TIME SERIES ANALYSIS AND FORECASTING OF MACD TECHNICAL INDICATOR FOR STOCK TRADING

TEAM

KISHAN PATEL

VAIBHAV SHAH

CONTENTS

EXECUTIVE SUMMARY

In this project, analysis and forecast is performed for the technical indicator called MACD, used by traders to determine BUY or SELL position in stock trading. For analysis purposes, data is collected for weekly adjusted closing stock price of Alphabet Inc. (GOOGLE) for last 5 years from yahoo finance website and two time series data MACDLine and SignalLine for analysis and forecasting has been created.

From visualization of time series data, it is identified that both time series have polynomial trend but no seasonality. Also, correlogram plot shows that initial lags have very high positive correlation and shows high trend but lag 52 has very high negative correlation and shows no seasonality.

For predictability of the data, AR (1) model is used, where the β value is close to 1 but standard error is low. It is verified that lag-1 difference of both time series using correlogram for checking all lags are above both significance levels. Data is partitioned with a 60:40 ratio for training and validation purpose for developing models and validating the same.

Forecasting techniques used are 1) Two-level combined (Regression trend with 3rd order polynomial + AR(3) of Residuals), 2) Auto Holt-Winter's Model, 3) ARIMA and 4) Auto ARIMA to analyze and forecast both time series data, where accuracy of models is compared based on MAPE and RMSE error metrics. Finally, it has been concluded that the best model is Auto ARIMA with order 2 autoregressive model AR (2) for **MACD-Line** and order 2 autoregressive model AR (2) with order 1 moving average MA (1) for error lags for **Signal-Line** and recommendation is provided for future use of forecasting model.

INTRODUCTION

What Is Moving Average Convergence Divergence (MACD)?

MACD, short for moving average convergence/divergence, is a trading indicator used in technical analysis of stock prices, created by Gerald Appel in the late 1970s. It is designed to reveal changes in the strength, direction, momentum, and duration of a trend in a stock's price. MACD indicator consist of 2 lines 1) MACDLine, 2) SignalLine.

KEY Points about MACD:

- The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA. The result of that calculation is the **MACDLine**.
- A 9 period EMA of the MACD called the **SignalLine**.
- MACD helps investors understand whether the bullish or bearish movement in the price is strengthening or weakening.

These 2 lines are plotted and the crossover of these 2 lines functions acts as a trigger for BUY and SELL signals. Traders may BUY the security when the MACD crosses above its signal line and SELL/SHORT the security when the MACD crosses below the signal line.

- **MOTIVATION:**

As per study, there are 60% of adults in USA taking part in stock market trading and many brokerage systems provides their system for technical analysis. As technical analysis is one of the key foundations for day to day, short-term and medium-term trading. There are numerous sources available on internet world about forecasting the price of stock. However, there isn't any system available to forecast technical indicators which helps traders who uses technical indicators for their trading.

Hence the prime motive of this project is to study this new area and apply time series analysis knowledge to predict one of the key technical indicator called **MACD** to help trading world to perform more risk free trading and act smartly in the most volatile world called STOCK MARKET.

As a part of this project, two individual time series MACDLine and SignalLine will be forecasted for future eight periods.

EIGHT STEPS OF FORECASTING

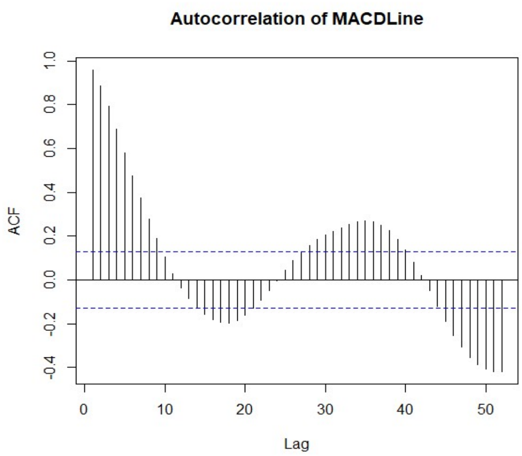
1. DEFINE GOAL

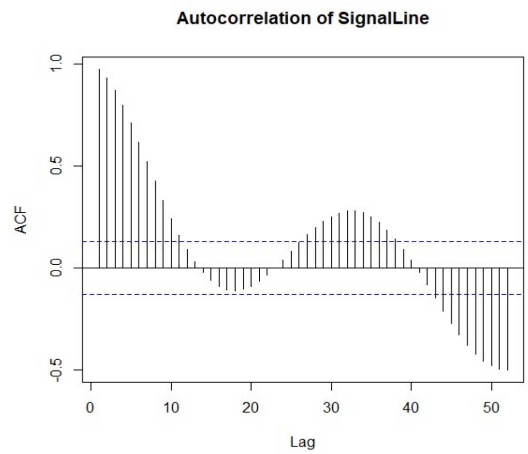
The goal of this project is to analyze the MACD technical indicator values and point forecast two time series MACD-Line and Signal-Line for next 8 weeks. The objective is to create a model which will analyze historical data based on trend and forecast weekly indicator values to help a trader to take early and better decision in stock market trading. As stock market is highly random in nature and this model will use weekly stock price data, model needs to be reevaluated at the end of every two weeks for better accuracy of forecasts. This project will be developed in R Language.

2. GET DATA

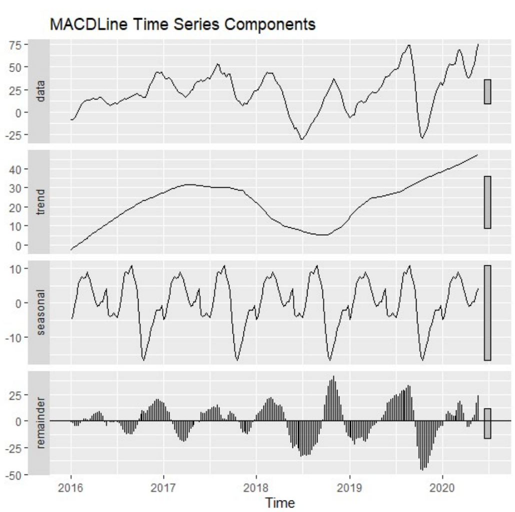
For this project, Alphabet Inc. (GOOGLE) from Nasdaq stock exchange has been selected. Collection of data is done from https://finance.yahoo.com/quote/GOOGL/history?p=GOOGL website from 2nd week of Nov-2015 to 2nd week of Nov-2020 with frequency as weekly adjusted closing price and saved in .csv format and also used ***quantmod*** library to generate difference of 26 & 12 Days EMA for MACD Line and 9 days EMA for Signal Line.

