I) Introduction:

For this Forecasting project, I will discuss Inflation and find the best model that will forecast Inflation. This dataset has a data starting from March 2006 - October 2022. It shows monthly historical variables that have a direct relationship with inflation, such as Consumer Price Index (CPI), Minimum Wage, and Unemployment Rate.

This Dataset has 4 columns and 201 rows. Below are the variables for the Inflation Dataset:

- **Observation date:** date (YYYY-MM-DD)
- **CPI:** Consumer Price Index (Unit = percentage change from Year Ago)
- **Fed Min Wage / Hour:** Federal Minimum Wage per Hour (Unit = Percentage)
- **Unemployment Rate:** Unemployment Rate in the United States (Unit = Percentage)

Since this dataset has 16 years of monthly data, we will use the last 4 years as our validation period and the initial 12 years as our training period for our forecasting.

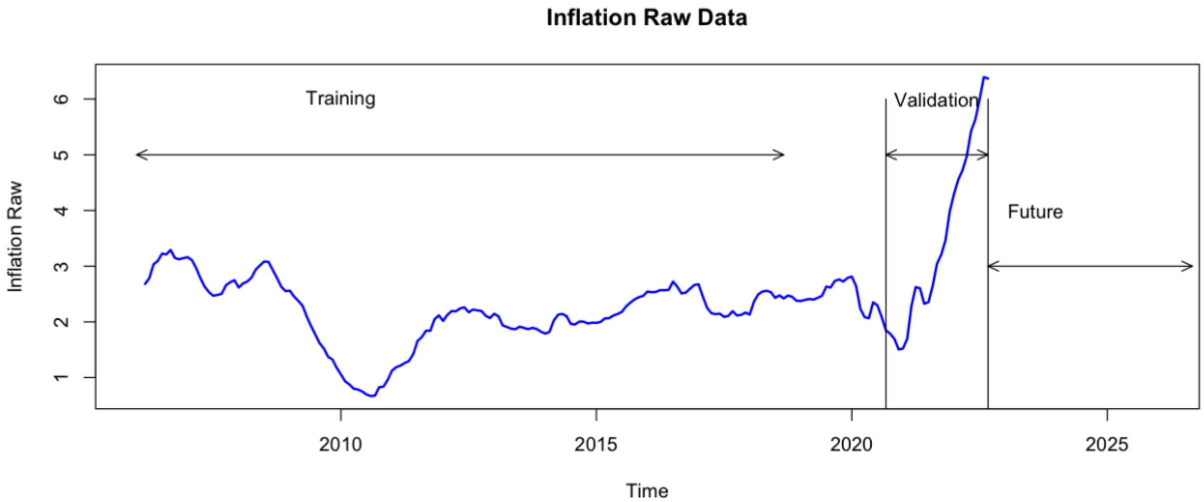
Before we dive into the Forecasting models, it's fairly important to do some preliminary work to ensure that the data is clean and ready to go. Below are the lists of preliminary work that's being utilized:

- Ensure that the data is clean, this includes
  - Remove any duplication on the date (observation_date) variables
  - Remove any N/A values
- Taking Logs on my primary variable, which is CPI (Consumer Price Index)

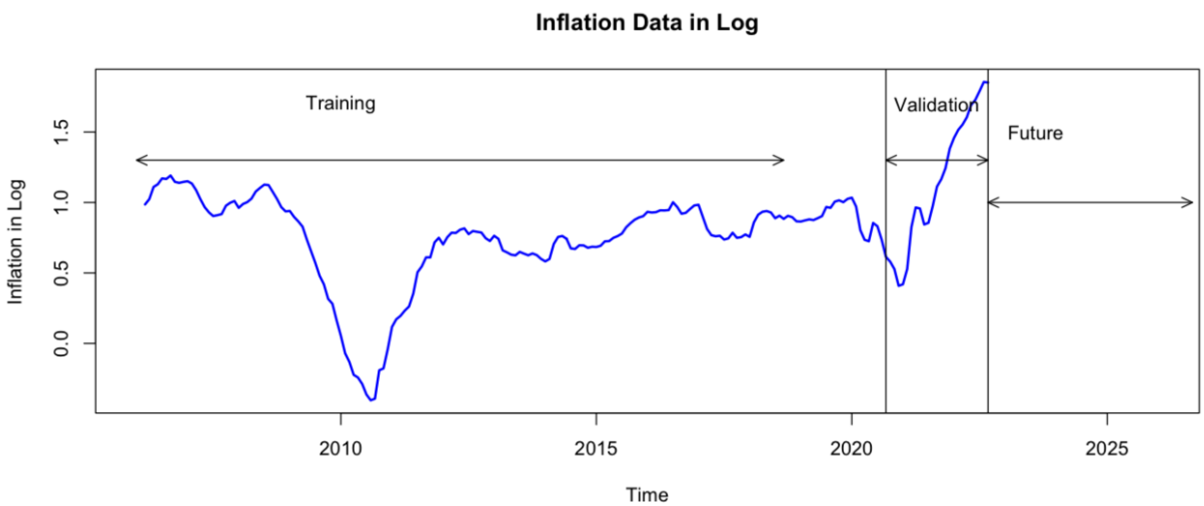
Next, we're going to look at the time series of our primary variable (CPI). We're also going to define the partitioning of the monthly Inflation data.

