DSC5211C QUANTITATIVE RISK MANAGEMENT

Semester 2, AY 2018/19

FORECASTING INDIA'S STOCK INDEX

Forecasting Assessment Report

Group 12

Name

Cho Zin Tun

Toh Pei Xuan

Zhang Qian Ru

1 March 2019

1. Introduction

1.1. Motivation of problem

Stock market prices have been widely forecasted as any successful attempts could bring profits. By predicting future stock prices, one can decide on the investment strategy (especially the hedging strategy) to take such that positive yield could be derived.

Investment strategies can come in both long term and short-term strategies. Short term strategies require accurate forecast of day-to-day market movements, while Long term investment depends on the overall trend and should not be affected by day-to-day fluctuations.

Stock prices are unlikely driven by its history alone. Other macro factors affecting the whole economy, both external and internal, would likely impact the price of the index as well. By considering such factors, a better investment strategy could be formed.

Hence the aim of this project was to forecast monthly stock index using both historical data of the index and other macro factors.

1.2. Description of problem

This project focuses on the India stock market using NIFTY 50 index. The NIFTY 50 index is National Stock Exchange of India's benchmark broad based stock market index for the Indian equity market. It represents the weighted average of 50 Indian company stocks in 12 sectors and is one of the two main stock indices used in India (Wikipedia, n.d.).

Historical index of NIFTY 50 was used to forecast NIFTY 50 using Holt Winter’s Exponential Smoothing model and ARIMA model for shorter term forecasts. Other than that, a cointegration model was also built to take other macro factors into account to forecast NIFTY 50 for the long-term. Macro factors include both global growth and Indian internal indication.

2. Data

2.1. Data Source

The adjusted close price of NIFTY 50 Index (Nifty50) was used as the variable of interest in this project. Table 1 contains the other macro factors (both internal and external) that were considered.

The data of these indexes was collected from Bloomberg terminal and are monthly indexes that range from March 2014 to November 2018 (57 data points). Since variables are index values, log transformation was applied for the whole dataset.

Internal	External
Valuation	US Non-Farm Payrolls
Electricity Generation	US Retail Sales
Cement Production	US Durable Goods Order
Interest Rate	US New House Sales
Sentiment	China Manufacturing PMI
India Inflation	Global Liquidity
India Earning Outlook	Geopolitical Risk
	US 10-Year 2-Year yield Difference
	US 2-Year yield
	US 10-year Yield

Table 1: Macro Factors

2.2. Feature Selection

Feature selection was done in two steps. Indexes with missing value or had not been updated timely were first removed. Feature selection using XG Boosting model was subsequently done on the other variables to select the most significant macro indexes. The final set of factors used are given in Table 2.

Index	Description
NFP TCH	**US Non-Farm Payrolls** US Employees on Non-farm Payrolls Total Month over Month Net Change
RSAOREST	**US Retail sales** US retail sales full-service restaurants
GVLQUSD	Global Liquidity US Government Security Liquidity Index
USYC2Y10	**US 10-year 2-year yield** Market matrix us sell 2-year & buy 10-year bond yield spread
TACKINIP	**India Inflation** Thumbtack Indiana Inflation Pass-Through Index
USGG10YR	**US 10-year 2-year yield** US Generic 10-year yield
WTEMINTR	**Earning outlook** India Earning Outlook Index

Table 2: Description of macro factors that was used

3. Model

This section describes three models used to forecast the Nifty50 series. The data from March 2014 to December 2018 was split into 2 parts. Models were built based on data from March 2014 to Dec 2017 (46 data points). These models were then used to do out-of-sample forecasting for the 2018 data (11 periods) and the forecasted values were compared against the Jan 2018 to Nov 2018 (11 points) actual data available.

3.1. Holt-Winters Exponential Smoothing model

Exponential Smoothing model is based on weighted averages of past observations. These weights reduce exponentially as the observations get older. Holts or Holts-Winter Exponential Smoothing model can also take into account overall trends and seasonality of the series. The trend parameter measures the overall growth or decline in the series, while the seasonality parameter is useful in determining patterns for different seasons in that series. For this project, seasonality is based on each month, hence 12 was used as the number of seasons.

3.1.1. Decomposing Nifty50 series

Before implementing the Exponential Smoothing model, Nifty50 data was decomposed using a function in R to observe if there were possible trends or seasonality in the series (Figure 1). This aided in determining the type of Exponential Smoothing model that should be used.

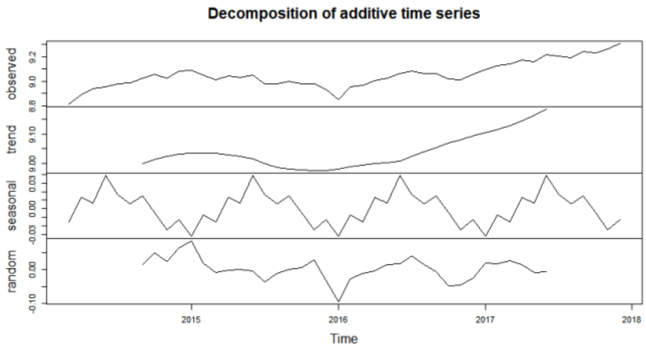

Figure 1: Nifty50 decomposed into trend seasonal and random

The plot suggested that there is trend and seasonality in this series. As such, the Holt-Winters Exponential Smoothing model with trend and seasonality (HWES) was used to model this Nifty50 series.