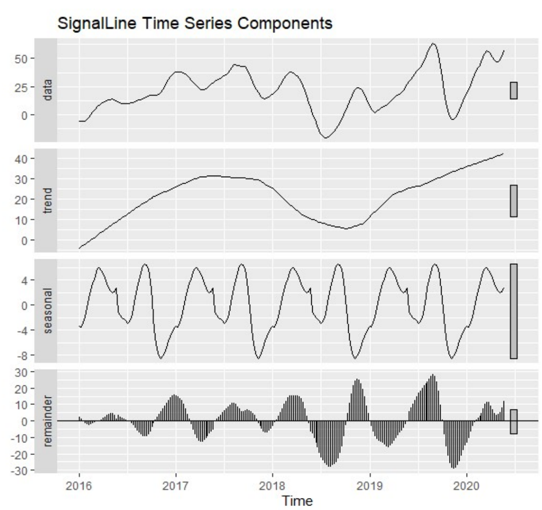
3. EXPLORE AND VISUALIZE SERIES





From Correlogram it is observed that both time series data has very high positive correlation in initial lags and shows strong trend but lag 52 has high negative autocorrelation in seasonality. Similarly, a polynomial trend pattern can also be seen in the below components chart for both time series data. Based on these charts, it is concluded that both time series have trend components but no seasonality component.





3.1. CHECK PREDICTABILITY

During initial analysis, the predictability check of data with following two approaches has been performed:

3.1.1. Approach 1: Apply AR(1) To Check Predictability Of Data

AR(1) model has been applied to both time series macdline.ts and signalline.ts and performed analysis on the summary.

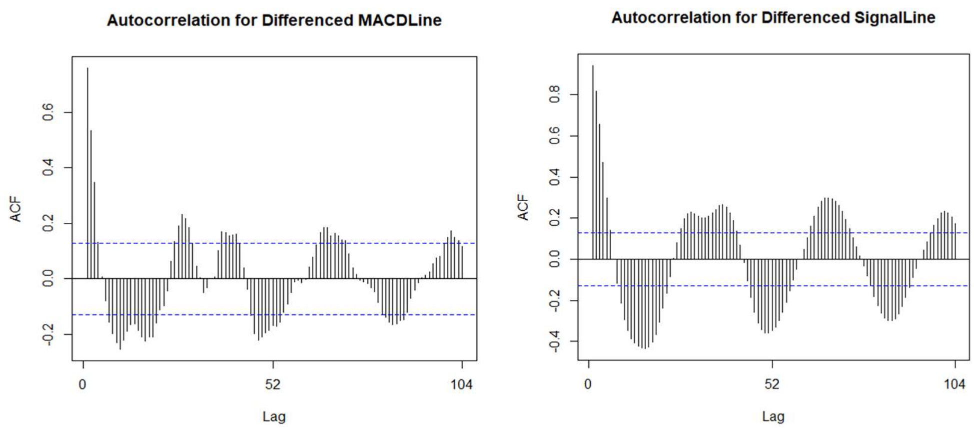
```
Series: macdline.ts
ARIMA(1,0,0) with non-

Coefficients:
         ar1     mean
      0.9849  26.5753
s.e.  0.0121  17.1285
```

```
> summary(signal.ar1)
Series: signalline.ts
ARIMA(1,0,0) with non-

Coefficients:
         ar1     mean
      0.9935  23.4291
s.e.  0.0068  18.4115
```

Time Series	**Beta**	**Comments**
macd.ar1	0.9849	As beta is high for both time series but std error is quite low, it is concluded that these time series are **predictable.**
signal.ar1	0.9935	

3.1.2. Approach 2: By Differencing Of Lag-1

With diff() function, a difference data of both time series has been created and applied the difference result in the acf() function with lag.max = 104 to check the significance. From below correlograms of both time series it can be concluded that the maximum data lags are above upper and lower significance levels. Hence this time series data is **predictable**.



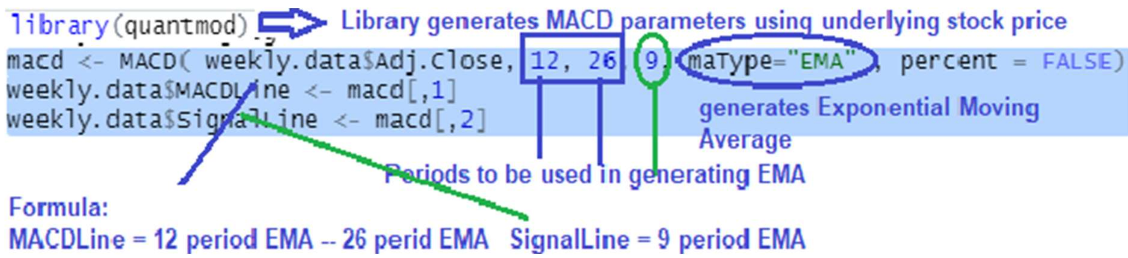
4. DATA PREPROCESSING

The original data downloaded from finance.yahoo.com contains the weekly data of [Date, Open, High, Low, Close, Adj Close, Volume] of GOOGLE, stored in 'GOOG.csv' file. 'AdjClose' price is used for analysis while rest all other columns are kept as it is in the original file. original data frame: **goog** <- data read from drive and saved in this data frame.

weekly.data - is a data frame that will be used for entire analysis with 262 observations and two variables selected as Date, AdjClose from original data frame.

4.1. GENERATE MACDLINE & SIGNALLINE DATA

As the main goal isn't to forecast the stock price, but to generate a statistical indicator using AdjClose price, two more columns needs to be added to the data frame which is called 'MACDLine' & 'SignalLine'. It is generated using a special r library called '**quantmod**' by using a function macd(), as explained in below snapshot. Also added these two lists to main dataframe '**weekly.data**' which will have 262 observations with 4 variables.



Last 6 observations: -

```
> tail(weekly.data)
          Date Adj.Close MACDLine SignalLine
257 2020-10-05   1515.22 37.99418   48.62008
258 2020-10-12   1573.01 41.26973   47.15001
259 2020-10-19   1641.00 48.78944   47.47790
260 2020-10-26   1621.01 52.53030   48.48838
261 2020-11-02   1761.75 66.08966   52.00863
262 2020-11-09   1749.84 75.00984   56.60887
```

4.2. REMOVE NA VALUES

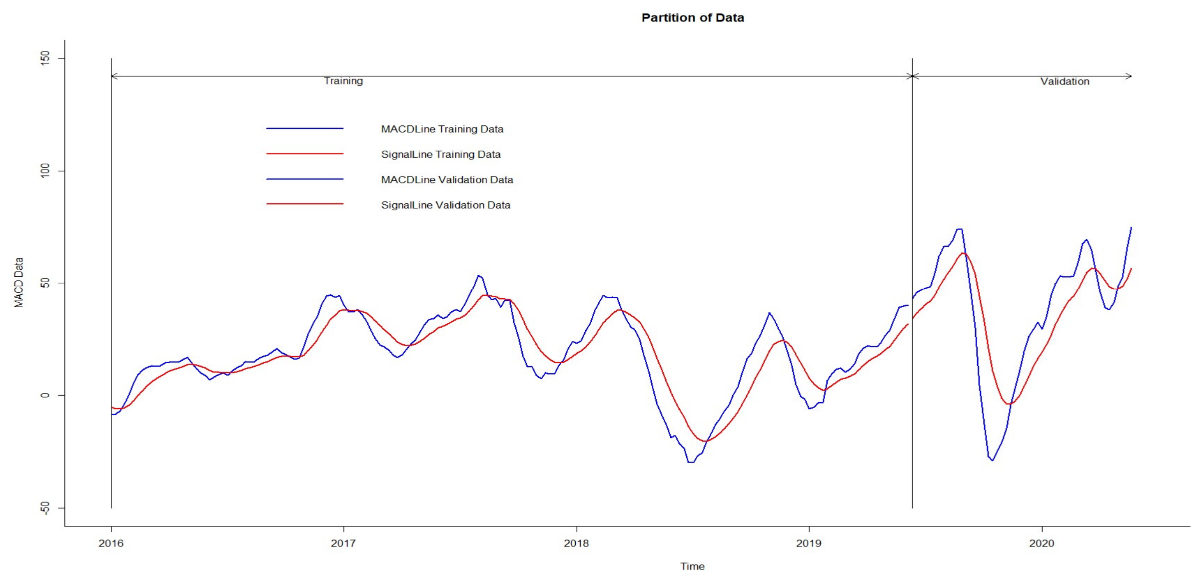
After generating the Moving Average data, the ‘NA’ rows are removed from ‘weekly.data’ main data frame as the initial 33 periods are being used in generating MACDLine and SignalLine. After removing ‘NA’ values, there are a total of 229 observations with 4 variables left in the weekly.data.