- Training period is from March 2006 to March 2020
- Validation period is from April 2020 to October 2022

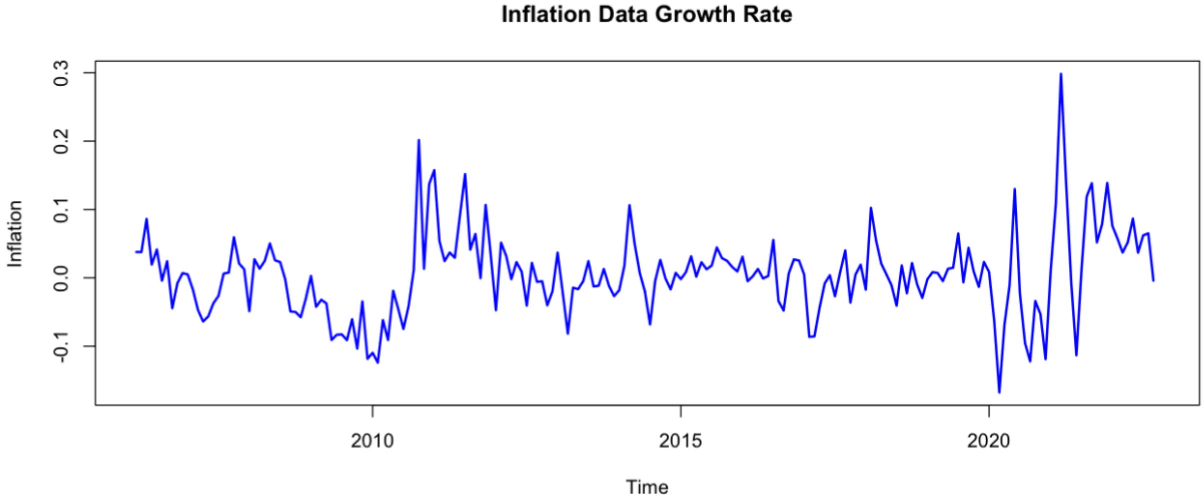
Time Series for Inflation in Raw Data



Time Series for Inflation in Log



Inflation Data Growth Rate



As we can see from our chart, Log is very helpful at transformation because Logs are a variance - stabilizing transformation. It also helps to make the data appear more linear.

After that, we're going to take a look at the Summary Statistics of our primary variable.

- Inflation.raw.ts: This indicates the time series of our raw primary variable
- Inflation.ts : This indicates the time series of our log primary variable
- Dinflation.ts : This indicates the time series of our growth rate primary variable

	inflation.raw.ts	inflation.ts	dinflation.ts
median	2.29719600	0.83168925	0.005365616
mean	2.36667404	0.79497187	0.004537670
SE.mean	0.06298694	0.02667562	0.004280337
CI.mean.0.95	0.12421135	0.05260478	0.008440900
var	0.78950365	0.14160611	0.003645936
std.dev	0.88854018	0.37630587	0.060381591
coef.var	0.37543834	0.47335746	13.306737588

Next, to determine which our forecast model is the best fit, we will be using 8 models on our forecasting method and see which one has the best fit. We will be paying attention to the MAE and RMSE.

- **MAE** stands for Mean Absolute Error, which refers to the absolute values of errors in our calculations, resulting in average errors that make more sense.