3.1.2. Results

Results of the HWES parameters are given in Figure 2. The value of alpha is large, suggesting that older observations have lower weights and are less important. A small beta suggested that the slope (trend) of the model only changed slightly with time. Lastly, a large seasonal parameter of 1 suggested that past seasonal component had little or no importance.

```
Smoothing parameters:
 alpha: 0.7866857
 beta : 0.04256005
 gamma: 1
```

Figure 2: Results of HWES model

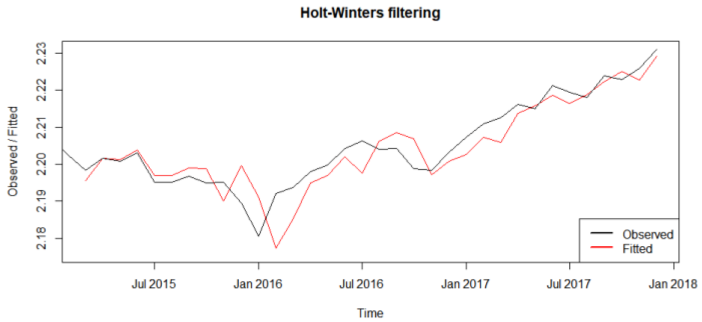

Figure 3: Plot of observed and fitted values of the HWES Model

The fitted values from the model were also plotted against the observed values as given in Figure 3. Values were fitted reasonably well when the increase and drops were small. However at large increase and drops, the fitted values seemed to be lagging behind.

3.1.3. Out-of-sample Forecasting

The plot in Figure 4 shows the actual figures and the forecast for the next 11 months using the above model (with the corresponding prediction bounds). The forecast looks reasonable when compared to the actual values. Also, the forecast is generally moving upwards, similar to the overall trend of the actual values.

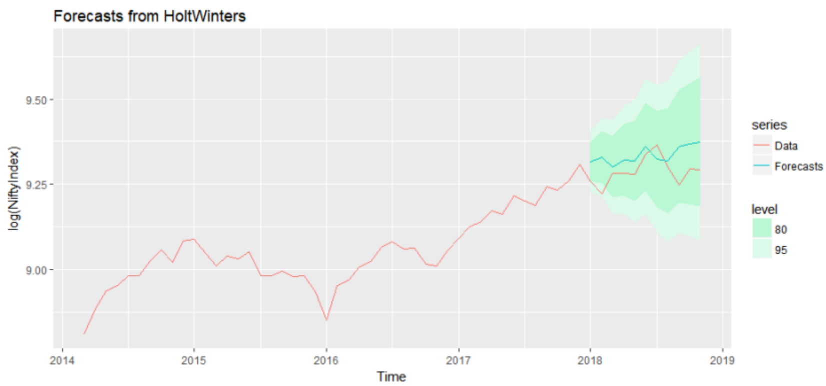

Figure 4: Plot of actual and forecasted values of the HWES Model

3.1.4. Residuals

Residuals are the actual values minus the forecasted values by the model. These residuals were plotted to check if assumptions were satisfied. These residuals have to be uncorrelated and normally distributed with zero mean and constant variance.

The autocorrelations (ACF) for the residuals were plotted (Figure 5) and suggested that the residuals were not correlated as all lags were within the significance bounds. The Box-Ljung test (Figure 6) had a p-value more than 0.05 which also suggests that residuals were uncorrelated.

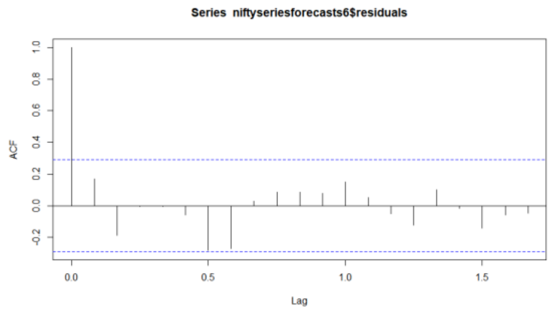

Figure 5: ACF plot of residuals of the HWES Model

```
        Box-Ljung test

data:  niftyseriesforecasts6$residuals
X-squared = 15.735, df = 20, p-value = 0.7329
```

Figure 6: Results of Box-Ljung test of the HWES Model

The residual plots and histogram of the residuals were also plotted below.