4.3. CREATING Two TIME SERIES DATA FRAMES

Two main time series data frames macdline.ts and signalline.ts are created with frequency = 52 for the analysis and forecasting models containing **229** observations.

5. PARTITION SERIES

A data partition of 60:40 ratio is created. Total 179 data records were used for training purposes and 50 data records for validation purposes. Training data from 20-06-2016 to 25-11-2019 and validation data till date 09-11-2020. Below plot shows the training and validation partition.



6. APPLYING FORECAST METHODS

6.1. MODEL 1: REGRESSION WITH 3rd POLYNOMIAL TREND + AR(3) RESIDUALS

Linear Regression was used to check the significance of trend and seasonality in a data. By applying different degrees of polynomial regression with and without seasonality, it was observed that seasonality was not affecting the model in training and hence seasonality has been removed. Thus, 3 degree polynomial trend is used.

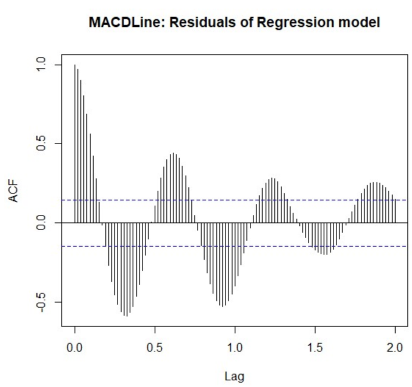
```
> summary(macdline.reg)

Call:
tslm(formula = macdline.train ~ trend + I(trend^2) + I(trend^3))

Residuals:
    Min      1Q  Median      3Q     Max 
-38.639 -10.615   2.001   9.627  29.567 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.487e+01  4.429e+00  -3.357 0.000965 ***
trend        1.959e+00  2.125e-01   9.217  < 2e-16 ***
I(trend^2)  -2.439e-02  2.739e-03  -8.903 6.79e-16 ***
I(trend^3)   8.248e-05  1.001e-05   8.244 3.82e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.51 on 175 degrees of freedom
Multiple R-squared:  0.3441,	Adjusted R-squared:  0.3328 
F-statistic:  30.6 on 3 and 175 DF,  p-value: 5.948e-16
```

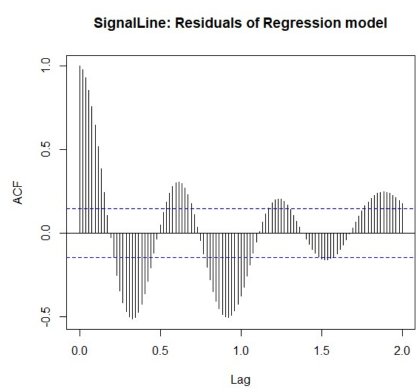
```
> summary(signalline.reg)

Call:
tslm(formula = signalline.train ~ trend + I(trend^2) + I(trend^3))

Residuals:
     Min       1Q   Median       3Q      Max 
-30.3996  -7.1155   0.9076   6.4910  21.9838 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.669e+01  3.480e+00  -4.794 3.48e-06 ***
trend        1.847e+00  1.670e-01  11.063  < 2e-16 ***
I(trend^2)  -2.150e-02  2.153e-03  -9.989  < 2e-16 ***
I(trend^3)   6.866e-05  7.862e-06   8.733 1.94e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.4 on 175 degrees of freedom
Multiple R-squared:  0.4575,	Adjusted R-squared:  0.4482 
F-statistic: 49.19 on 3 and 175 DF,  p-value: < 2.2e-16
```

Looking at the summary of both time series, MACDLine and SignalLine, coefficients of are statistically significant where MACDLine explains goodness of fit around 33% of historical data in a model, whereas SignalLine explains around 44% of data in its model.

- Regression equation MACDLine: $\mathbf{Y_t = -14.87 + 1.959t - 0.0244t^2 + 0.000082t^3}$
- Regression equation SignaLine: $\mathbf{Y_t = -16.69 + 1.847t - 0.0215t^2 + 0.000069t^3}$

After applying ACF to residuals of the regression model, there were significant lags where lag-1 has highest significance. Thereafter, AR(3) was applied, order 3 autoregressive model, to the forecast of the regression model which incorporated residuals in a model except lag 1, as shown in figure below:





MACDLine AR(3) Summary:

```
> summary(macdline.res.ar3)

Call:
arima(x = macdline.reg.pred$residuals, order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3  intercept
      1.5340  -0.4298  -0.1598     0.3929
s.e.  0.0734   0.1326   0.0734     3.1229

sigma^2 estimated as 5.595:  log likelihood = -410.24,  aic = 830.48

Training set error measures:
                      ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set -0.01015356 2.365448 1.825698 -47.15701 79.54055 0.6835358 -0.02367865
```

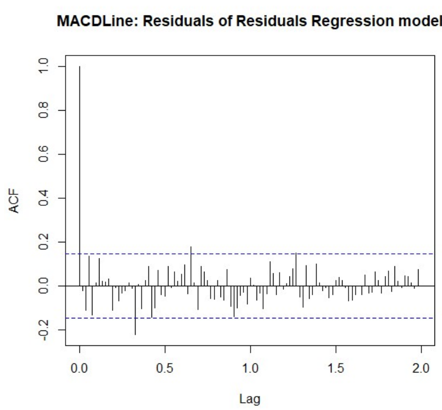
SignalLine AR(3) Summary:

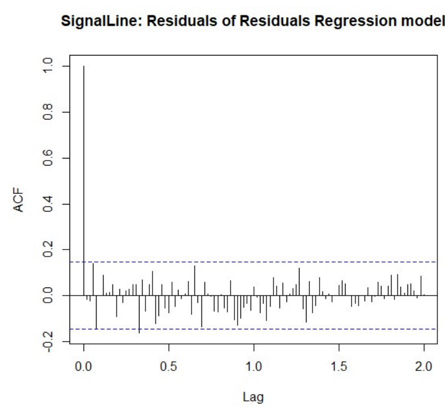
```
> summary(signalline.res.ar3)

Call:
arima(x = signalline.reg.pred$residuals, order = c(3, 0, 0))

Coefficients:
         ar1      ar2     ar3  intercept
      2.3073  -1.7085  0.3828     0.8981
s.e.  0.0687   0.1348  0.0691     1.9040

sigma^2 estimated as 0.2207:  log likelihood = -123.47,  aic = 256.95

Training set error measures:
                      ME      RMSE       MAE        MPE     MAPE     MASE        ACF1
Training set -0.006320388 0.4697634 0.3631651 0.03757129 8.412291 0.225784 -0.01501609
```

- Equation of AR(3) for MACDLine: $\mathbf{Y_t = 0.3929+1.534\beta t\text{-}1 - 0.4298\beta t\text{-}2 - 0.1598\beta t\text{-}3}$
- Equation of AR(3) for SignalLine: $\mathbf{Y_t = 0.8981+2.3073\beta t\text{-}1 - 1.7085\beta t\text{-}2 +0.3828\beta t\text{-}3}$

Finally, regression forecast and AR(3) forecast is combined to form a two level forecast with 3 degree polynomial regression model and AR model for residuals.

