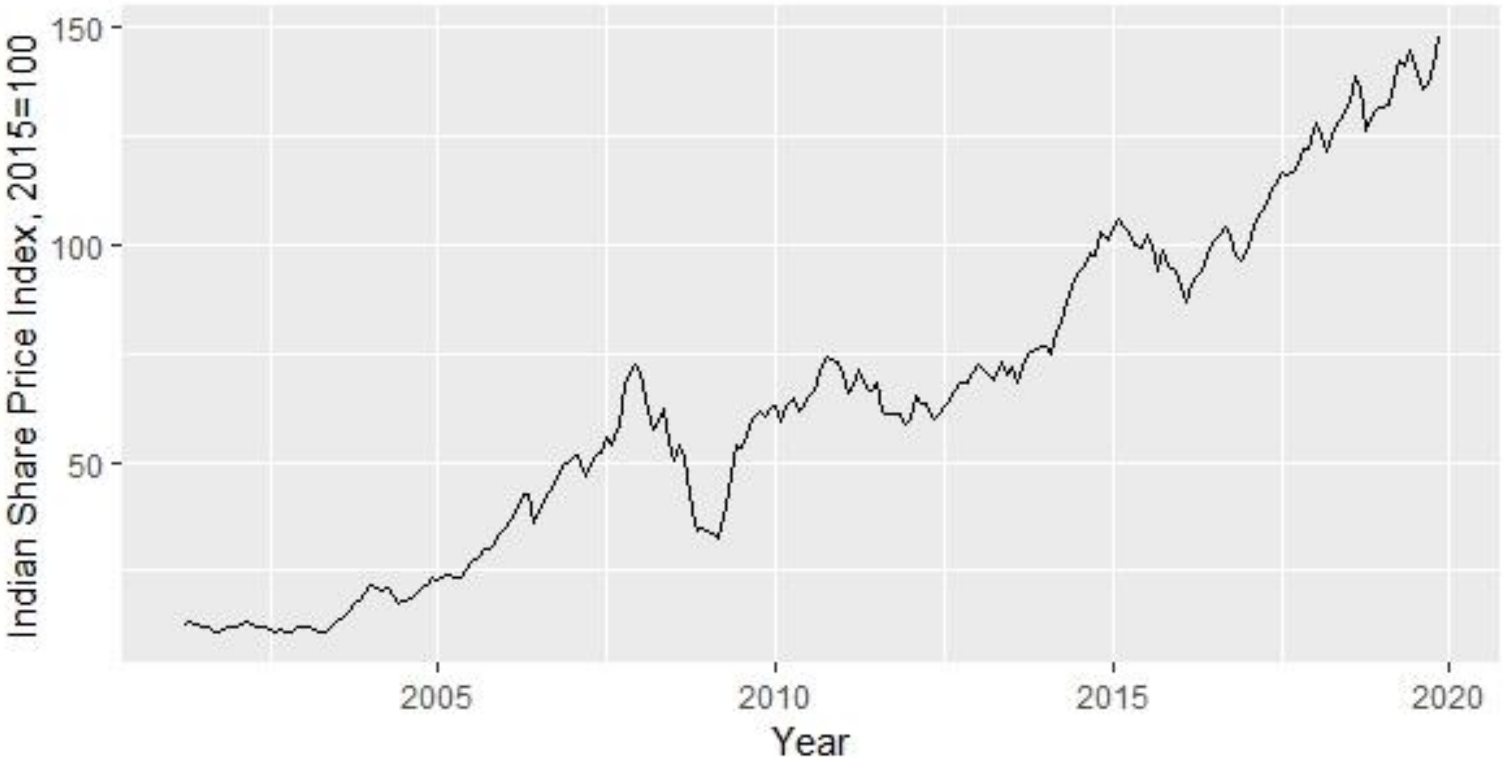
KU LEUVEN

Advanced Time Series Analysis Project

Shrikrishna Warkhedi

Masters of Artificial Intelligence

Univariate Time Series Analysis

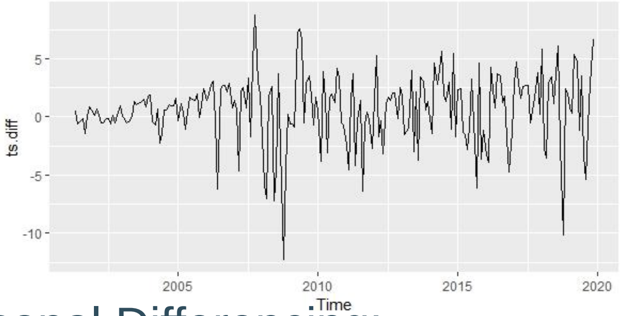


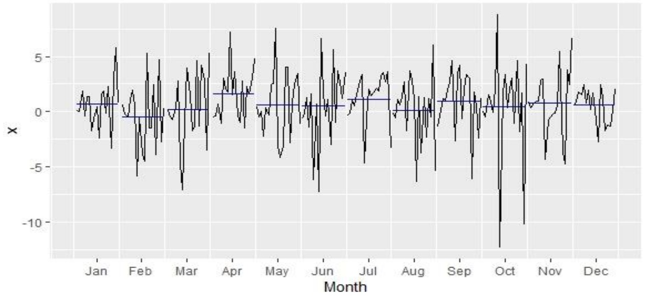
- Indian stock market indices data has been considered. Data ranges from year 2001 to 2019 at monthly level with 224 data points.
- Source: https://data.oecd.org/price/share-prices.htm
- Unit taken as 2015=100
  According to the website, data is log transformed and seasonally adjusted.