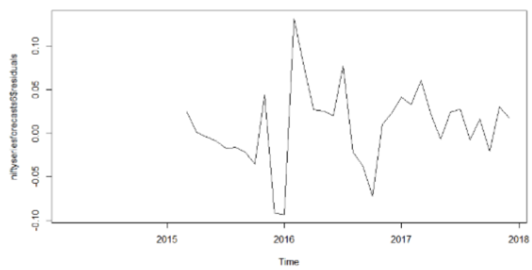

Figure 7: Residual Series Plot of the HWES Model

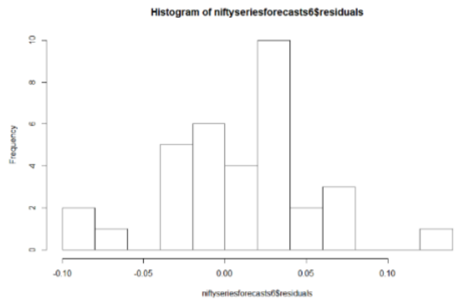

Figure 8: Histogram of Residuals of the HWES Model

Figure 7 shows the residual plots and suggest that the residuals do not follow any particular trends. Also, the histogram of the residuals in Figure 8 centred around 0 (slightly above 0) and showed a shape similar to a normal distribution.

In conclusion, the residuals of this HWES model satisfies various underlying assumption and suggest that the HWES forecast model is valid.

3.2. ARIMA model

Autoregressive integrated moving average (ARIMA) model is a statistical model for forecasting and analysing time series data that takes into account correlations of historical data. An ARIMA(p,d,q) model consists of 3 components. The Autoregressive component represented by p is the number of periods of lagged values to be included in the model. d represents the order of differencing transformations done to obtain a stationary time series. The moving average component represented by q, is the number of periods of lagged error component to be included in the model.

3.2.1. ARIMA Model parameters Selection

As ARIMA modelling can only be performed on stationary time series, the data would have to be checked to ensure assumptions could be satisfied. The Nifty50 data showed overall increasing trend based on time series data plot in Figure 1. Therefore, difference had to be taken for Nifty50 to attempt to make the series stationary.

A First order difference was first taken for this series. The Augmented Dickey–Fuller (ADF) test was done to test for stationarity and results can be found in Figure 9.

```
        Augmented Dickey-Fuller Test

data:  niftyseriesdiff1
Dickey-Fuller = -3.3555, Lag order = 3, p-value = 0.07556
alternative hypothesis: stationary
```

Figure 9: Dickey-Fuller Test of First Order Difference

The P-value of the test is more than 0.05 and hence the null hypothesis (that the series is non-stationary) cannot be rejected. This suggest that the first order difference is still non-stationary. As such, the second order difference was taken. The ADF test was similarly done and P-value of the test was less than 0.05. Also, ACF and partial autocorrelation function (PACF) were plotted (Figure 10).

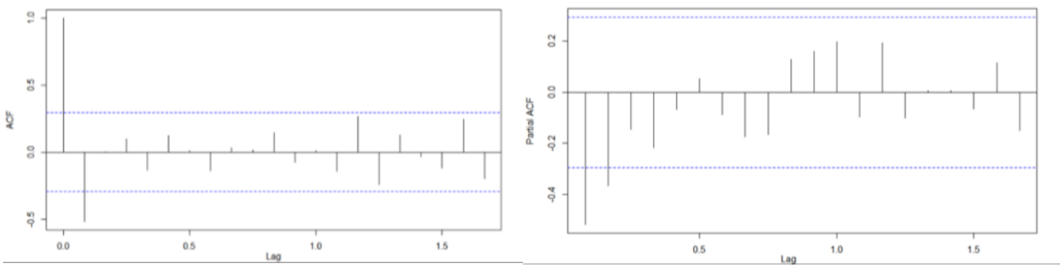

Figure 10: ACF and PACF plot of second order difference

The plots suggest that this series is stationary. The first two lags of ACF and PACF plots were significantly different from zero. This suggests that various ARIMA model from AR(0) to AR(2) and MA(0) to MA(2) could be possible. Root mean square error (RMSE) for each of the 9 models (using second order difference) were derived (Table 3).

ARIMA(p,d,q)	0,2,0	1,2,0	2,2,0	0,2,1	1,2,1	2,2,1	0,2,2	1,2,2	2,2,2
RMSE	0.0521	0.0445	0.0409	0.0372	0.0381	0.0380	0.0372	0.0373	0.0373

Table 3: RMSE of various ARIMA models

ARIMA (0,2,1) and ARIMA (0,2,2) had similar RMSE. Since ARIMA (0,2,1) requires less parameters than ARIMA (0,2,2), ARIMA (0,2,1) was chosen to be used as the ARIMA model for this series.

3.2.2. Results

Results of the ARIMA(0,2,1) model is given in Figure 11. The MA(1) coefficient is significant (P-value less than 0.05) and suggests that this parameter cannot be dropped from the model.

```
    Estimate Std. Error z value  Pr(>|z|)
ma1 -0.99997    0.15181 -6.5869 4.492e-11 ***
```

Figure 11: coefficient of Arima(0,2,1) model

3.2.3. Out-of-sample Forecasting

The plot in Figure 12 shows the actual figures and the out-of-sample forecast for the next 11 months using the above model (with the corresponding prediction bounds). The forecast showed a general upward trend. When compared to the actual values, the actual values had fluctuations, while the forecasted values do not. In general, the forecasted trend was in the same direction as the actual trend.

