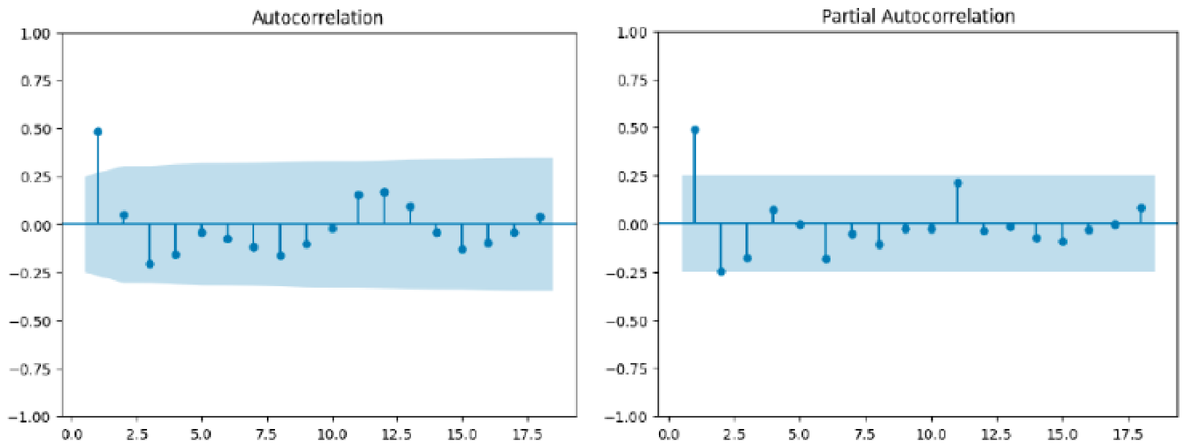
   The predicted value of the training data was then subtracted from original data to obtain the detrended data and has been graphed below.


Linear Model Trend with Data
1
0.5
0
−0.5
−1
−1.5
−2
IR
2008
2009
2010
2011
2012
2013
Time
Reference Data
Linear Model Trend

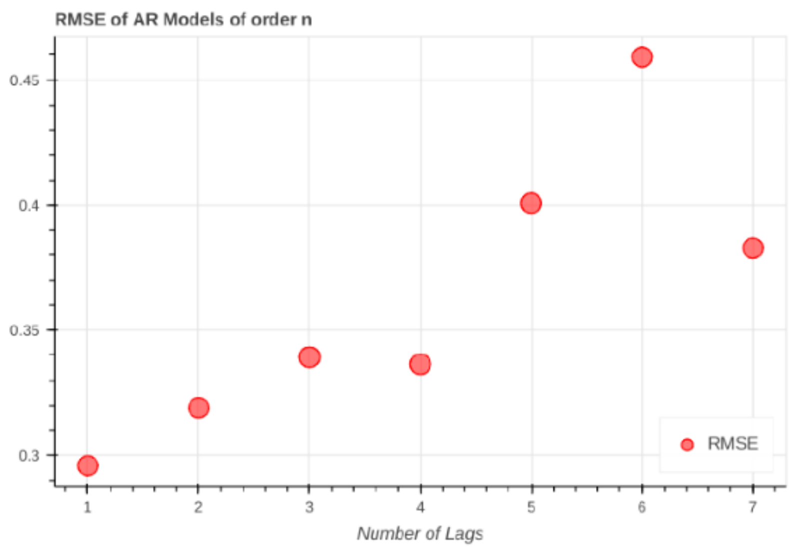

IR Detrended Data
Detrended IR Data
2008
2009
2010
2011
2012
2013

(3 points) Statement of and justification for the chosen AR(p) model. Include plots and reasoning.

To choose an AR(p) model, the autocorrelation and partial autocorrelation functions of the residuals after detrending were plotted. It can be seen from both plots that the highest lag where the autocorrelation extends beyond the statistically significant boundary is at lag 1. This would indicate that the optimal model would be an AR(1) model.



This can be confirmed by calculating the RMSE of the fit with different lags.