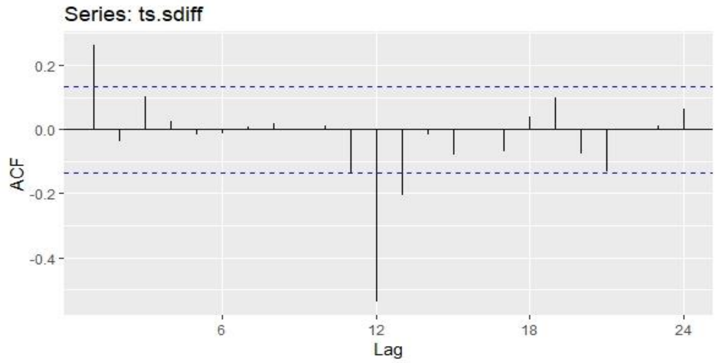
KU LEUVEN

- It is clear from the plot in previous slide that series is not stationary and needs differencing. It is also confirmed by Dickey-Fuller test with p-value > 5% thus failing to reject Null Hypothesis.
- After differencing, the plot becomes somewhat stationary. We check for seasonal effects and notice some seasonality. Taking seasonal differences results in stationary series.
- From the correlograms it is clear that data is not white noise and also confirms the seasonality with a strong autocorrelation at lag 12. Box-Ljung Test with p value < 5% confirms white noise too.

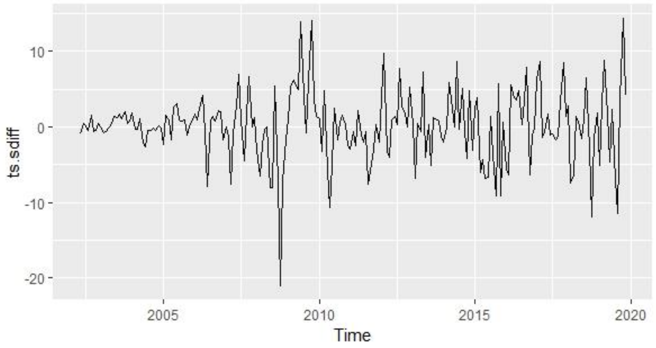
First Differencing:

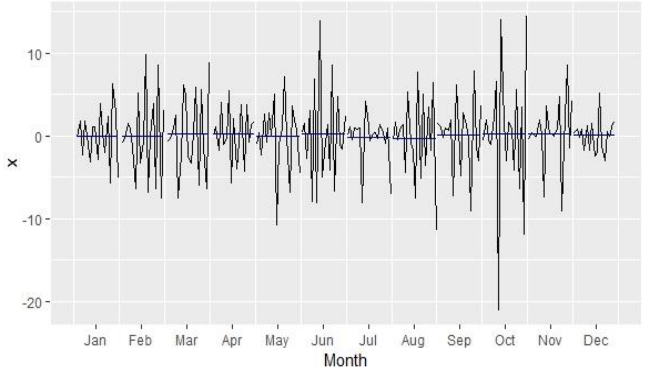


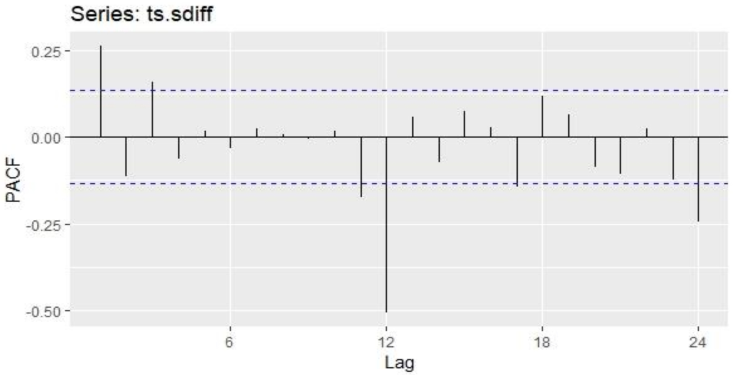




Seasonal Differencing:







Models:

- Multiple models were inspected. Table1 compares them and Table 2 contains their estimates with prediction intervals. Green marked components are significant. Box-Ljung test shows that models 1 and 2 are invalid. (p-values < 5%)

Num	Model	AIC	BIC	ME	RMSE	MAE	MPE	MAPE	MASE	
1	ARIMA(2,1,1)(0,1,0)[12]	1143.41	1156.62	0.0106	3.9331	2.769	0.15	4.78	0.25	Train Set
				7.745	10.52	9.483	5.42	6.7	0.84	Test Set
2	ARIMA(2,1,0)(0,1,0)[12]	1153.77	1163.67	0.0082	4.0742	2.825	0.15	4.91	0.251	Train Set
				8.5981	11.154	10.01	6.03	7.07	0.89	Test Set
3	ARIMA(2,1,1)(0,1,1)[12]	1031.76	1048.27	0.1552	2.7325	1.951	0.32	3.44	0.173	Train Set
				5.4168	6.6884	5.424	3.79	3.8	0.482	Test Set
4	ARIMA(2,1,1)(1,1,1)[12]	1031.61	1051.43	0.1727	2.7008	1.923	0.34	3.39	0.171	Train Set
				5.1522	6.3406	5.169	3.61	3.62	0.46	Test Set
5	ARIMA(1,1,2)(0,1,1)[12]	1031.77	1048.28	0.1546	2.7325	1.95	0.32	3.44	0.173	Train Set
				5.4132	6.6856	5.421	3.79	3.8	0.482	Test Set
6	ARIMA(2,1,1)	1055.67	1069.11	0.457	2.8227	2.116	0.79	3.91	0.188	Train Set
				7.8468	8.9878	7.911	5.51	5.56	0.704	Test Set