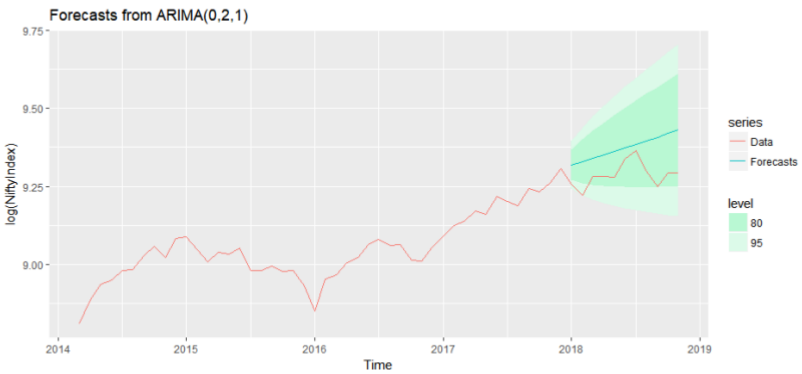

Figure 12: Forecast Result for ARIMA(0,2,1) model

3.2.4. Residuals

Similar to section 3.1.4, the residuals have to be uncorrelated and normally distributed with zero mean and constant variance.

The ACF for the residuals were plotted (Figure 13) and suggested that the residuals were not correlated as all lags were within the significance bounds. The Box-Ljung test (Figure 14) had a p-value more than 0.05 which also suggest that residuals were not correlated.

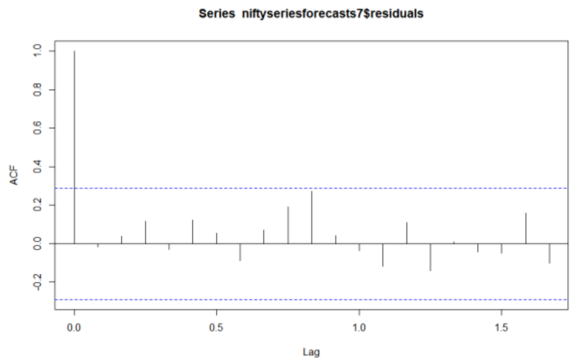

Figure 13: ACF plot of residuals of the ARIMA(0,2,1) Model

```
        Box-Ljung test

data:  niftyseriesforecasts7$residuals
X-squared = 15.922, df = 20, p-value = 0.7215
```

Figure 14: Results of Box-Ljung test of the ARIMA(0,2,1) Model

The residual plots and histogram of the residuals were also plotted below.

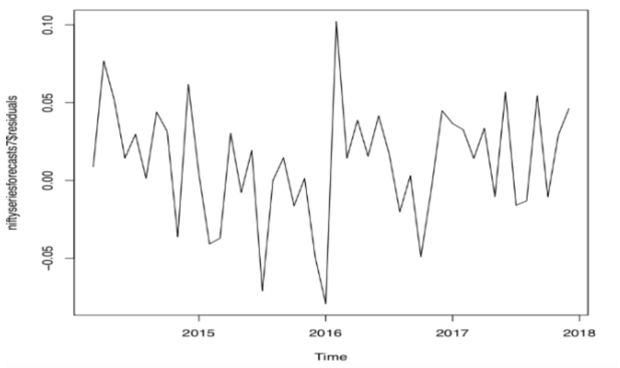

Figure 15: Residual Series Plot of the ARIMA(0,2,1) Model

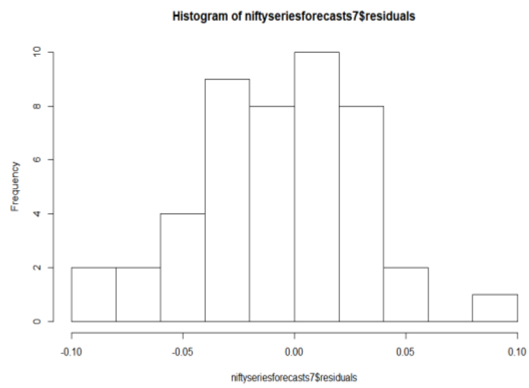

Figure 16: Histogram of Residuals of the ARIMA(0,2,1) Model

Figure 15 shows the residual plots and suggest that the residuals do not follow any particular trends. Also, the histogram of the residuals in Figure 16 centred around 0 and showed a shape similar to a normal distribution.

In conclusion, the residuals of this ARIMA(0,2,1) model satisfies various underlying assumptions and suggest that the ARIMA(0,2,1) forecast model is valid.

3.3. Model Comparison Between Univariate Models

In this project, Holt-Winter exponential smoothing model and Arima model ARIMA(0,2,1) were used to model univariate time series for short-term forecasting. Both series were compared to determine which model was a better one.