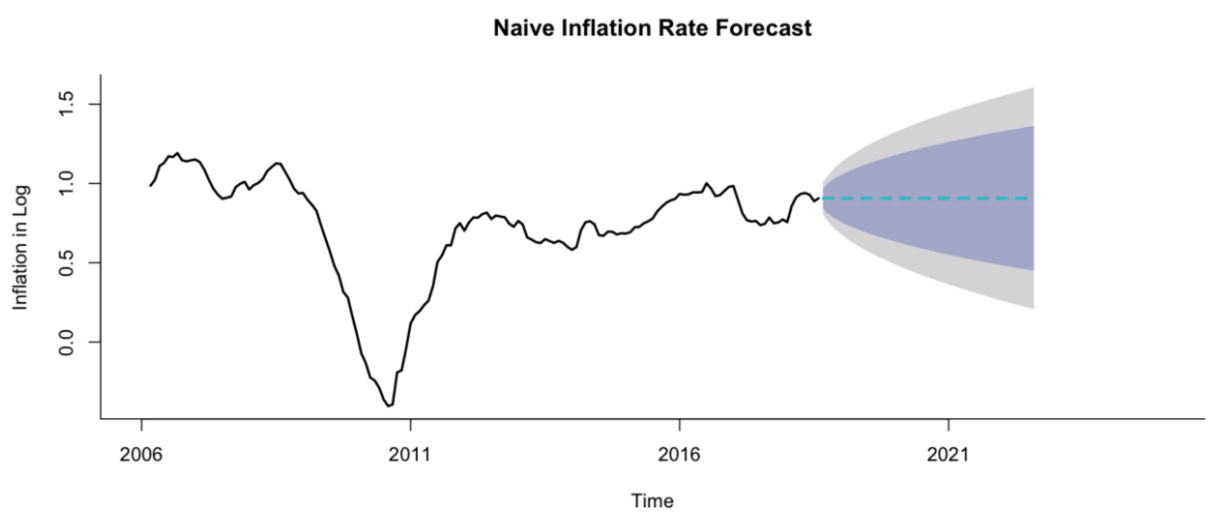
- **RMSE** stands for Root Mean Square Error. This is the standard deviation of our residuals. Since residuals measure how far the regression line data points are, the purpose of RMSE is to measure how spread out these residuals are. It will tell us how concentrated the data is around the line of best fit.
- **MAPE** stands for Mean Absolute Percentage Error is the sum of individual absolute errors divided by the demand in each period separately. MAPE is the average of the percentage errors. It presents the average deviation between the forecasted value and actual values. As an ideal rule of time, we tried to aim MAPE that is <10% which indicates the best fit.

Next, let's dive into Forecasting.

II) Summary

1) <u>Naive Forecast</u>

The Naive Forecast approach considers what happened in the previous period and predicts the same thing will happen again. In other words, a naive forecast is the most recent value of the series. Naive forecast can also act as a baseline when evaluating other forecasting method performance. Hence, on why it's always important to include Naive Forecast in our forecasting method. Naive forecast is simple – as a result, it doesn't accounts any seasonality and trend on the time series data. The chart below shows Naive Forecast on Inflation data on a monthly basis. Based on the Naive forecast, it shows that the number of Inflation is the stay almost constant or the same as the previous period.



Forecast Model	RMSE	MAE	MAPE
Naive Forecast	0.3588	0.243	24.68561

2) Moving Average

Moving Average is averaging the values across a window over consecutive periods of generating a series of averages. There are two type of moving average, which are:

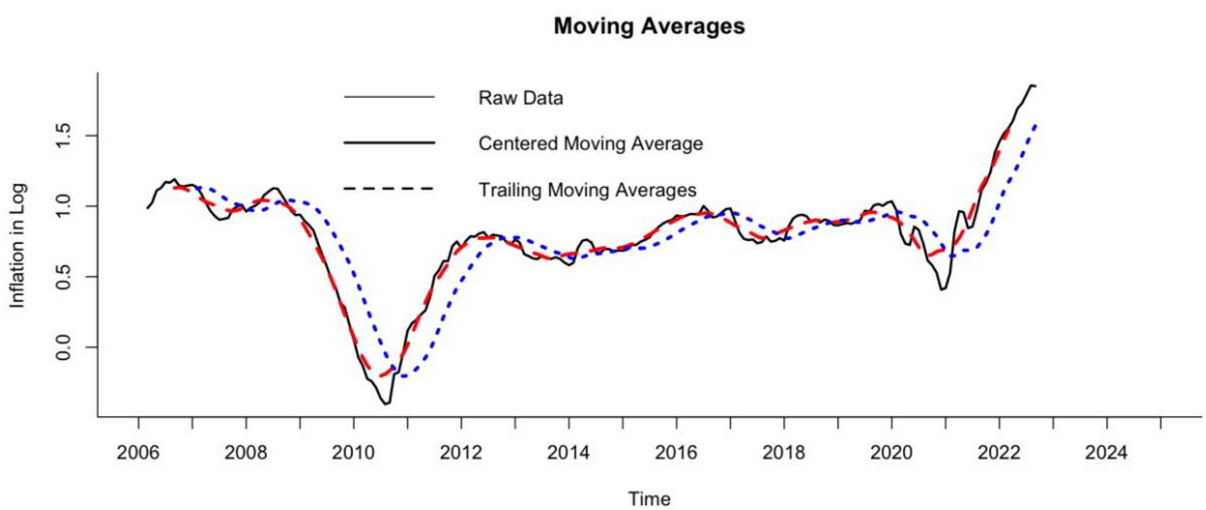
- **Centered moving average:** useful for visualizing trends since averaging can suppress seasonality and noise

$$MA_t = \left(y_{t-(w-1)/2} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+(w-1)/2}\right) / w. \tag{5.1}$$