Table 1

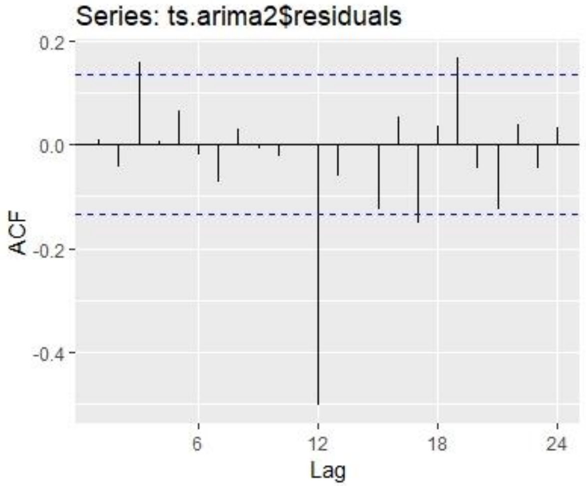
	Coefficients					Prediction Interval				
	ar1	ar2	ma1							
	-0.639	0.1692	0.9925			-0.78	0.03	0.95		
s.e.	0.0709	0.0706	0.0229			-0.50	0.31	1.04		
	ar1	ar2								
	0.2793	-0.0541				0.14	-0.19			
s.e.	0.0706	0.0704				0.42	0.08			
	ar1	ar2	ma1	sma1						
	-0.349	-0.0164	0.6194	-1		-0.79	-0.22	0.20	-1.17	
s.e.	0.2229	0.1015	0.2126	0.0856		0.09	0.18	1.04	-0.83	
	ar1	ar2	ma1	sar1	sma1					
	-0.3154	-0.0408	0.5768	-0.111	-1	-0.75	-0.24	0.16	-0.26	-1.29
s.e.	0.2206	0.0991	0.2121	0.0748	0.151	0.12	0.15	0.99	0.04	-0.71
	ar1	ma1	ma2	sma1						
	-0.3299	0.6012	-0.018	-1		-1.05	-0.12	-0.29	-1.17	
s.e.	0.3677	0.3663	0.1388	0.0856		0.39	1.32	0.25	-0.83	
	ar1	ar2	ma1							
	-0.3584	0.013	0.6448			-0.77	-0.18	0.26		
s.e.	0.2092	0.0993	0.1974			0.05	0.21	1.03		

Table 2

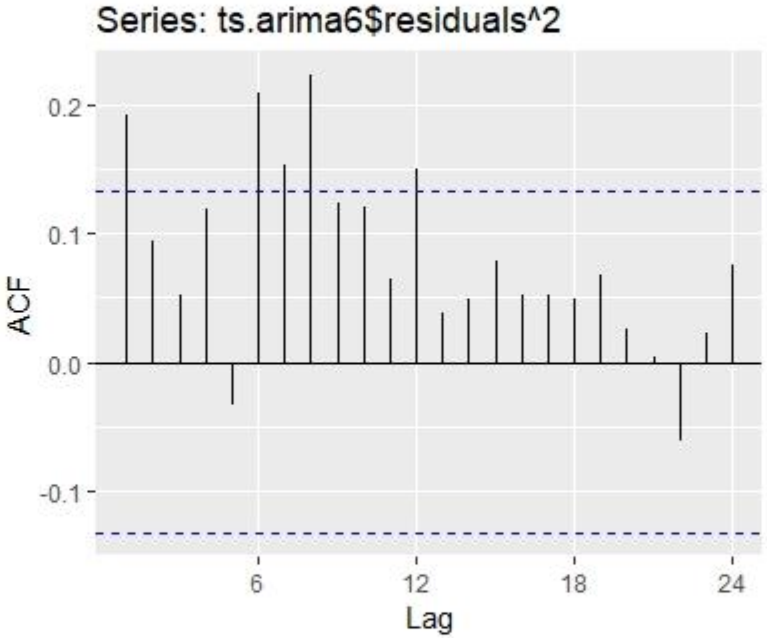
- We observe that model number 3, 4 and 5 give the best AIC, BIC and forecast error values. Table 3 show p values of DM tests for valid models.
  Most of them are not significantly different. Model 4 and 6 are significantly different.
- We select model 3 ARIMA(2,1,1)(0,1,1)[12] for now.
- Model was estimated using train data set consisting of 214 data points. Forecasts are done on test set with 10 data points.