Both series were used to derive out of sample forecast (11 data points from Jan 2018 to Nov 2018) and were plotted together (Figure 17).

Figure 17: Actual Data Versus Forecasted Data for Arima(0,2,1) and HWES

HWES model showed forecasts that were closer to the actual data based on the plot. Also, the mean squared error of these 11 points for HWES model, 0.046, is smaller than that of the ARIMA(0,2,1) model, 0.102.

This comparison showed that the Holt-Winter Exponential Smoothing model was a better short-term model as it provided better out-of-sample forecast. This could be due to the seasonal component in the HWSE model that could take into account fluctuations of this time series.

3.4. Cointegration Test

To identify long-run relationships between Nifty50 and macro factors described in the Section 2.2, a multivariate model was explored. Spurious regression problem could occur in time series when non-stationary variables are regressed on each other. Thus, instead of directly regressing Nifty50 with other non-stationary macro variables, cointegration was used to avoid spurious regression problem. Cointegration provides some linear combinations between variables, which represents long-run equilibrium relation tying the individual variables together. It helps to identify the degree to which two variables are sensitive to the same average impulse over a specific period and provides the information if the distance between them remains the same over time.

3.4.1. Non-stationary Checking

ADF, ACF and PACF tests were performed to check if the series are non-stationary. The tests showed that macro variables (described in section 2.2) RSAOREST, GVLQUSD, USYC2Y10, USGG10YR and WTEMINTR were non-stationary (P-Value of ADF test is more than 0.05). Therefore, these were used as endogenous variables in the cointegration testing model.

Index	NFP.TCH	RSAOREST	GVLQUSD	USYC2Y10	USGG10YR	WTEMINTR	TACKINIP
P-Val	0.0100	0.0706	0.2893	0.3734	0.4786	0.9574	0.04961

Table 4: P-value of ADF test for macro indexes

3.4.2. Johansen Cointegration Test

Johansen cointegration test was implemented to detect multiple cointegrating vectors. The Johansen test performs the testing for cointegration by iterating the number of independent linear combinations (K) for an N time series variable set to achieve a stationary process. Out of two forms of Johansen test, trace test was used to determine the presence of cointegration between variables. The null hypothesis for trace test examines if K (number of linear combinations) is equal to given value (K_0) and the alternative hypothesis is K greater than K_0.

3.4.3. Johansen Cointegration Test – Implementation

EView was used to execute the Johansen Cointegration Test.

Date: 03/02/19 Time: 13:27
Sample (adjusted): 2014M05 2018M11
Included observations: 55 after adjustments
Trend assumption: No deterministic trend (restricted constant)
Series: LOG(NIFTY INDEX ADJUST CLOSE) LOG(GVLQUSD INDEX) LOG(RSAOREST INDEX) LOG(US
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.548088	118.9321	103.8473	0.0035
At most 1	0.410859	75.24736	76.97277	0.0672
At most 2	0.288393	46.14742	54.07904	0.2100
At most 3	0.243688	27.43482	35.19275	0.2674
At most 4	0.183175	12.07326	20.26184	0.4423
At most 5	0.017037	0.945093	9.164546	0.9583

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Figure 18: Johansen Cointegration Trace Test

From Figure 18, the trace test indicated that the P-value of the null hypothesis of having no cointegration equation between variables is less than 0.05 and hence can be rejected, i.e there is cointegration between variables. The result suggests that there is 1 cointegrating equation at the 0.05 confidence level. Therefore, cointegration model with 1 cointegrating equation was built. Figure 19 and Figure 20 show the long-term and short-term vector error correction estimates of the cointegration model respectively. For long-term relationship, RSAOREST, GVLQUSD and WTEMINTR played an important role while USYC2Y10 and USGG10YR were not statistically significant. However, USYC2Y10 and USGG10YR were not dropped since they had significant effect on GVLQUSD in the short-term model.

Vector Error Correction Estimates
Date: 03/02/19 Time: 13:53
Sample (adjusted): 2014M05 2017M12
Included observations: 44 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
LOG(NIFTY_INDEX_AD	1.000000
LOG(GVLQUSD_INDEX(	-0.293192 (0.03166) [-9.25941]
LOG(RSAOREST_INDE	-0.353490 (0.12358) [-2.86051]
LOG(USGG10YR_INDE	-0.013377 (0.04596) [-0.29107]
LOG(USYC2Y10_INDEX	0.023035 (0.03076) [0.74880]
LOG(WTEMINTR_INDE	-1.009987 (0.05910) [-17.0886]
C	-2.626304 (0.66530) [-3.94756]

Figure 19: Vector Error Correction Estimates (long-term)