From below correlogram, it is identified that data is left with only noise and no more predictable.



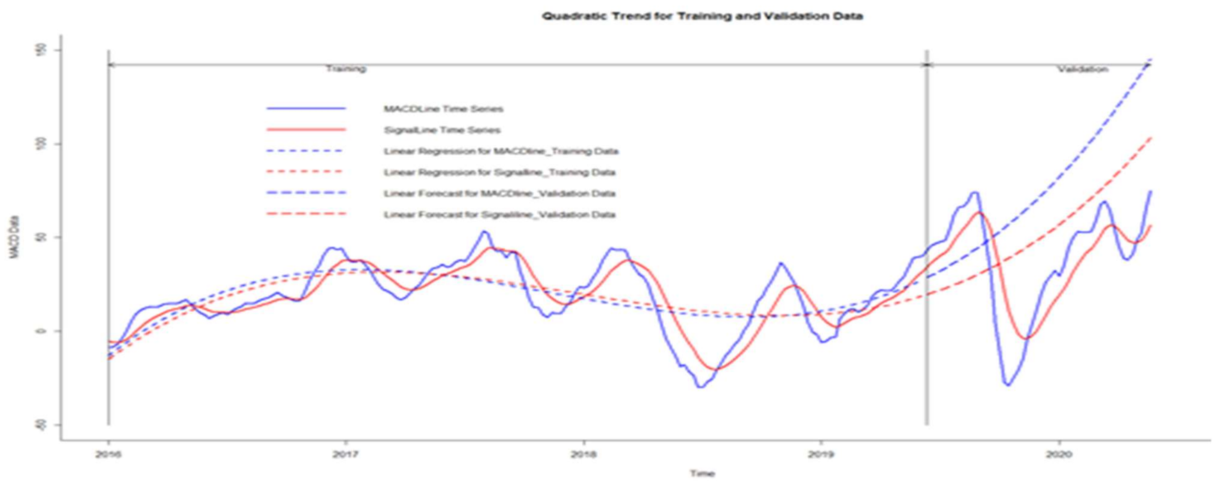



```
> round(accuracy(macdline.2level.valid.pred, macdline.valid),3)
              ME   RMSE    MAE     MPE    MAPE  ACF1 Theil's U
Test set -39.324 55.042 48.563 -57.784 232.981 0.959     5.876
> round(accuracy(signalline.2level.valid.pred, signalline.valid),3)
             ME   RMSE    MAE      MPE    MAPE ACF1 Theil's U
Test set -19.19 32.467 29.308 -456.098 749.587 0.97     11.23
```

Model	RMSE	MAPE	Theil's U
MACD: 3 degree polynomial Regression + AR(3)	55.042	232.981	5.876
Signal: degree polynomial Regression + AR(3)	32.467	749.587	11.23

CAL STATE EAST BAY

Looking at the accuracy above, it can be seen that both have very high MAPE(MACD MAPE = 232.981 and Signal MAPE = 749.587) and RMSE for MACD line(55.042) and Signal line (32.467). **Theil's U** is also > 1 for both lines which clearly indicates that the Naive Forecast model will be able to forecast better than this model.



Graph shows the forecast on validation data. It is observed that data grows exponentially and is **overestimated.** Also, it does not show seasonality.

6.2. MODEL 2: HW MODEL - AUTOMATIC SELECTION OF MODEL

After analyzing, Linear Regression model with AR(3) residuals, another forecasting model of Holt-Winter's model is developed with automatic selection of model variables. The Holt-Winter's Model is used for level, trend and seasonal components. As the time series data has no seasonality, it was decided to develop the Hold Winter's model with Automated selection of model parameters.

Learning: When ets() function was applied, it was noticed that ets() can't handle data with Frequency > 24. Hence, it was decided to develop a HW model with stlf() function and parameter etsmodel = 'ZZZ' on training data. stlf() function also generates the point forecast. Hence, generated the forecast for validation data.

```
> macdline.hw.zzz$model
ETS(A,Ad,N)

Call:
 ets(y = na.interp(x), model = etsmodel,

  Smoothing parameters:
    alpha = 0.9621
    beta  = 0.7321
    phi   = 0.8038

  Initial states:
    l = -6.3757
    b = 0.875

  sigma:  2.3318

     AIC     AICc      BIC
1238.575 1239.064 1257.700
```

```
> signalline.hw.zzz$model
ETS(A,Ad,N)

Call:
 ets(y = na.interp(x), model = etsmodel,

  Smoothing parameters:
    alpha = 0.868
    beta  = 0.868
    phi   = 0.8863

  Initial states:
    l = -5.3045
    b = -0.6576

  sigma:  1.136

      AIC      AICc       BIC
 981.1259  981.6142 1000.2502
```

From above summary it is observed that the model has no seasonality as it is also concluded in the initial analysis during exploration and visualization. So, this is a **Holt's Model.**

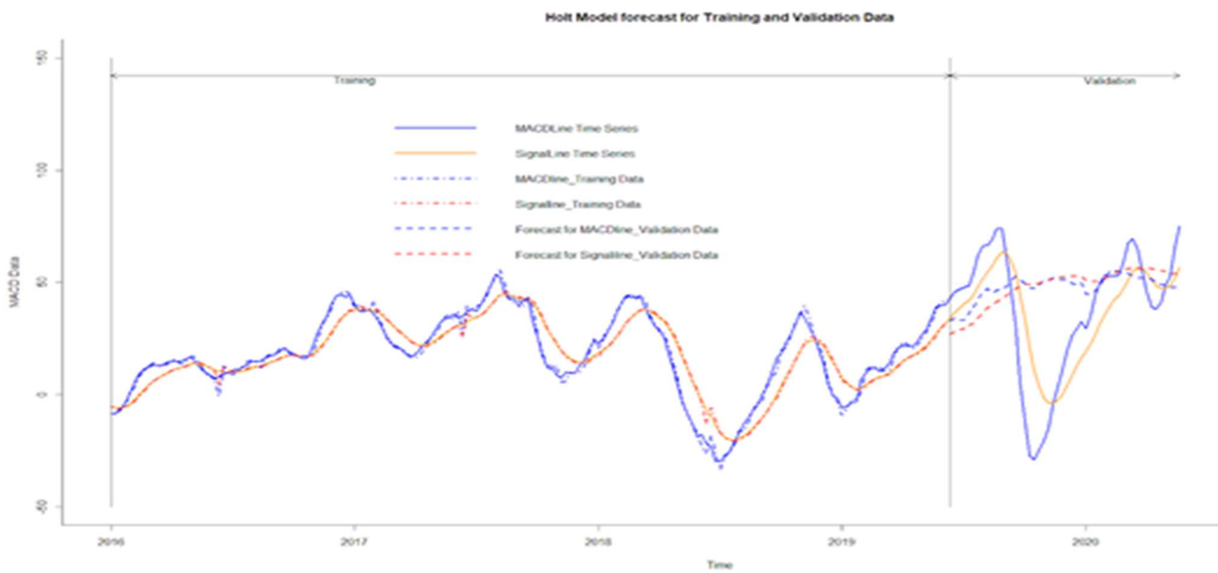
The model which is selected is ets(A, Ad, N) for both time series macdline and signalline.