Model	4	5	6
3	0.21	0.95	0.09
4		0.2	0.01
5			0.09

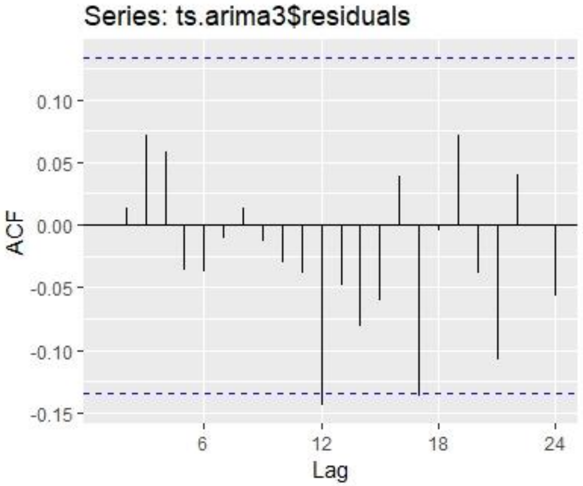
Table 3



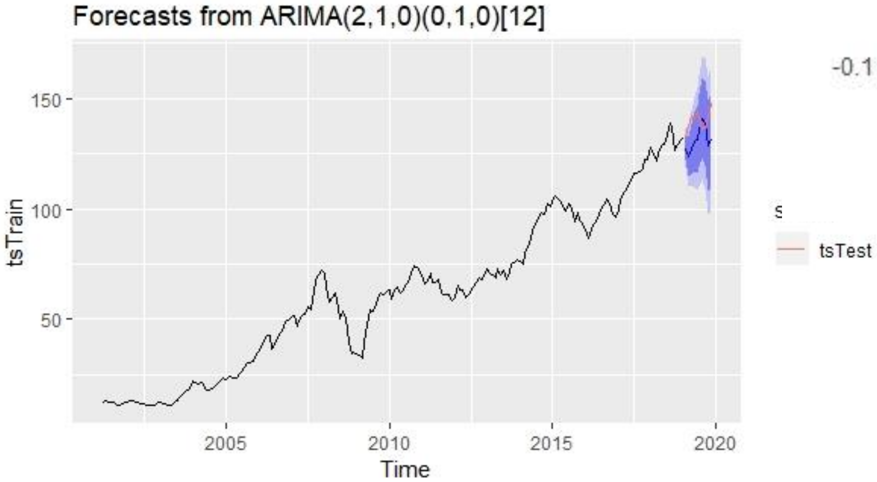
Residuals for the Invalid model, we clearly see autocorrelations and seasonal effect on 12th lag



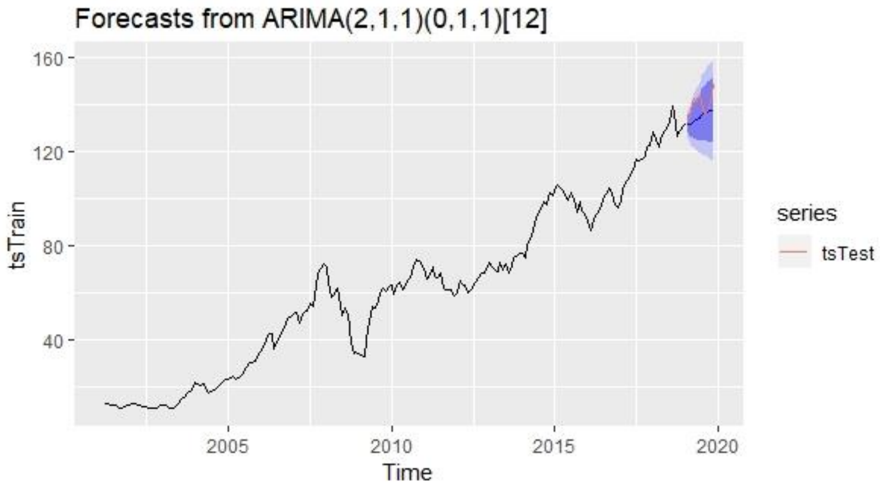
Selected model unfortunately suffers from Heteroscedasticity as seen from the correlogram of squared residuals. GARCH model can be explored.



Residuals for the selected model, we see no significant autocorrelations. 12th lag due to seasonality is a little bit significant.



Invalid model Forecast using test data. We clearly see, inverted forecasts.



Selected model forecasts better and closer to the actuals.

Error Analysis:				
Estimate	Std. Error	t value	Pr(>|t|)	
mu 6.188e+01	8.58E-01	72.135	< 2e-16	***
omega 5.849e+00	2.40E+00	2.435	0.014888	*
alpha1 1.000e+00	2.70E-01	3.7	0.000216	***
beta1 1.000e-08	2.35E-01	0	1	

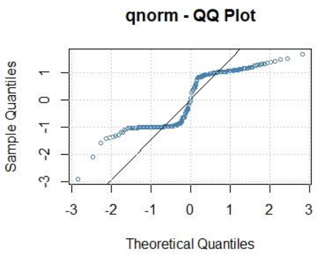
Standardised	Residuals	Tests:				
					Statistic	p-Value
	Jarque-Bera	Test	R	Chi^2	20.11422	4.29E-05
	Shapiro-Wilk	Test	R	W	0.835348	2.60E-14
	Ljung-Box	Test	R	Q(10)	1261.451	0
	Ljung-Box	Test	R	Q(15)	1675.366	0
	Ljung-Box	Test	R	Q(20)	2052.595	0
	Ljung-Box	Test	R^2	Q(10)	19.8123	0.031079
	Ljung-Box	Test	R^2	Q(15)	33.45622	0.004057
	Ljung-Box	Test	R^2	Q(20)	40.61941	0.004168
	LM Arch	Test	R	TR^2	22.32668	0.034018
Information	Criterion	Statistics:				
	AIC	BIC	SIC	HQIC		
8.855151	8.918067	8.85447	8.880575			