Error Correction:	D(LOG(NIFT	D(LOG(GVL	D(LOG(RSA	D(LOG(USG	D(LOG(USY	D(LOG(WTE
CointEq1	0.209766 (0.17576) [1.19347]	2.453691 (0.52920) [4.63662]	-0.205872 (0.23790) [-0.86537]	-0.046420 (0.43270) [-0.10728]	0.457028 (0.47030) [0.97177]	0.077898 (0.07665) [1.01628]
D(LOG(NIFTY_INDEX_	-0.177092 (0.26739) [-0.66231]	-3.656643 (0.80507) [-4.54201]	0.434237 (0.36192) [1.19982]	-0.212509 (0.65828) [-0.32283]	-0.648915 (0.71547) [-0.90697]	1.157312 (0.11661) [9.92475]
D(LOG(GVLQUSD_IND	-0.022739 (0.05068) [-0.44872]	0.018317 (0.15258) [0.12005]	-0.077882 (0.06859) [-1.13545]	-0.101022 (0.12476) [-0.80974]	-0.047763 (0.13560) [-0.35224]	0.035927 (0.02210) [1.62565]
D(LOG(RSAOREST_IN	0.083315 (0.11119) [0.74929]	0.045372 (0.33479) [0.13553]	-0.671610 (0.15050) [-4.46243]	-0.115764 (0.27374) [-0.42289]	0.329167 (0.29753) [1.10634]	-0.065331 (0.04849) [-1.34726]
D(LOG(USGG10YR_IN	0.053670 (0.10485) [0.51189]	0.783881 (0.31568) [2.48314]	0.257234 (0.14191) [1.81261]	-0.142454 (0.25812) [-0.55189]	-0.216144 (0.28055) [-0.77043]	-0.003796 (0.04572) [-0.08303]
D(LOG(USYC2Y10_IND	-0.131076 (0.09771) [-1.34141]	-1.187695 (0.29421) [-4.03693]	-0.130978 (0.13226) [-0.99030]	0.184482 (0.24056) [0.76688]	0.365692 (0.26146) [1.39863]	-0.001030 (0.04261) [-0.02417]
D(LOG(WTEMINTR_IN	-0.056259 (0.12426) [-0.45277]	0.404741 (0.37412) [1.08185]	-0.129853 (0.16818) [-0.77209]	0.275642 (0.30590) [0.90108]	-0.180043 (0.33248) [-0.54151]	0.003898 (0.05419) [0.07193]
R-squared	0.050956	0.528024	0.464450	0.123334	0.084124	0.910207
Adj. R-squared	-0.102943	0.451488	0.377604	-0.018828	-0.064396	0.895646
Sum sq. resids	0.056111	0.508672	0.102799	0.340083	0.401751	0.010672
S.E. equation	0.038942	0.117251	0.052710	0.095872	0.104202	0.016983
F-statistic	0.331102	6.898980	5.347969	0.867563	0.566415	62.50966
Log likelihood	84.18822	35.68980	70.86836	44.54732	40.88120	120.7025
Akaike AIC	-3.508556	-1.304082	-2.903107	-1.706696	-1.540055	-5.168297
Schwarz SC	-3.224707	-1.020233	-2.619259	-1.422848	-1.256206	-4.884448
Mean dependent	0.009595	0.010899	0.006041	-0.002166	-0.033174	0.010853
S.D. dependent	0.037081	0.158316	0.066813	0.094982	0.101001	0.052573
Determinant resid covariance (dof adj.)		2.53E-16				
Determinant resid covariance		8.95E-17				
Log likelihood		438.3428				

Figure 20: Vector Error Correction Estimates (short-term)