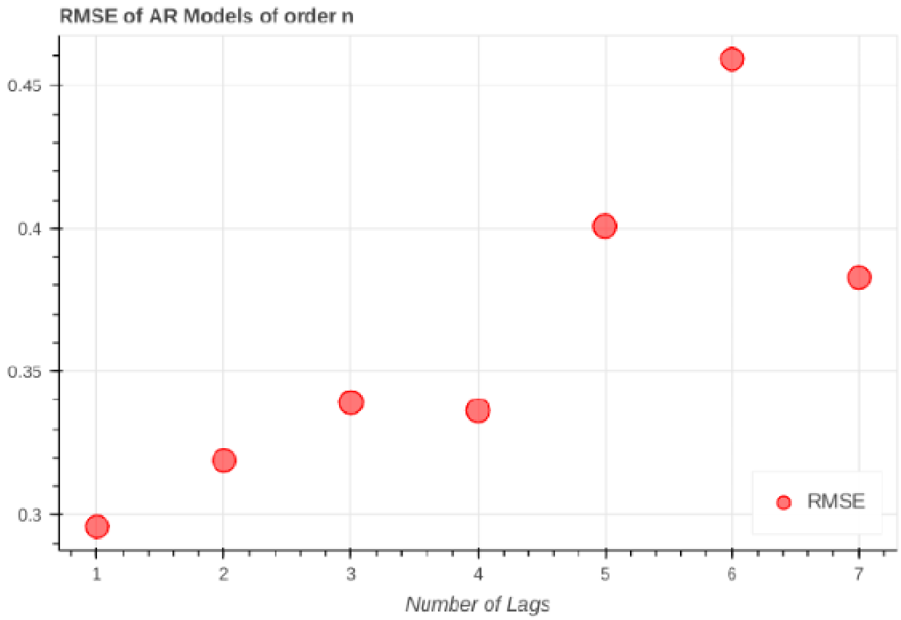


(3 points) Description of the final model; computation and plots of the 1 month-ahead forecasts for the validation data. In your plot, overlay predictions on top of the data

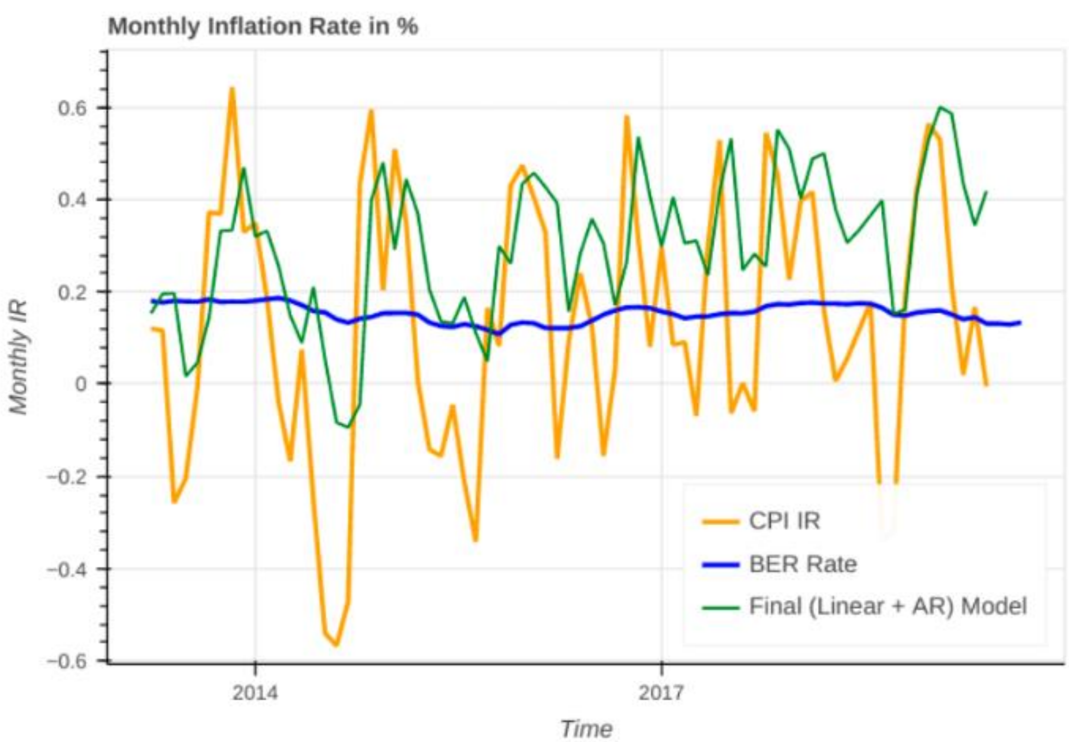
- The final model is an AR(1) model used to predict using the training and testing data.
- AutoReg was fit onto detrended training data, predictions were made from the AR model.
- The linear model provided the test predictions and the test data was then detrended by removing linear prediction.
- Lag values were computed using past detrended training data
- Coefficient 0.48651929962 and an intercept of -0.01158861882 were obtained and used to update previous values
- Obtain final values by adding back training and test prediction with the linear prediction to reinstate the trend