- **Trailing moving average:** useful for forecasting

$$F_{t+k} = \left(y_t + y_{t-1} + \cdots + y_{t-w+1}\right) / w.$$

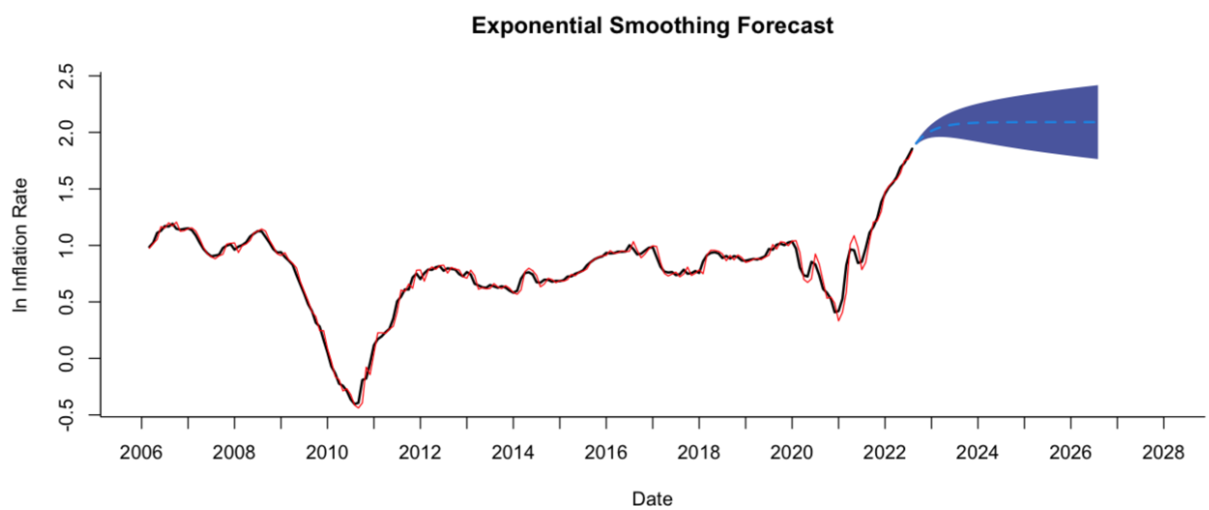
In the results below, it looks like the moving average is pretty adequate on capturing all the seasonality in the data. Centered Moving Average looks like a better fit compare to the Trailing Moving Average, because trailing moving average does lag a little. We can also see from the Diagnostic results below that the Centered Moving Average is a better fit because the MAPE between predicted and actual values is 2.6 percent.



Forecast Model	RMSE	MAE	MAPE
Centered Moving Average	0.06615756	0.04791397	2.616992
Trailing Moving Average	0.2065949	0.1537883	7.458179

3) Exponential Smoothing Forecast

Holt Winter models allow trend, level, and seasonality patterns to change over time. We are using the ZAN model and alpha = 0.2 to fit an exponential smoothing model over the training period. Based on the result, we can see the this forecast fits really well.



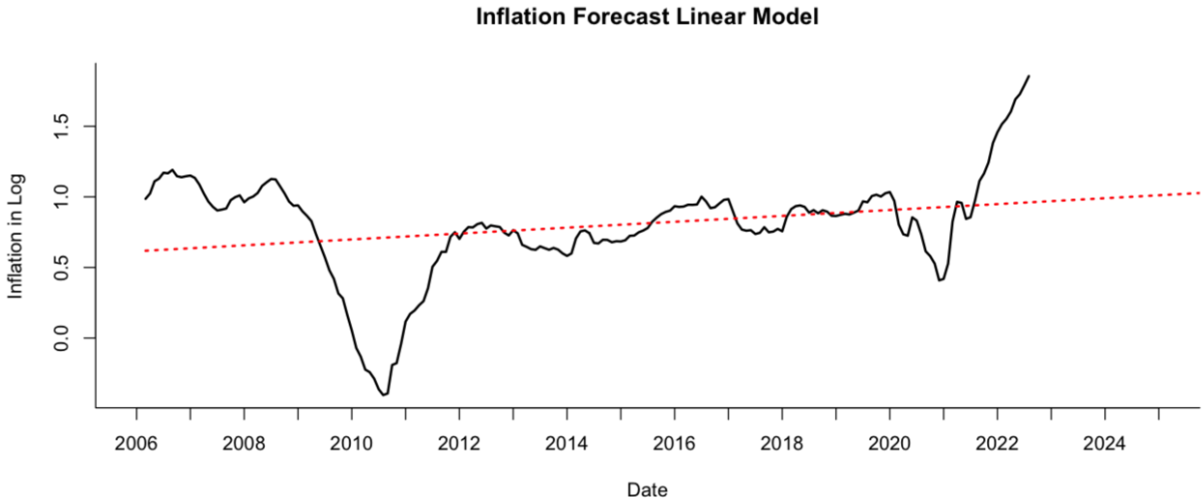
Forecast Model	RMSE	MAE	MAPE
Exponential Smoothing	0.05105135	0.03674695	8.151348