Resulted Vector Autoregression (VAR) Model can be found in Figure 21.

```
VAR Model - Substituted Coefficients:
===============================
D(LOG(NIFTY_INDEX_ADJUST_CLOSE_)) = 0.209766127299*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) - 0.177092083068*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) - 0.0227392243964*D(LOG(GVLQUSD_INDEX(-1))) +
0.0833147437881*D(LOG(RSAOREST_INDEX(-1))) + 0.0536699248362*D(LOG(USGG10YR_INDEX(-1))) - 0.131075524616*D(LOG(USYC2Y10_INDEX(-1))) - 0.0562589397377*D(LOG(WTEMINTR_INDEX(-1)))

D(LOG(GVLQUSD_INDEX)) = 2.45369074891*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) - 3.65664266143*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) + 0.0183174727297*D(LOG(GVLQUSD_INDEX(-1))) +
0.0453721236118*D(LOG(RSAOREST_INDEX(-1))) + 0.783881016227*D(LOG(USGG10YR_INDEX(-1))) - 1.18769505233*D(LOG(USYC2Y10_INDEX(-1))) + 0.404740979913*D(LOG(WTEMINTR_INDEX(-1)))

D(LOG(RSAOREST_INDEX)) =  - 0.205872313779*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) + 0.434236547208*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) - 0.0778822326307*D(LOG(GVLQUSD_INDEX(-1))) -
0.671609535697*D(LOG(RSAOREST_INDEX(-1))) + 0.257234134213*D(LOG(USGG10YR_INDEX(-1))) - 0.130977848226*D(LOG(USYC2Y10_INDEX(-1))) - 0.129853483179*D(LOG(WTEMINTR_INDEX(-1)))

D(LOG(USGG10YR_INDEX)) =  - 0.0464198278803*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) - 0.212509247775*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) - 0.101022269881*D(LOG(GVLQUSD_INDEX(-1))) -
0.115763619197*D(LOG(RSAOREST_INDEX(-1))) - 0.14245444468*D(LOG(USGG10YR_INDEX(-1))) + 0.184482342812*D(LOG(USYC2Y10_INDEX(-1))) + 0.275642347198*D(LOG(WTEMINTR_INDEX(-1)))

D(LOG(USYC2Y10_INDEX)) = 0.457027619057*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) - 0.648914705918*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) - 0.0477628941944*D(LOG(GVLQUSD_INDEX(-1))) +
0.329166578298*D(LOG(RSAOREST_INDEX(-1))) - 0.216144474173*D(LOG(USGG10YR_INDEX(-1))) + 0.365691576606*D(LOG(USYC2Y10_INDEX(-1))) - 0.180043214133*D(LOG(WTEMINTR_INDEX(-1)))

D(LOG(WTEMINTR_INDEX)) = 0.0778984257786*( LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1)) - 0.293191844208*LOG(GVLQUSD_INDEX(-1)) - 0.353489862804*LOG(RSAOREST_INDEX(-1)) - 0.0133771929198*LOG(USGG10YR_INDEX(-1)) +
0.0230346332186*LOG(USYC2Y10_INDEX(-1)) - 1.00998743462*LOG(WTEMINTR_INDEX(-1)) - 2.62630350091 ) + 1.15731151469*D(LOG(NIFTY_INDEX_ADJUST_CLOSE_(-1))) + 0.0359269128356*D(LOG(GVLQUSD_INDEX(-1))) -
0.0653305638368*D(LOG(RSAOREST_INDEX(-1))) - 0.00379625047737*D(LOG(USGG10YR_INDEX(-1))) - 0.00103016235472*D(LOG(USYC2Y10_INDEX(-1))) + 0.0038976102903*D(LOG(WTEMINTR_INDEX(-1)))
```

Figure 21: VAR model

3.4.4. Robustness Test

Various robustness tests were further performed to check the stability of the model.

3.4.4.1. Impulse-Response Analysis

An impulse-response analysis was carried out to uncover the dynamic relationship between variables within VAR models. It measures the time profile of the effect of shock on the future value on a variable. A positive shock in macro variables except RSAOREST will cause Nifty50 to drop for the first 2 months before stabilising (Figure 22).

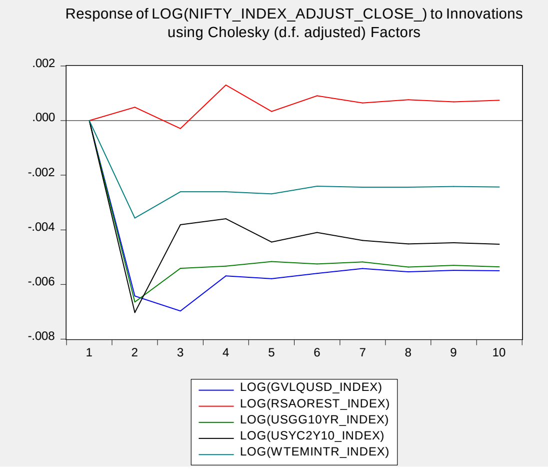

Figure 22: Impulse-Response Analysis

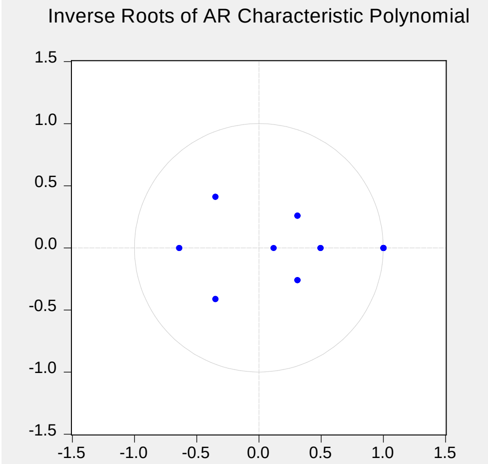

Figure 23: Inverse Roots of AR Characteristic Polynomial

3.4.4.2. AR Root Test

AR Root Test was performed as additional robustness test to check if the VAR model is stationary. Since all the roots were within or on the unit circle (Figure 23), it could be concluded that the model is stable.

3.4.4.3. Residual Checking

Residual for each series used in the model is given in Figure 24. The graphs suggested that there was no trend in the residuals and it could be concluded that residuals were stationary.