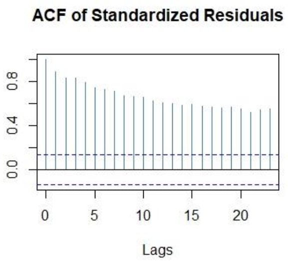
On the left side, garch(1,1) was tested. Except beta1 other coefficients are significant. Jarque-Bera and Shapiro-Wilk test show us that standardized residuals are not normal. Qqplot proves the same. ACF of standard and squared standard residuals show that model is not valid.

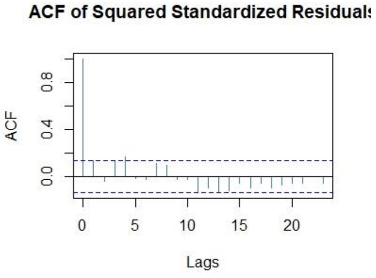
On the right side, we estimate ~arma(2,1)+garch(1,1) model using QMLE conditional distribution. All coefficients except mu are significant. Q-tests on standardized and squared std residuals have pvalue > 5% thus failing to reject Null hypothesis thereby concluding model to be valid. Normal QQ-Plot, ACFs without significant auto correlations are displayed below.

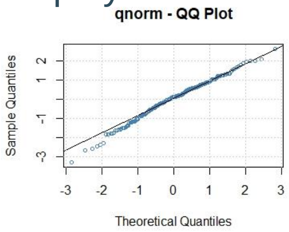
Error A	nalysis:				
	Estimate	Std. Error	t value	Pr(>|t|)	
mu	0.09155	0.25395	0.361	0.71847	
ar1	0.64199	0.119791	5.359	8.36E-08	***
ar2	0.371258	0.122066	3.041	0.00235	**
ma1	0.665184	0.096682	6.88	5.98E-12	***
omega	0.032902	0.000116	283.838	< 2e-16	***
alpha1	0.377361	0.046357	8.14	4.44E-16	***
beta1	0.707985	0.032096	22.058	< 2e-16	***

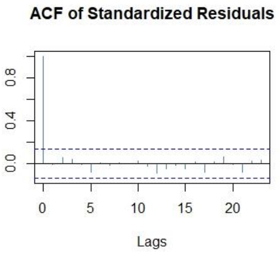
Standardised	Residuals	Tests:				
					Statistic	p-Value
	Jarque-Bera	Test	R	Chi^2	6.109018	0.047146
	Shapiro-Wilk	Test	R	W	0.98862	0.087207
	Ljung-Box	Test	R	Q(10)	2.742385	0.986854
	Ljung-Box	Test	R	Q(15)	5.489005	0.987114
	Ljung-Box	Test	R	Q(20)	8.167846	0.990698
	Ljung-Box	Test	R^2	Q(10)	14.89167	0.136061
	Ljung-Box	Test	R^2	Q(15)	18.27333	0.248569
	Ljung-Box	Test	R^2	Q(20)	21.39475	0.374234
	LM Arch	Test	R	TR^2	16.3183	0.177089
Information	Criterion	Statistics:				
	AIC	BIC	SIC	HQIC		
4.610608	4.72071	4.608557	4.655099			