4) Linear Model Forecast

Linear Regression Trend captures the values either increasing or decreasing linearly in time. In order to fit the linear relationship between Inflation and Time, we set the y variable as Inflation (based on CPI) and the predictor as time (t) in the regression model.

$$y_t = \beta_0 + \beta_1 t + \epsilon$$

Based on the result, it looks like for the inflation dataset, it looks like it is increasing linearly in time.



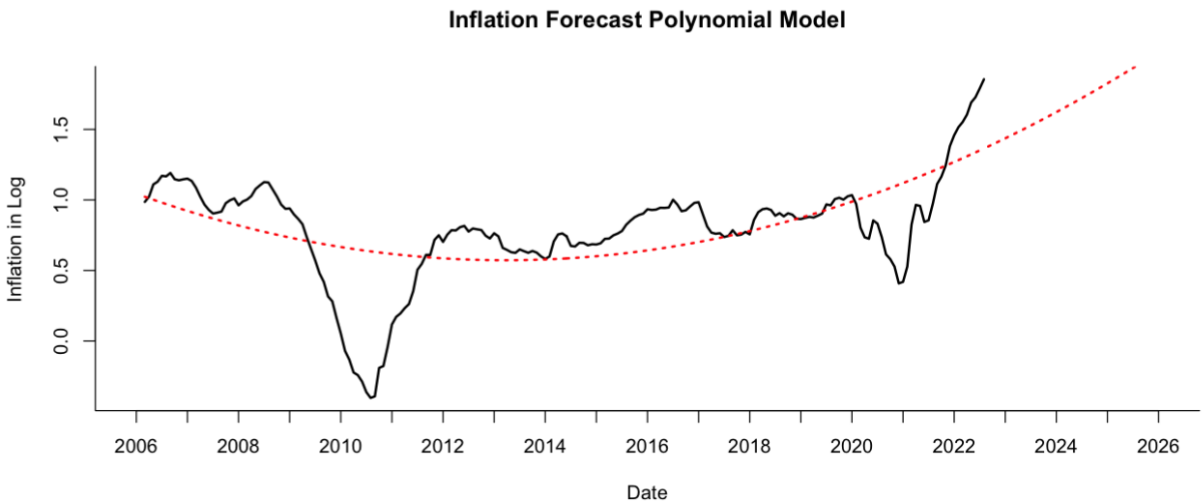
Forecast Model	RMSE	MAE	MAPE
Linear Model	0.3550611	0.2480476	19.07588

5) Polynomial Model Forecast

Polynomial Trend is fitted by adding the predictor t^2 (the squared of t) and fit a multiple linear regression with two predictors of t and t^2

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon.$$

Below, we can see that the polynomial trend captures the U Shape trend in the data



Forecast Model	RMSE	MAE	MAPE
Polynomial Model	0.3040002	0.2184891	20.21839

6) Random Walk with Drift

Here, we're going to do a Dickey Fuller Test. Below is the result:

```
        Augmented Dickey-Fuller Test

data:  inflation.ts
Dickey-Fuller = -2.7531, Lag order = 5, p-value = 0.2607
alternative hypothesis: stationary
```

Based on the results, with p-value 0.2607, we fail to reject the null hypothesis and conclude that the Inflation time series data is non-stationary. We can say that it has a constant autocorrelation structure and trends, but does not have a constant variance (the degree of spread in a data set about the mean value of that data) over a time period. Hence, we can conclude that there is evidence of random walk.