Figure 24: Residual Series Plot of cointegration test model

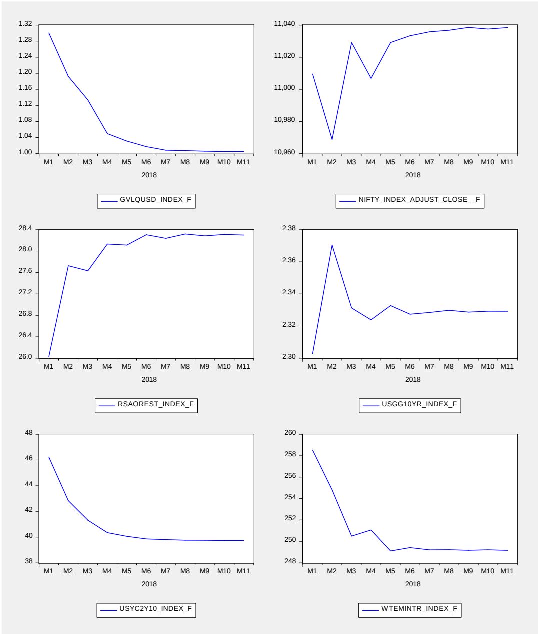

Figure 25: Out-of-sample forecast (for 2018)

3.4.5. Out-of-sample-Forecasting

Out-of-sample forecasting (11 datapoints for Jan 2018 to Nov 2018) was evaluated to determine how good the model performed. In Figure 25 (plot on the top right-hand corner), it can be seen that Nifty50 dropped sharply (by about 40 points) in February then increased sharply in March (by about 60 points) before gradually stabilising from May onwards.

The plot in Figure 26 shows the actual figures and the forecast for the next 11 months using the above model. The plot showed that there was generally a flat trend in the forecast. This was in contrast to Figure 25 as log had been taken for the plot in Figure 26. Due to the large values of the Nifty50 series, the difference (between 20 to 60 points) became negligible after taking log.

When compared to the actual values, the actual values had fluctuations, while the forecasted values do not. In general, the forecasted values do not seem to be too accurate. This may be because this model is used to identify long-run relationships and may not provide accurate forecasts for the short-run.

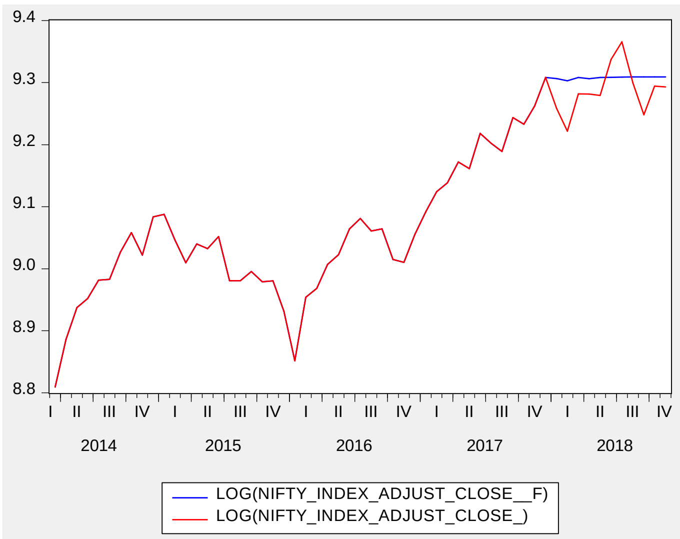

Figure 26: Forecast Result for cointegration model

3.4.6. Interpretation of model

Based on Figure 22, the result indicated that growth in US retail index would lead to positive response of India stock market. The growth in US retail index could indicate a good global economic environment and hence has a positive impact on India stocks.

Results also showed that an increase in India inflation would cause a decrease in India stock market. A possible interpretation would be that since high inflation would reduce consumers’ purchasing power, this would lead to decline in revenues and profits; and as the economy slows down, stock market performance will be negatively impacted.

An increasing "sell 2-year and buy 10-year bond yield spread" in the US market (USYC2Y10), indicates a negative outlook on the global economy which could also negatively impact India's stock prices.

Lastly, for US Non-farm payrolls, Indian earning outlook and global liquidity, Figure 22 indicates that these variables have a negative relationship against stock prices. This is in contrast to expectations and could be due to various factors that this project could be improved on (discussed in section 4).

4. Conclusion

This project built three models to forecast Nifty50 time series data. The univariate model produced overall upward trend forecasts for the short run, with some fluctuations in between. While for the long run, the cointegration model produced forecasts that were flat and stable and with little variations. Based on various residual and robustness test, the models built were stable and satisfied the various underlying assumptions.

Nevertheless, the model could be further improved from various aspects. Firstly, more macro factors could be studied together. By including more of such macro factors, they may better account for movement in the stock index. Secondly, the time window selected in this project was based on monthly data. The study can be extended further to be based on quarterly and daily basis if data based on such windows were available. This would allow different analysis and forecast to be made for different types of investment strategies. Lastly, the models were built based on 4 years data. Further studies could be done with more amount of data (e.g. 20 years data), especially if the model is for the purpose of capturing macro trends and forecasting for longer term investment.

References

Coghlan, A. (2017). *A Little Book of R For Time Series.*

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time with R.* Springer.

Wikipedia. (n.d.). Retrieved from https://en.wikipedia.org/wiki/NIFTY_50