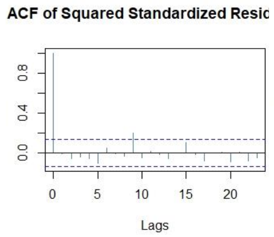


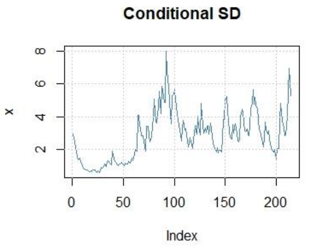




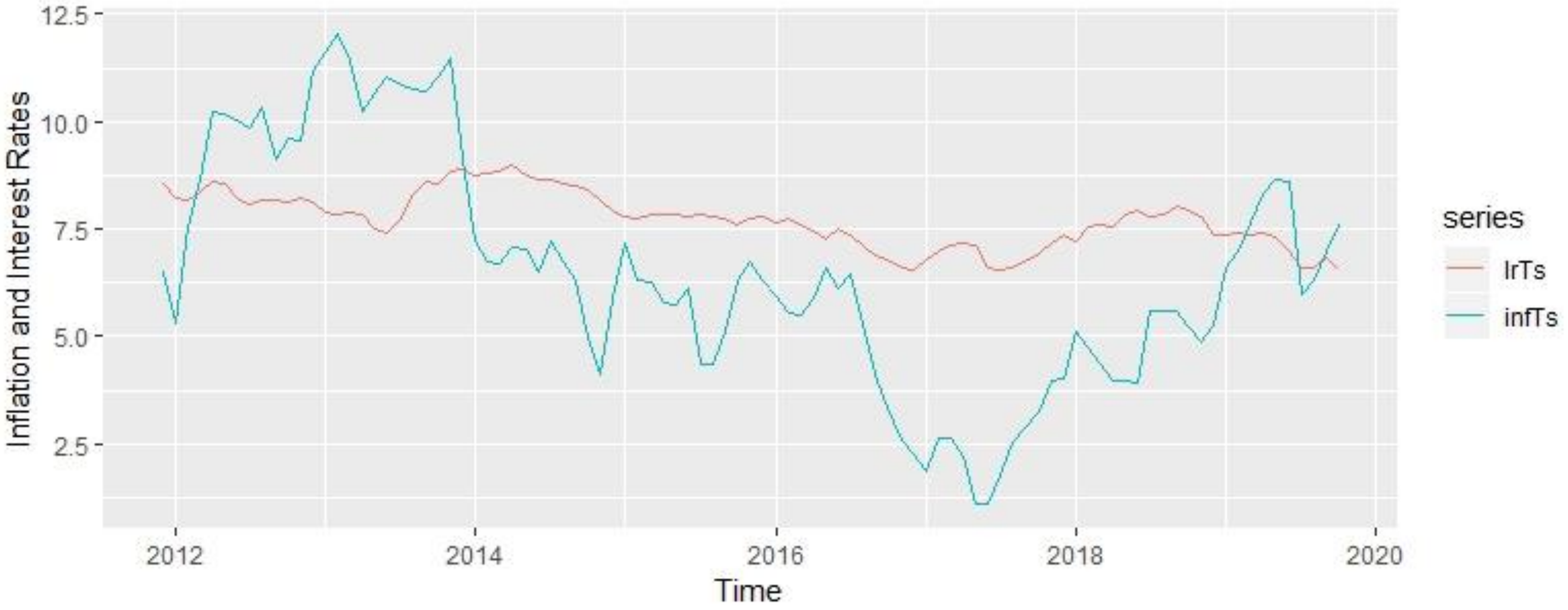






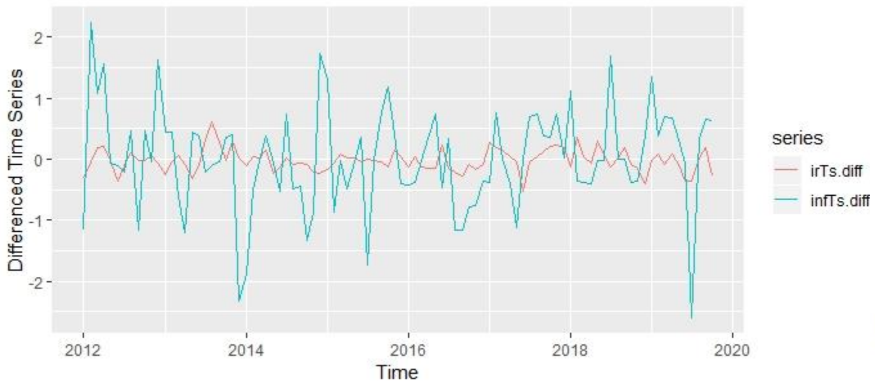


Multivariate Time Series Analysis



- Indian Inflation(CPI) and Long term interest rates has been considered for analysis. Time period is December 2011 to October 2019. In total, 95 data points.
- Data source: https://data.oecd.org/interest/long-term-interest-rates.htm https://data.oecd.org/price/inflation-cpi.htm
- Inflation is expressed here as annual growth %
- Long term interest rates are expressed in percent per annum.

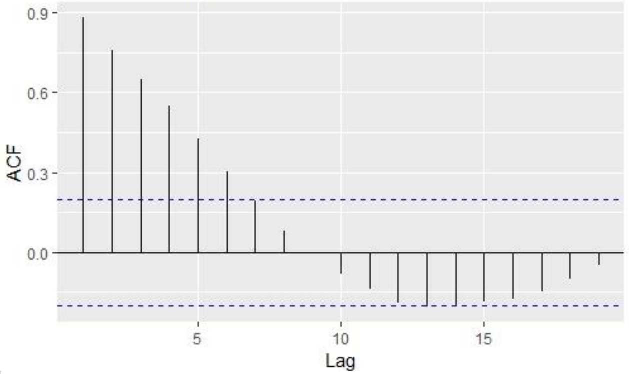
From the multivariate graph in the previous slide it is clear that series are not stationary which is also proven by ADF test. First order differencing stationarizes the series as seen from the below plot.
Linear Regression: IR ~ Inflation shows that though the coefficients are significant there are significant autocorrelations in the residual correlogram. P value is less than 5% in Box Ljung test rejecting Null hypothesis thereby making the model invalid. Linear Regression of differenced IR on differenced Inflation turns out to be valid with no significant autocorrelations on the correlogram and p value > 5%. Coefficients are not significant though.