Optimal values for smoothing constants of level ($\boldsymbol{\alpha}$) and trend ($\boldsymbol{\beta}$) for both models are:

Time Series	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	**comments**
macdline	0.9621	0.7321	It can be concluded that alpha and beta values tend to be more locally adjusted than globally. Also, the level and trend changes over time.
signalline	0.868	0.868	

Using this HW model, the point forecast in the validation period (from Start = c(2019, 24) End = c(2020, 21)) is presented below:

```
> macdline.hw.zzz$mean
Time Series:
Start = c(2019, 24)
End = c(2020, 21)
Frequency = 52
 [1] 32.90461 34.10907 33.13452 33.04712 36.22069 38.83466 42.22330 46.04308 47.09478 45.46189 46.38241 47.66634 47.98071 50.56999 53.05197 51.21857
[17] 49.38252 47.83102 47.21514 48.69090 51.01434 51.22339 52.01833 51.47961 51.19365 51.28255 49.45369 48.88793 49.95409 44.73392 44.47879 46.42974
[33] 48.12326 52.79258 54.57400 55.59620 54.87356 53.87217 54.00566 52.63210 52.22382 51.83183 52.20694 51.40446 49.91531 49.16442 48.77646 47.82717
[49] 47.61726 47.55704
> signalline.hw.zzz$mea
Time Series:
Start = c(2019, 24)
End = c(2020, 21)
Frequency = 52
 [1] 26.90881 28.04826 28.80407 29.43689 30.62706 32.15412 34.10356 36.49621 38.66832 40.06318 41.41322 42.79699 44.01075 45.54035 47.30791 48.38982
[17] 48.96591 49.13847 49.15698 49.48779 50.23601 50.89349 51.59391 52.07634 52.42169 52.82657 52.79963 52.66290 52.77988 50.89833 50.31528 50.24711
[33] 50.54337 51.71750 53.03538 54.30772 55.18368 55.69693 56.14670 56.24587 56.25558 56.24122 56.31354 56.16296 55.75530 55.28083 54.83427 54.29766
[49] 53.83797 53.46852
```



From above forecast plot, it can be observed that the data for training is fitted well but the model is not able to forecast validation data accurately. Hence the forecast for validation data is not good.

```
> round(accuracy(macdline.hw.zzz, macdline.valid),3)
                 ME   RMSE    MAE    MPE    MAPE  MASE  ACF1 Theil's U
Training set  0.072  2.299  1.721 10.321  23.729 0.065 0.001        NA
Test set     -9.469 31.384 22.760  4.784 154.752 0.865 0.945     4.495
> round(accuracy(signalline.hw.zzz, signalline.valid),3)
                  ME   RMSE    MAE      MPE    MAPE  MASE  ACF1 Theil's U
Training set   0.045  1.120  0.666   28.047  33.624 0.030 0.015        NA
Test set     -11.460 25.765 19.712 -498.372 875.177 0.873 0.971    12.847
```

HW Model (A, Ad, N)	RMSE	MAPE	Theil's U
MACDLine	31.384	154.752	4.495
SignalLine	25.765	875.177	12.847

Also, from below accuracy measures it can be seen that for the test set **MAPE** of MACDLine is (154.752) and SignaLine is (875.177), **RMSE** is (31.384) and (25.765) respectively. **Theil's U** is also > 1, which clearly indicates that the Naive Forecast model will be able to forecast better than this model.

So, it is concluded that **<u>Holt model is not good</u>** for our forecast and will develop another model.

6.3. MODEL 3: ARIMA MODEL

Autoregressive Integrated Moving Average (ARIMA) models can perform on data with level, trend and seasonality. As our data does not contain seasonality, non-seasonal ARIMA is selected to predict validation data. ARIMA(3,0,1) is applied to both MACD and Signal line which means order 3 autoregressive model **AR(3)**, order 0 differencing incorporate trend in our model and order 1 moving average **MA(1)** for error lags. Reason of choosing value p=3(AR(3)) in the model is because of the AR(3) model which was incorporated before in the two level forecast and AR(3) in that model uses significance of all lags.

```
> summary(macd.train.arima.seas)
Series: macdline.train
ARIMA(3,0,1) with non-zero mean

Coefficients:
         ar1     ar2      ar3     ma1     mean
      0.8374  0.6749  -0.5692  0.8778  18.1628
s.e.  0.1212  0.1971   0.1026  0.0944   5.7191

sigma^2 estimated as 6.084:  log likelihood=-415.44
AIC=842.87   AICc=843.36   BIC=862

Training set error measures:
                     ME     RMSE      MAE     MPE     MAPE       MASE       ACF1
Training set 0.06873495 2.431851 1.837178 9.11687 21.48782 0.06983323 -0.1251236
```

Equation of ARIMA model for MACDLine:

$$\mathbf{y_t - y_{t-1} = 0.8374(y_{t-1} - y_{t-2}) + 0.6749(y_{t-2} - y_{t-3}) - 0.5692(y_{t-3} - y_{t-4}) - 0.8778\varepsilon_{t-1}}$$

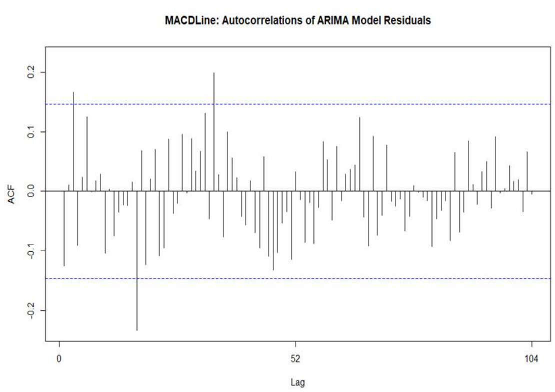
```
> summary(signal.train.arima.seas)
Series: signalline.train
ARIMA(3,0,1) with non-zero mean

Coefficients:
         ar1      ar2     ar3      ma1     mean
      2.5746  -2.2178  0.6369  -0.2412  17.9311
s.e.  0.1424   0.2770  0.1374   0.1883   4.1279

sigma^2 estimated as 0.2384:  log likelihood=-128.2
AIC=268.39   AICc=268.88   BIC=287.52

Training set error measures:
                    ME      RMSE       MAE       MPE     MAPE       MASE       ACF1
Training set 0.01005418 0.4814332 0.3680144 -1.900423 7.115434 0.01630552 0.01568876
```

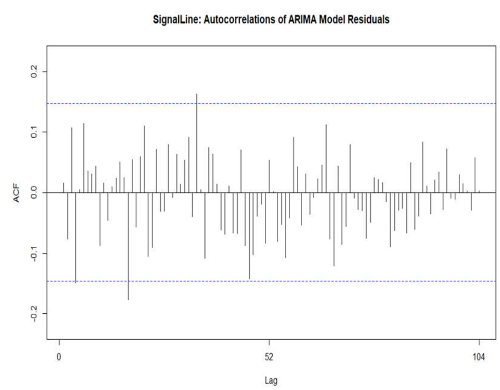
The equation of ARIMA model for SignalLine:

$$\mathbf{y_t - y_{t-1} = 2.5746(y_{t-1} - y_{t-2}) - 2.2178(y_{t-2} - y_{t-3}) + 0.6369(y_{t-3} - y_{t-4}) - 0.2412\varepsilon_{t-1}}$$

After looking at the summary, point forecast is done on validation data for both MACD and Signal Line to check the training of a model and how it predicts on validation data.