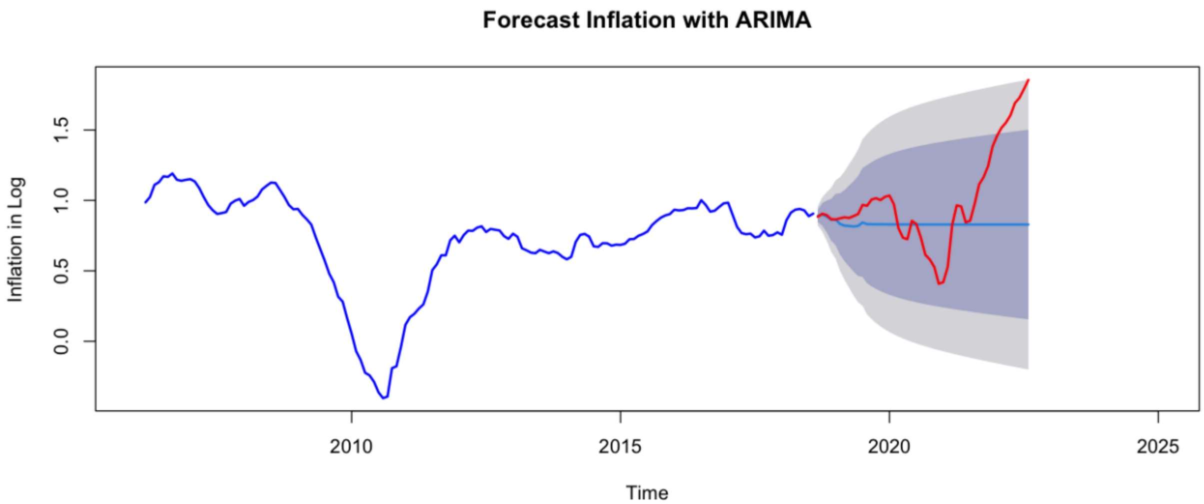
We're going to perform a Random Walk forecast with drift. The drift parameter μ equal to the average monthly growth rate in Inflation.

Diagnostics:

Forecast Model	RMSE	MAE	MAPE
Random Walk with drift	0.059	0.0408	11.382

7) **ARIMA Forecast**

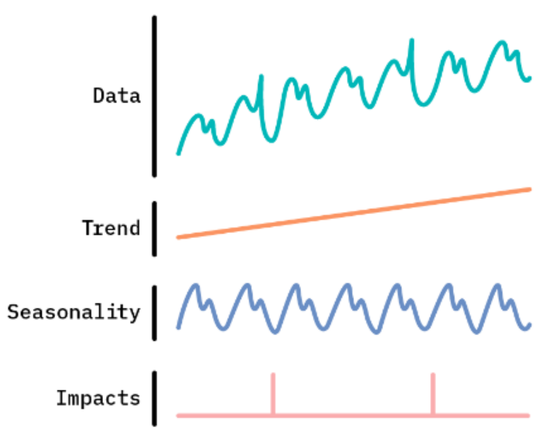
ARIMA stands for Autoregressive Integrated Moving Average. The model is a subset of linear regression models that "attempt" to use our part observations of the target variable (in this case CPI) to forecast future values. The key differentiation of the ARIMA model is that they do not consider exogenous variables in the basic form. ARIMA forecast is purely made from the past values of the target variable. Using auto.arima on my forecast, we can find the best fit ARIMA model for our time series data, which details – p (number of lags in the observations), d (number of times to differ the observations), or q(size of the moving average). In this case, my ARIMA Model is of the order of (2,1,2). Based on the result, we can see that the ARIMA model mimic the previous Inflation historical dataset, with a slight dip/decrease on one period of time, then increasing again over time.



Forecast Model	RMSE	MAE	MAPE
ARIMA	0.3863966	0.2608475	24.61019

8) <u>Structural Forecast</u>

Structural Forecast takes into consideration trend, seasonality, and external variables to the time series. Below is a quick visual of the approach on structural forecast that address these features below:



Below is the estimation of the structural model for Inflation using Wage and Unemployment Rate as our explanatory variables.