Non Differenced

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.88792    0.14603   47.17  < 2e-16 ***
infTs        0.12861    0.02074    6.20 1.53e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5463 on 93 degrees of freedom
Multiple R-squared:  0.2925,	Adjusted R-squared:  0.2849
F-statistic: 38.45 on 1 and 93 DF,  p-value: 1.534e-08
```



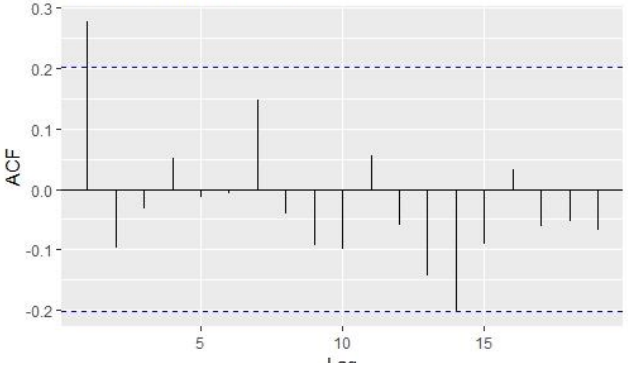

```
	Box-Ljung test

data:  fitLm$residuals
X-squared = 259.19, df = 15, p-value < 2.2e-16
```

Differenced

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.02180    0.01923  -1.133    0.260
infTs.diff   0.02563    0.02280   1.124    0.264

Residual standard error: 0.1865 on 92 degrees of freedom
Multiple R-squared:  0.01355,	Adjusted R-squared:  0.002829
F-statistic: 1.264 on 1 and 92 DF,  p-value: 0.2639
```

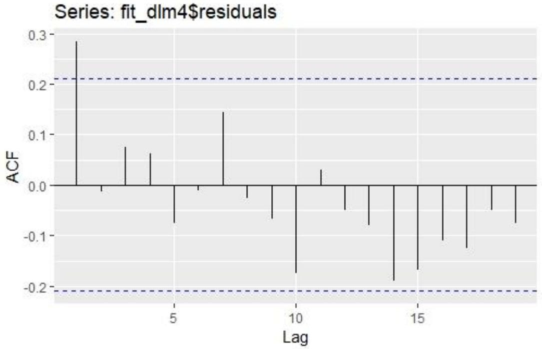



```
	Box-Ljung test

data:  fitLmD$residuals
X-squared = 13.188, df = 10, p-value = 0.2133
```

DLM, ADLM, Granger Causality, Engle Granger

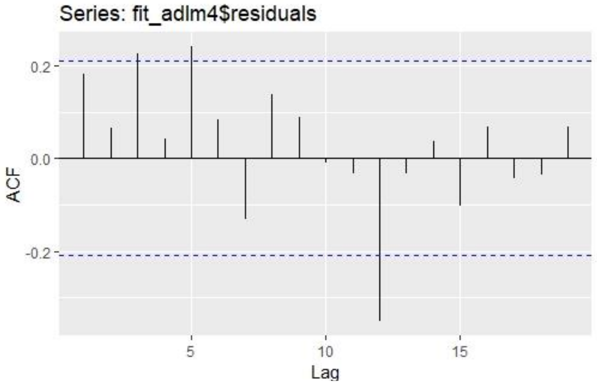
DLM models upto lag 4 have been found to be valid with no significant autocorrelation found in ACF of residuals and pvalue from Box-Ljung test being greater than 5%




```
                Box-Ljung test

data:  fit_dlm4$residuals
X-squared = 14.338, df = 10, p-value = 0.1581
```

ADLM models upto lag 4 have been found to be valid with significant autocorrelation at lag 5 and 12 (indicating some amount of seasonality) found in ACF of residuals and pvalue from Box-Ljung test being less than 5%. R squared is almost 15%. Thus explaining 15% variability of Δinflation by the regressors infTs.diff1 and infTs.diff3 are significant, i.e. Inflation in lag 1 and 3. Which means inflation effects are quarterly.




```
        Box-Ljung test

data:  fit_adlm4$residuals
X-squared = 18.99, df = 10, p-value = 0.0404
```

```
Residuals:
     Min       1Q   Median       3Q      Max 
-2.32152 -0.42501  0.08879  0.43060  1.57318 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -0.05668    0.08444  -0.671   0.5041  
infTs.diff.1 -0.24522    0.10896  -2.251   0.0272 *
infTs.diff.2  0.01618    0.11599   0.139   0.8894  
infTs.diff.3 -0.21637    0.11536  -1.876   0.0644 .
infTs.diff.4 -0.10926    0.10991  -0.994   0.3232  
irTs.diff.1  -0.19914    0.51413  -0.387   0.6996  
irTs.diff.2   0.03936    0.54893   0.072   0.9430  
irTs.diff.3   0.18997    0.54962   0.346   0.7305  
irTs.diff.4  -0.67687    0.50212  -1.348   0.1816  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.779 on 78 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1498,	Adjusted R-squared:  0.0626 
F-statistic: 1.718 on 8 and 78 DF,  p-value: 0.1073
```

Granger Causality of ΔintRate on Δinflation using ADLM 4: Anova gives pvalue > 5% thus concluding ΔintRate has no explanatory power in predicting Δinflation.
Engle Granger test shows that ADF(1) = -1.94 is greater than the critical value -3.41 thus ΔintRate and Δinflation are not co-integrated. Therefore we do not estimate ECM.

```
Analysis of Variance Table

Model 1: infTs.diff.0 ~ infTs.diff.1 + infTs.diff.2 + infTs.diff.3 + infTs.diff.4 + 
    irTs.diff.1 + irTs.diff.2 + irTs.diff.3 + irTs.diff.4
Model 2: infTs.diff.0 ~ infTs.diff.1 + infTs.diff.2 + infTs.diff.3 + infTs.diff.4
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     78 47.338                           
2     82 48.608 -4   -1.2705 0.5234 0.7188
```

```
        ADF test

data:  res_fit_ci
ADF(1) = -1.9443, p-value = 0.3108
alternative hypothesis: true delta is less than 0
sample estimates:
      delta 
-0.08296297
```