ACF was applied on ARIMA model's residuals to check the significance of lags after applying the ARIMA model.





For the MACD and Signal line, it is observed that maximum lags except 3 lags are in between threshold lines. So the ARIMA model incorporates significance of lags in its model which helps the model to perform better.

```
> round(accuracy(macd.train.arima.seas.pred, macdline.valid), 3)
                 ME   RMSE    MAE    MPE   MAPE MASE   ACF1 Theil's U
Training set  0.069  2.432  1.837  9.117 21.488 0.07 -0.125        NA
Test set     16.014 32.287 28.670 56.026 97.136 1.09  0.935     2.001
> round(accuracy(signal.train.arima.seas.pred, macdline.valid), 3)
                 ME   RMSE    MAE    MPE   MAPE  MASE  ACF1 Theil's U
Training set  0.010  0.481  0.368 -1.900  7.115 0.016 0.016        NA
Test set     17.111 31.830 28.671 58.361 92.185 1.270 0.934     1.845
```

Model	**RMSE**	**MAPE**	**Theil's U**
MACD ARIMA on Training set	2.432	21.488	NA
MACD ARIMA on Test set	32.287	97.136	2.001
Signal ARIMA on Training set	0.481	7.115	NA
Signal ARIMA on Test set	31.830	92.185	1.845

From the above table, it is observed that both line training is done well with MAPE 21.488 for MACD and MAPE 7.115 for Signal line. The test set does not perform well where MAPE is 97.136 for MACD and 92.185 for Signal line.

6.4. **MODEL 4: AUTO ARIMA MODEL**

After analyzing the ARIMA model with non-seasonality components, the idea was to verify the model for optimal values of model parameters with the Auto ARIMA model also. So, an optimal ARIMA model is generated with automatic selection of (p,d,q)(P,D,Q) parameters using the auto.arima() function.

After developing Auto ARIMA model on **macdline** training data below is the summary.

```
> summary(macd.train.auto.arima)
Series: macdline.train
ARIMA(3,0,2) with non-zero mean

Coefficients:
         ar1    ar2      ar3     ma1     ma2     mean
      1.1765  0.345  -0.5673  0.4125  -0.332  18.2961
s.e.  0.1717  0.282   0.1283  0.1761   0.094   4.0947

sigma^2 estimated as 5.789:  log likelihood=-410.58
AIC=835.16   AICc=835.81   BIC=857.47

Training set error measures:
                    ME     RMSE      MAE      MPE     MAPE       MASE       ACF1
Training set 0.0434947 2.365384 1.796012 6.238635 19.35259 0.06826847 -0.0114908
```

As per above summary, ARIMA model developed is **(3,0,2)** for order components. The model consists of order 3 AutoRegressive model AR(3) and order 2 Moving Average MA(2) for error lags. However, there is no differencing parameter to remove trend components.

Based on coefficients values of ar1, ar2 and ar3 for AR(3) and ma1, ma2 for MA(2) model the equation of the model is: $\boldsymbol{y_t - y_{t-1} = 18.2961 + 1.1765\ (y_{t-1} - y_{t-2}) + 0.345\ (y_{t-2} - y_{t-3}) - 0.5673\ (y_{t-3} - y_{t-4}) + 0.4121\ e_{t-1} - 0.332\ e_{t-2}}$

Below is the summary of Auto ARIMA model for 2nd time series signalline:

```
> summary(signal.train.auto.arima)
Series: signalline.train
ARIMA(2,1,2)

Coefficients:
         ar1      ar2      ma1      ma2
      1.8885  -0.9262  -0.5137  -0.2752
s.e.  0.0333   0.0312   0.0853   0.0841

sigma^2 estimated as 0.2357:  log likelihood=-123.77
AIC=257.54   AICc=257.89   BIC=273.45

Training set error measures:
                     ME      RMSE       MAE      MPE     MAPE       MASE         ACF1
Training set 0.02528533 0.4786784 0.3620062 3.019117 6.348028 0.01603932 -0.007517503
```

From the above summary, it can be seen that, the ARIMA model developed for signaline time series is **(2,1,2)** and its different then macdline model as both time series are individual. The model contains order components as order 2 AutoRegressive model AR(2), order 1 differencing to remove linear trend and order 2 Moving Average MA(2) for error lags.

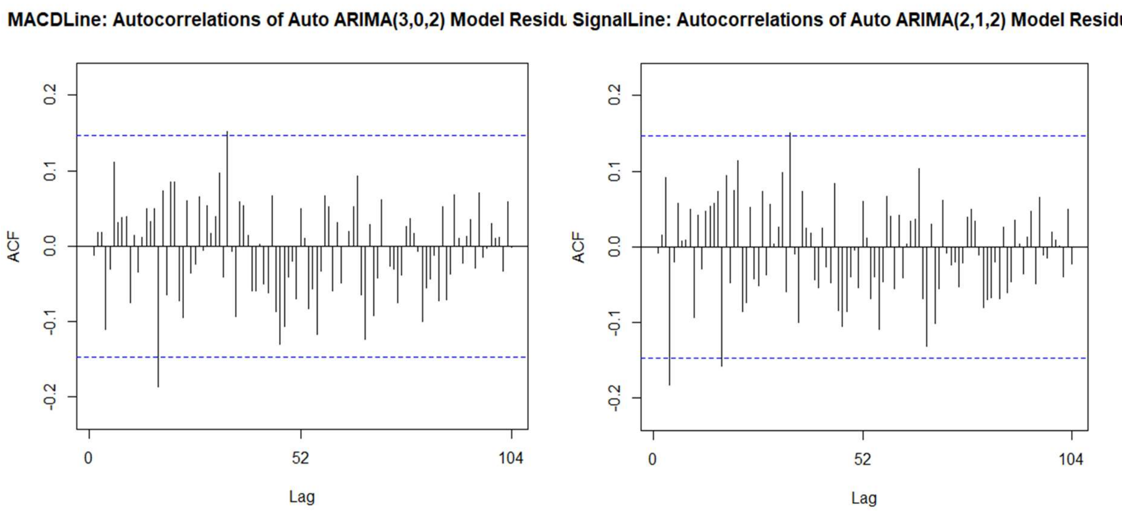
Based on coefficients values of ar1, ar2 for AR(2) and ma1, ma2 for MA(2) model the equation of the model is: $\boldsymbol{y_t - y_{t-1} = 1.8885\ (y_{t-1} - y_{t-2}) - 0.9262\ (y_{t-2} - y_{t-3}) - 0.5137\ e_{t-1} - 0.2752\ e_{t-2}}$

As it was concluded initially above seasonality, the Auto ARIMA model has also not developed a model for seasonality components.