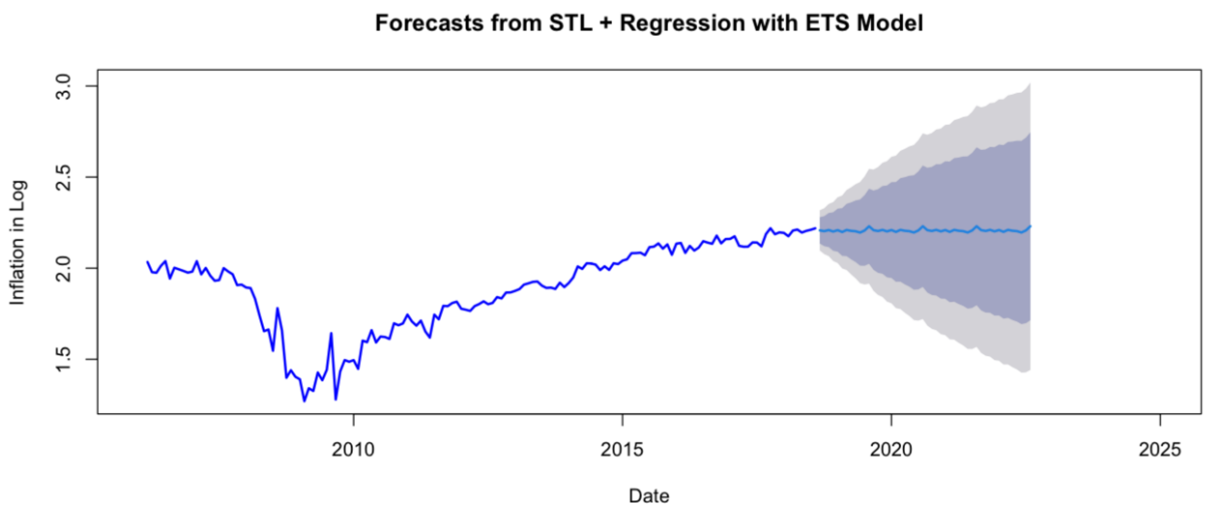
```
tslm(formula = inflation.ts ~ wage.ts + unemployment.ts + trend)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.82332 -0.09935  0.01579  0.14893  0.74332 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -2.836032   0.499132  -5.682 4.79e-08 ***
wage.ts          0.220648   0.025499   8.653 1.86e-15 ***
unemployment.ts -0.116649   0.008293 -14.067  < 2e-16 ***
trend           -0.011897   0.001453  -8.188 3.41e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.239 on 195 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.6026,	Adjusted R-squared:  0.5965 
F-statistic: 98.56 on 3 and 195 DF,  p-value: < 2.2e-16
```

Next, we're going to use the estimated model from above to forecast Inflation Data for a 4 year (48 month) validation period



Forecast Model	RMSE	MAE	MAPE
Structural Forecast	0.05532834	0.03696628	2.139591

III) Two Best Out of Sample Forecast

Below are all of the diagnostic results for all of the Forecast Model that we conducted from above.

Forecast Model	RMSE	MAE	MAPE
Naive Forecast	0.3588	0.243	24.68561
Random Walk with drift	0.059	0.0408	11.382
ARIMA	0.3863966	0.2608475	24.61019
Structural Forecast	0.05532834	0.03696628	2.139591
Centered Moving Average	0.06615756	0.04791397	2.616992
Trailing Moving Average	0.2065949	0.1537883	7.458179
Exponential Smoothing	0.05105135	0.03674695	8.151348
Linear Model	0.3550611	0.2480476	19.07588
Polynomial Model	0.3040002	0.2184891	20.21839

Based on the data provided above, we're going to pick the best 2 Out of Sample Forecasts from each week and do some discussions.

In Week 4 and 5, we conducted Naive, Centered Moving Average, Trailing Moving Average, Linear, Polynomial and exponential smoothing methods forecasts. Even though Centered Moving Average has the lowest MAPE – I would not pick this as my best forecast. The reason is Moving Average has some pitfalls – it only draws trends from past information only and not taking any other factors that might impact the forecast. Exponential smoothing on the other hand takes trend, level, and seasonality patterns into account. Hence, for week 4 and 5, I will pick Exponential Smoothing.

In Week 6, we conducted Random Walk with drift, ARIMA, and Structural Forecast. Based on the results, Structural Forecast is the best fit because it captures the trend from other external variables such as Wage and Unemployment Rate that has direct relationship with inflation

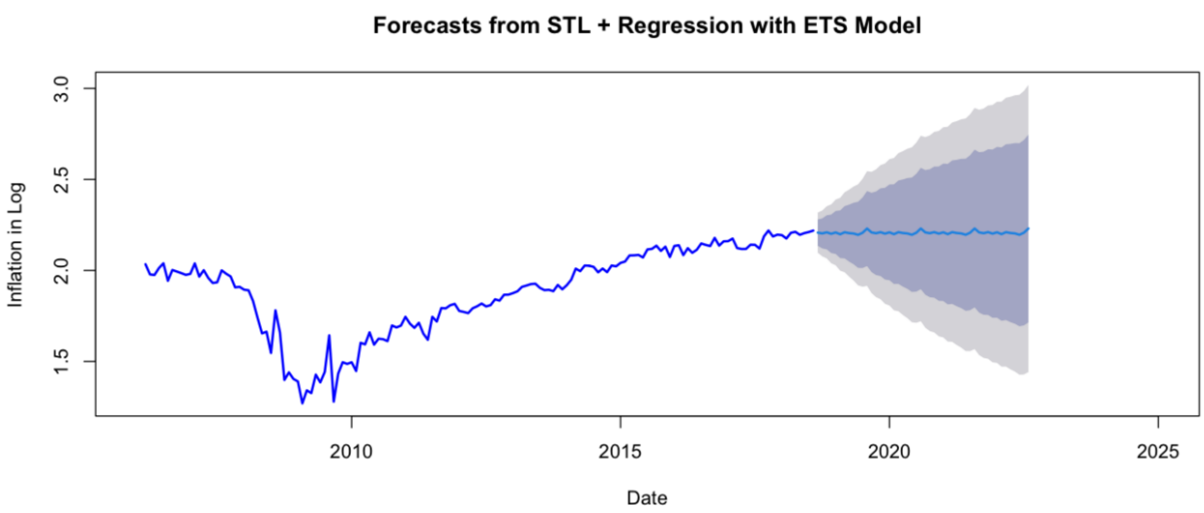
Based on all discussions and results, we're going to conclude our best 2 Out of Sample Forecasts, which are Exponential smoothing and Structural Forecast.