2. *(3 points) Which AR(p) model gives the best predictions? Include a plot of the RSME against different lags p for the model.*

   AR(1) model provides the best predictions confirmed by the lowest RSME of 0.29574

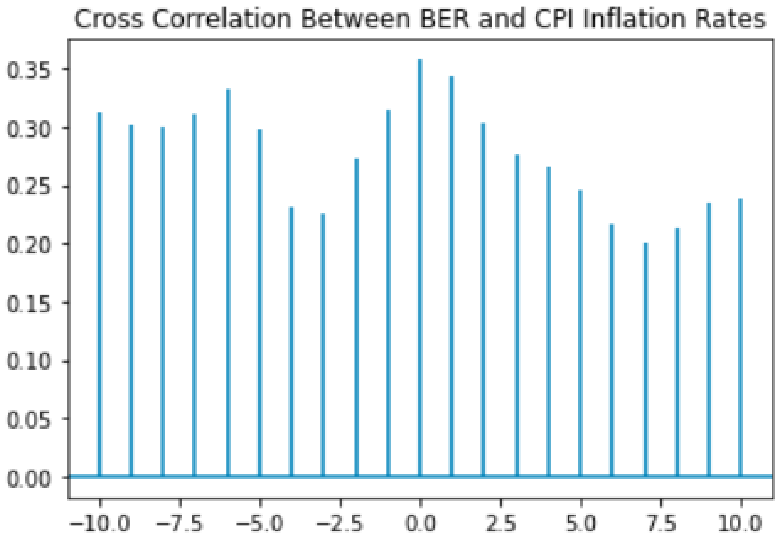


3. *(3 points) Overlay your estimates of monthly inflation rates and plot them on the same graph to compare. (There should be 3 lines, one for each datasets, plus the prediction, over time from September 2013 onward.)*



External Regressors and Model Improvements

4. *(4 points) Plot the cross correlation function between the CPI and BER inflation rate, by which find r, i.e., the lag between two inflation rates. (As only one external regressor term is involved in the model, we only consider the peak in the CCF plot.)*



From the above plot, we can see the highest peak is 0 and so r = 0

5. *(3 points) Fit a new AR model to the CPI inflation rate with these external regressors and the most appropriate lag. Report the coefficients, and plot the 1 month-ahead forecasts for the validation data. In your plot, overlay predictions on top of the data.*

SARIMAX Results

Dep. Variable:	y	No. Observations:	60
Model:	SARIMAX(1, 0, 0)	Log Likelihood	-26.128
Date:	Wed, 19 Apr 2023	AIC	58.256
Time:	19:57:42	BIC	64.539
Sample:	0	HQIC	60.714
	- 60		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
x1	1.0324	0.683	1.512	0.130	-0.306	2.370
ar.L1	0.5198	0.082	6.329	0.000	0.359	0.681
sigma2	0.1391	0.020	7.033	0.000	0.100	0.178

6. *(3 points) Report the mean squared prediction error for 1 month ahead forecasts.*

RMSE = 0.2604187092936659

Improving your Model

(5 points) What other steps can you take to improve your model from part III? What is the smallest prediction error you can obtain? Describe the model that performs best. You might consider including MA terms, adding a seasonal AR term, or adding multiple daily values (or values from different months) of BER data as external regressors.

The CPI doesn't show any seasonality so that is excluded. From analysis conducted, it might help in improving the model if we could obtain additional external regressors that could potentially correlate better with the CPI data so we can achieve better forecast data.