After developing models for both time series on training data, it was predicted the point forecast of validation period for both time series.

```
> macd.train.auto.arima.pred$mean
Time Series:
Start = c(2019, 24)
End = c(2020, 21)
Frequency = 52
 [1] 40.55826 39.89650 38.95061 37.41875 35.66559 33.61111 31.45819 29.21105 26.99006 24.82317 22.78240 20.89384 19.19719 17.70727 16.44043 15.39852
[17] 14.58089 13.97818 13.57812 13.36336 13.31459 13.41010 13.62748 13.94385 14.33688 14.78510 15.26856 15.76901 16.27031 16.75848 17.22184 17.65102
[33] 18.03886 18.38035 18.67245 18.91389 19.10498 19.24738 19.34388 19.39812 19.41444 19.39762 19.35268 19.28474 19.19885 19.09986 18.99230 18.88034
[49] 18.76766 18.65749
> signal.train.auto.arima.pred$mean
Time Series:
Start = c(2019, 24)
End = c(2020, 21)
Frequency = 52
 [1] 33.55343 34.85940 35.71157 36.11134 36.07706 35.64206 34.85232 33.76375 32.43941 30.94658 29.35390 27.72871 26.13462 24.62935 23.26303 22.07685
[17] 21.10217 20.36009 19.86137 19.60682 19.58801 19.78824 20.18380 20.74537 21.43956 22.23043 23.08106 23.95501 24.81766 25.63734 26.38638 27.04178
[33] 27.58577 28.00609 28.29606 28.45436 28.48477 28.39557 28.19897 27.91028 27.54718 27.12884 26.67508 26.20561 25.73926 25.29337 24.88321 24.52159
[49] 24.21855 23.98116
```

Correlogram using acf(with lag.max = 104) was also verified for residuals of both time series and from the plot it is observed that almost all lags are now within upper and lower bounds of significance levels and hence, it can be concluded that there is only noise present in the data.



So the accuracy of models is measured as below:

For macdline, MAPE is (84.798) and RMSE is (31.553) and for signaline MAPE is (90.879) and RMSE is (28.188).

```
> round(accuracy(macd.train.auto.arima.pred, macdline.valid), 3)
                 ME   RMSE    MAE    MPE   MAPE  MASE   ACF1 Theil's U
Training set  0.043  2.365  1.796  6.239 19.353 0.068 -0.011        NA
Test set     17.835 31.553 28.019 58.399 84.798 1.065  0.928     1.617
> round(accuracy(signal.train.auto.arima.pred, macdline.valid), 3)
                 ME   RMSE    MAE    MPE   MAPE  MASE   ACF1 Theil's U
Training set  0.025  0.479  0.362  3.019  6.348 0.016 -0.008        NA
Test set     11.766 28.188 24.674 45.869 90.879 1.093  0.930     2.022
```

Model	RMSE	MAPE	Theil's U
MACD Auto ARIMA on Training set	2.365	19.353	NA
MACD Auto ARIMA on Test set	31.553	84.798	1.617
Signal Auto ARIMA on Training set	0.479	6.348	NA
Signal Auto ARIMA on Test set	28.188	90.879	2.022

From the above table, it is observed that both line training is done well with MAPE 19.353 for MACD and MAPE 6.348 for Signal line. The test set does not perform well where MAPE is 84.798 for MACD and 90.879 for Signal line.

7. COMPARISON OF MODELS

After developing 4 different models on training data and predicting the forecast on validation data of both time series **macd line** and **signal line**, below is the comparison of all 4 models.

MACDLine Accuracy:

```
> round(accuracy(macdline.2level.valid.pred, macdline.valid),3)
              ME   RMSE    MAE     MPE    MAPE  ACF1 Theil's U
Test set -39.324 55.042 48.563 -57.784 232.981 0.959     5.876
> round(accuracy(macdline.hw.ZZZ$mean, macdline.valid),3)
             ME   RMSE   MAE   MPE    MAPE  ACF1 Theil's U
Test set -9.469 31.384 22.76 4.784 154.752 0.945     4.495
> round(accuracy(macd.train.arima.seas.pred$mean, macdline.valid), 3)
             ME   RMSE   MAE    MPE   MAPE  ACF1 Theil's U
Test set 16.014 32.287 28.67 56.026 97.136 0.935     2.001
> round(accuracy(macd.train.auto.arima.pred$mean, macdline.valid), 3)
             ME   RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 17.835 31.553 28.019 58.399 84.798 0.928     1.617
```

SignalLine Accuracy:

```
> round(accuracy(signalline.2level.valid.pred, signalline.valid),3)
             ME   RMSE    MAE      MPE    MAPE ACF1 Theil's U
Test set -19.19 32.467 29.308 -456.098 749.587 0.97     11.23
> round(accuracy(signalline.hw.ZZZ$mean, signalline.valid),3)
             ME   RMSE    MAE      MPE    MAPE  ACF1 Theil's U
Test set -11.46 25.765 19.712 -498.372 875.177 0.971    12.847
> round(accuracy(signal.train.arima.seas.pred$mean, macdline.valid), 3)
             ME  RMSE    MAE    MPE   MAPE  ACF1 Theil's U
Test set 17.111 31.83 28.671 58.361 92.185 0.934     1.845
> round(accuracy(signal.train.auto.arima.pred$mean, macdline.valid), 3)
             ME   RMSE    MAE    MPE   MAPE ACF1 Theil's U
Test set 11.766 28.188 24.674 45.869 90.879 0.93     2.022
```

TIME SERIES	MODELS	MODEL PARAMETER	RMSE	MAPE
MACDLine	two-level-regression	trend^3 + AR(3)	55.042	232.981
	Auto HW	(A, Ad, N)	31.384	154.752
	ARIMA	(3,0,1)	32.287	97.136
	Auto ARIMA	(3,0,2)	31.553	84.798
SignalLine	two-level-regression	trend^3 + AR(3)	32.467	749.587
	Auto HW	(A, Ad, N)	25.765	875.177
	ARIMA	(3,0,1)	31.83	92.185
	Auto ARIMA	(2,1,2)	28.188	90.879

From the above table, it can be concluded that the MAPE of **Auto ARIMA** model for both time MACDLine is (**84.798**) and SignalLine is (**90.879**) series is comparatively low and better than all other models. The RMSE of MACDLine for Holt's Model is (**31.384**) and for Auto ARIMA is (**31.553**) while RMSE of SignalLine for Holt's Model is (**25.765**) and for Auto ARIMA is (**28.188**). Though the RMSE is lower of Holt's Model compared to Auto ARIMA but with comparison of MAPE, Auto ARIMA wins the battle.

So, it is concluded that Auto ARIMA is the recommended model to implement when forecasting MACD technical indicator for stock trading decisions.

Now, finally Auto ARIMA model is applied on the entire data set of both time series.

8. IMPLEMENT FORECAST - AUTO ARIMA ON ENTIRE DATA

Finally, in order to create a forecast, a model must be developed on the entire data set.