Below are the results.

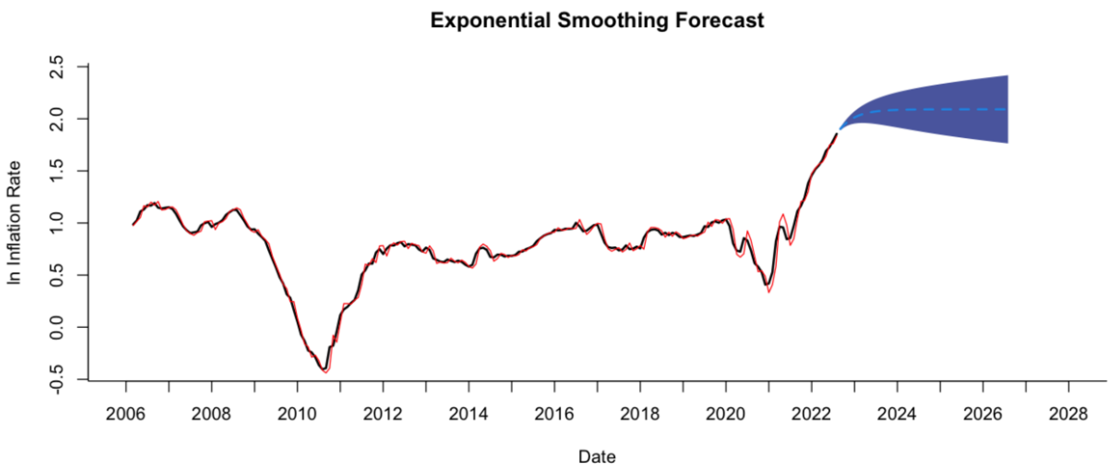
Forecast Model	RMSE	MAE	MAPE
Structural Forecast	0.05532834	0.03696628	2.139591
Exponential Smoothing	0.05105135	0.03674695	8.151348

IV) Conclusion

Structural Forecast – this model takes two regressors into our model which are Wage and Unemployment Rate. Our Primary variable is CPI (Consumer Price Index). Based on the forecast results below, we can see that these two regressors fit really well to CPI. It make senses because as CPI goes up, wages will go up as well because people need more money to be able to afford consumer goods and services, which indicates a positive correlation. Inflation and unemployment on the other hand traditionally had an inverse relationship. When one rises, the other drops. Low levels of unemployment typically corresponded with higher inflation, while high unemployment corresponded with lower inflation. Low unemployment indicates more consumers have more income to purchase goods and the demand for goods rises. When that happens, prices will follow. On the other hand, high unemployment, consumers purchase fewer goods as they're trying to save up, which puts downward pressure on prices and reduces inflation. Below is the plot for Structural Forecast:



Exponential Smoothing – is a technique for smoothing time series data using the exponential window function. This model allows trend, level, and seasonality patterns to change over time. Hence why it is one of the best fit for our dataset. Based on the results below, it captures all the trends and patterns on our dataset accurately. Hence, the forecast takes trend and seasonality into accounts.



V) References

Shmueli, G., & Lichtendahl, K. C. (2018). Practical time series forecasting with R: A hands-on guide. Axelrod Schnall Publishers.

Shumway, R,H., & Stoffer, D, S. (2017). Time Series Analysis and Its Applications: With R Examples (Springer Texts in Statistics). Springer.

Consumer Price Index for All Urban Consumers: All Items in U.S. City Average. (2022, November 10). Fred Economic Data. https://fred.stlouisfed.org/series/CPIAUCSL

Unemployment Rate. (2022, November). Fred Economic Data. https://fred.stlouisfed.org/series/UNRATE

Average Hourly Earnings of All Employees, Total Private.(2022, November). Fred Economic Data. https://fred.stlouisfed.org/series/CES0500000003