KU LEUVEN

```
Estimation results for equation IR:
===================================
IR = Inf..l1 + IR.l1 + const

        Estimate Std. Error t value Pr(>|t|)
Inf..l1  0.03956    0.02169   1.824  0.07152 .
IR.l1    0.27336    0.09925   2.754  0.00712 **
const   -0.01332    0.01828  -0.729  0.46813
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.1754 on 90 degrees of freedom
Multiple R-Squared: 0.1213,     Adjusted R-squared: 0.1017
F-statistic: 6.209 on 2 and 90 DF,  p-value: 0.002978
```

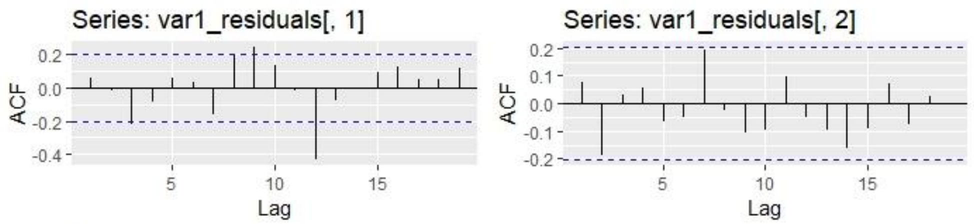
VAR(1) was tested to begin with. The estimation results presented here show that for equation Interest Rate, 12% variance of ΔintRate is explained by lagged obs of ΔintRate and Δinflation at lag 1.
Fstat < 5% show that regressors are jointly significant

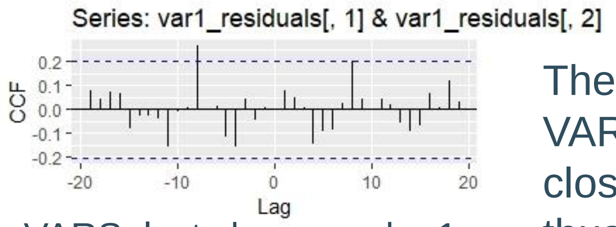
```
Estimation results for equation Inf.:
=====================================
Inf. = Inf..l1 + IR.l1 + const

        Estimate Std. Error t value Pr(>|t|)
Inf..l1  0.20929    0.10307   2.031   0.0452 *
IR.l1   -0.31047    0.47158  -0.658   0.5120
const    0.01778    0.08687   0.205   0.8382
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.8333 on 90 degrees of freedom
Multiple R-Squared: 0.04542,    Adjusted R-squared: 0.02421
F-statistic: 2.141 on 2 and 90 DF,  p-value: 0.1234
```

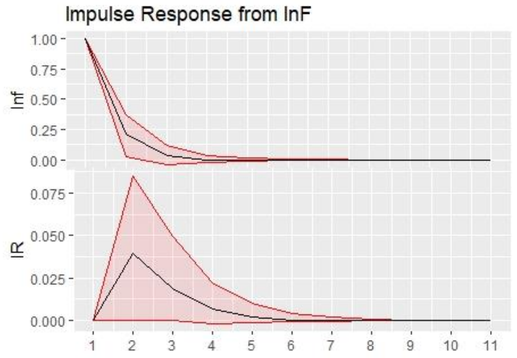
The estimation results presented here show that for equation Inflation, only 4% variance of Δinflation is explained by lagged obs Δinflation at lag 1.
Fstat >5% show that regressors aren't jointly significant



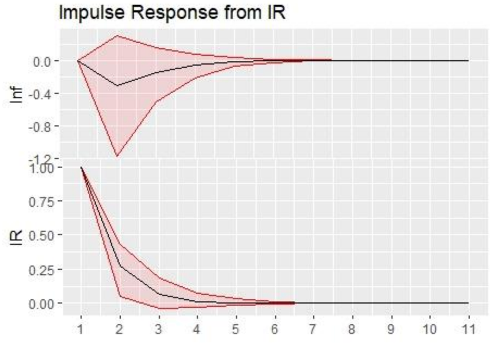


The residual plot for VAR(1) seem to be close to white noise thus we can say that the model is valid.

VARSelect choses order 1 which is same as before thus VAR(1) validity is confirmed.



Unitary impulse in Inf at t results in positive response in IR at time t+2.



Unitary impulse in IR at t results in negative response in IR at time t+3.

Conclusion:

In univariate analysis we observed that stock indices have a lot of volatility and seasonality. Though SARIMA is used to model the data, GARCH proves to be better in modelling volatility.
To improve the predictions maybe methods like Holt’s Winter Exponential smoothing or GARCH with SARIMA could be used which are out of the purview of this course.

For multivariate analysis, inflation and long term interest rate was compared. We know in general that these two share an inverse relation with each other. We can also see this in the plots and impulse response function.
DLM and ADLM show us that inflation is strongly dependent on its previous values which makes sense because inflation does not increase or decrease quickly. There is strong persistency and dependency on previous values. Though there is some amount of inverse correlation the time series are not cointegrated. Also Granger Causality test shows us that one series is not explaining the other.