So, Auto ARIMA model was applied on both entire time series and below are the results for MACDLine and SignalLine:

MACDLine:

```
> summary(macd.auto.arima)
Series: macdline.ts
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2    mean
      1.7537  -0.7923  23.837
s.e.  0.0407   0.0412   5.092

sigma^2 estimated as 9.311:  log likelihood=-581.46
AIC=1170.93   AICc=1171.11   BIC=1184.66

Training set error measures:
                     ME     RMSE      MAE      MPE     MAPE       MASE        ACF1
Training set 0.04078575 3.031327 2.221028 1.987432 18.67099 0.06969743 0.006948149
```

The ARIMA model for MACDLine time series is ARIMA(2,0,0) with coefficients consisting of AR1 coefficient and AR2 coefficient with value as 1.7537 and -0.7923 respectively.

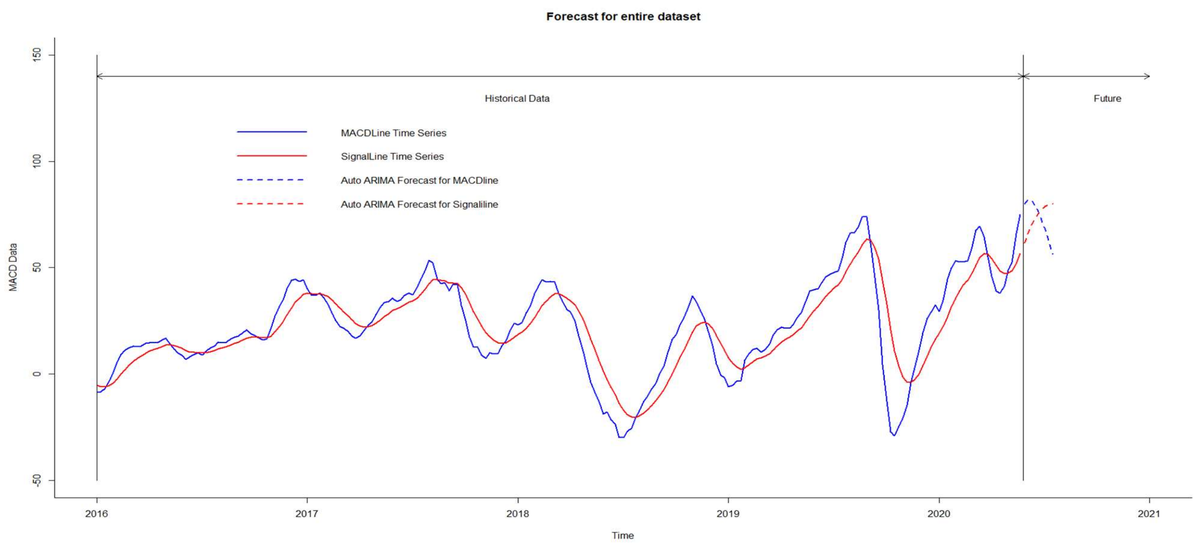
The ARIMA model for SignalLine time series is ARIMA(2,1,0) with coefficients consisting of AR1 coefficient and AR2 coefficient with value as 1.6287 and -0.7165 respectively.

SignalLine:

```
> summary(signal.auto.arima)
Series: signalline.ts
ARIMA(2,1,0)

Coefficients:
         ar1      ar2
      1.6287  -0.7165
s.e.  0.0459   0.0460

sigma^2 estimated as 0.3816:  log likelihood=-214.56
AIC=435.12   AICc=435.23   BIC=445.41

Training set error measures:
                     ME      RMSE     MAE    MPE     MAPE       MASE        ACF1
Training set 0.03074465 0.6136745 0.45204 3.2605 6.858592 0.01679098 -0.04178863
```

Below is the point forecast applied for future 8 weeks of data for MACDLine and SignalLine time series, which will help the trader to decide their position on BUY, SELL or HOLD the stock of Alphabet Inc. (GOOGLE) as explained in the introduction of MACD usage.

```
> final.forecast.df
  Forecast for MACDLine Forecast for SignalLine
1              80.09899                61.57893
2              81.95589                66.37750
3              81.17996                70.63181
4              78.34794                74.12259
5              73.99630                76.75977
6              68.60887                78.55378
7              62.60906                79.58613
8              56.35602                79.98213
```



The Accuracy measures of Auto ARIMA model with naïve forecast model (for the entire dataset) for both time series are presented below:

```
> #1. Naive
> #2. Auto ARIMA
> round(accuracy(naive(macdline.ts)$fitted, macdline.ts),3)
            ME  RMSE   MAE    MPE   MAPE  ACF1 Theil's U
Test set 0.367 4.884 3.516 19.167 43.402 0.759         1
> round(accuracy(macd.auto.arima.pred$fitted, macdline.valid), 3)
            ME RMSE   MAE    MPE  MAPE  ACF1 Theil's U
Test set 0.813  4.5 3.474 -8.258 16.56 0.174     0.247
>
> round(accuracy(naive(signalline.ts)$fitted, signalline.ts),3)
            ME RMSE MAE    MPE    MAPE  ACF1 Theil's U
Test set 0.271 2.76 2.1 90.611 107.351 0.942         1
> round(accuracy(signal.auto.arima.pred$fitted, signalline.ts), 3)
            ME  RMSE   MAE   MPE  MAPE   ACF1 Theil's U
Test set 0.031 0.614 0.452 3.261 6.859 -0.042     0.038
```

Time Series	Model	Model Parameters	MAPE	RMSE
MACDLine	Naïve	-	43.402	4.884
	Auto ARIMA	(2,0,0)	**16.56**	**4.5**
SignalLine	Naïve	-	107.351	2.76
	Auto ARIMA	(2,1,0)	**6.859**	**0.614**

After applying the model on entire data, the accuracy metrics are improved a lot with MAPE for MACDLine is 16.56% and SignalLine is only 6.859% with very low RMSE values as 4.5% and 0.614% respectively. Hence, we can conclude that the Auto ARIMA model as the best model for forecasting next 8 weeks of MACD indicator values.

CONCLUSION

After analyzing both time series data, it is concluded that these time series data are following trend which is forecasted using Auto ARIMA model.

As this data is till 09 Nov 2020, this forecast could be tested with actual data and it was observed that the forecast for 3rd and 4th week of Nov were very close to actual values which was also learned in course work that short time future forecast has high accuracy then longer time forecast values.

As stock market world is highly volatile and price action has high impact on MACD indicator, it is recommended to re-evaluating forecasting model after every 4 weeks for better accuracy.

This forecasting model of Auto ARIMA is developed for our selected security/stock of Alphabet Inc (GOOGLE). However, this model can be used for other stocks as well, but a check has to be developed that if an accuracy of overall model pass X% then only the forecasting result will be use otherwise forecasting result should not be use. Hence, this model is generic model for predicting MACD technical indicator.

REFERENCES

1. Rob J Hyndman and George Athanasopoulos, Forecasting Principles and Practice: - Chapter 9 ARIMA models
2. Yahoo Finance: https://finance.yahoo.com/
3. Jason Fernando, Moving Average Convergence Divergence (MACD): https://www.investopedia.com/terms/m/macd.asp