Time series and Regression analysis of Yahoo share price of Japan and USA

Name: Rochita Das

Registration No: A01−2112−0735−11

Roll No: 91/STS/141003

Contents:

Topics	Page No.
1. Motivation	3
2. Objectives	3
3. Data collection	4
4. Time series analysis	5 – 34
i. General approach	6 - 12
a. Detrend	7
b. Deseasonalisation	7
c. Stationary data	8
d. Mean stationary	8
e. Variance stationary	11
f. Confidence interval	12
g. Forecast	12
ii. Time series analysis of Yahoo Share price of USA	13 - 21
iii. Time series analysis of Yahoo Share price of Japan	22 - 34
5. Regression analysis	35 – 43
i. Exogenous variables	36 - 37
ii. VAR Model	38 - 39
iii. Simultaneous Equation Model	40 - 43
6. Reference	44
7. Acknowledgements	45
8. Appendix - I (snapshot of data)	46 - 47
9. Appendix - II (r – codes)	48 - 57

Motivation:

The **stock market** is the market, where shares of publicly held companies are issued and traded either through exchanges or over-the-counter markets. Also known as the equity market, the stock market is one of the most vital components of a free-market economy, as it provides companies to access the capital in exchange of giving investors a slice of ownership in the company. The stock market makes it possible to grow small initial sums of money into larger ones, and to become wealthy without taking the risk of starting a business or making the sacrifices that often are required in a high-paying career. But not all investors are successful in gaining a return on their investment, rather very few are successful in making money. This happens because the price of stocks is constantly fluctuating with time and at any moment, the price of a stock could fall below its buying price; selling the shares at this reduced price results the investors losing money. A natural solution would be for investors to sell their shares before they begin to diminish in value, ideally at a point when the stock's price is higher than when it was purchased by the investor. That is why, nowadays prediction of stock price is needed very much in this competitive economy.

Objectives:

- ✓ Firstly, **time series analysis** is done separately on both yahoo share price data of USA and Japan.
- ✓ Secondly, **regression analysis** is done considering some relevant exogenous variables and also considering mutual interaction between the two share prices.

Key words: ARIMA, GARCH, VAR, Causality, 2SLS

Data collection:

Data are collected on the following variables:

1. Yahoo share price of USA - Y_1 (endogenous variable)
2. Yahoo share price of Japan - Y_2 (endogenous variable)
3. Nasdaq (American stock exchange index) - X_1 (exogenous variable)
4. Nikkei (Tokyo stock exchange index) - X_2 (exogenous variable)
5. Yen – Dollar exchange ratio - X_3(exogenous variable)

For variables Y_1, Y_2, X_1, X_2 market **closing values** are considered.

- Y_1 , Y_2 are considered as **endogenous** variables as they affect each other.
- X_1, X_2, X_3 are considered as **exogenous** variable. For this reason, their values are considered at one lag before than endogenous variables.

✓ Data on Y_1, Y_2 are collected from 20/01/2010 to 26/01/2016.

✓ Data on X_1, X_2, X_3 are collected from 19/01/2010 to 25/01/2016.

✓ For prediction purpose, data on each variable are collected for 10 days after the last time point.

➢ Data are collected from the following sites:

finance.yahoo.com

https://research.stlouisfed.org/fred2/series/EXJPUS

A snapshot of data set is given in the appendix - 1.

TIME SERIES ANALYSIS

General approach:

A **time series** is a collection of observations made sequentially and typically equally spaced in time. (In this project the share price observations are considered as successive observations, though some of them are apart by holidays). Time series analysis is a major branch in statistics that mainly focuses on the characteristics of the data and extract meaningful statistical model in order to predict future values of the series. There are two methods in time series analysis, namely: **frequency-domain** and **time-domain**. The former is based mostly on Fourier Transform while the latter closely investigates the autocorrelation of the series and is of great use of Box-Jenkins and ARCH or GARCH methods to perform forecast of the series.

In **time domain**, it is often convenient to think of a time series as consisting of non-stationary components (which evolve over time), such as trend (T_t), seasonal variation (S_t) and stationary irregular component (I_t), (which does not evolve with time) for t^{th} time point.

By **trend** we mean a smooth, regular, long-term movement which exhibits the basic tendency of a series to either increase or decrease or remain steady with time. Any minor or irregular fluctuations are inconsistent with idea of trend. The **seasonal variation** is concerned with the periodic fluctuation in the series with in each year. Seasonal fluctuations are most often attributed to social customs or weather changes. Once the trend and seasonal variations have been accounted for, the remaining movement is attributed to **irregular components**.

Let Y_t be the t^{th} time series observation, then

$$\boldsymbol{Y_t = T_t + S_t + I_t} \qquad : \quad \text{Additive Model}$$

$$\boldsymbol{Y_t = T_t * S_T * I_t} \qquad : \quad \text{Multiplicative Model}$$

Generally, a ***multiplicative model*** with a ***log transformation*** is used which reduces the model to an additive one.

$$\boxed{\boldsymbol{logY_t = logT_t + logS_t + logI_t}}$$

The first step in modeling time series data is to convert the non-stationary time series to stationary one. This is important for the fact that a lot of statistical and econometric methods are based on this assumption and can only be applied to stationary time series. Non-stationary time series are erratic and unpredictable while stationary process is mean-reverting, i.e., it fluctuates around a constant mean with constant or changing variance.

Detrend:

- ✓ Firstly, the log transformed data is **plotted** and it is tried to find whether seasonality is present or not.
- ✓ Then **Mann-Kendall Test** for Monotonic Trend on log transformed data is performed for confirmation whether trend is present or not. The hypothesis of the test is,

$$\boldsymbol{H_0}\text{: No monotonic trend} \quad \text{vs} \quad \boldsymbol{H_1}\text{: Monotonic trend is present}$$

If trend is present, next step is to detrend the data. The method may be -

a) Differencing up to appropriate order
b) Fitting mathematical curve

After the trend is removed, it is checked whether seasonality is present in detrended data.

Deseasonalisation:

- ✓ Firstly, first few data are plotted week wise and looked for whether weekly seasonality is present or not.
- ✓ Month wise **boxplot** is drawn and it is tried to find whether monthly seasonality is present or not.
- ✓ For confirmation, **Kruskal-Wallis test by ranks** (non-parametric One-way ANOVA on ranks) is performed. The hypothesis of the test is,

$\boldsymbol{H_0}$: medians of all groups are identical (seasonality is absent)

vs $\boldsymbol{H_1}$: at least one population median of one group is different from the population median of at least one other group.

If seasonality is present, it is removed by **method of average**. In method of average, averages for each of the months are calculated and the averages are adjusted to zero and 12 for additive and multiplicative model respectively. These adjusted averages are called "**monthly indices**". From detrended data, the corresponding monthly indices are subtracted to get deseasonalised data.

Stationary data:

Now the detrended, deseasonalised data should look like **irregular**. This has become startionary data, say it as $\boldsymbol{Z_t}$ for $\mathbf{t^{th}}$ time point. Now one opt to model for mean stationary and then variance stationary i.e. volatility (if required) for $\boldsymbol{Z_t}$ data.

➢ Weak stationary:

The time series $\{Z_t, t \in I\}$ (where I is the time point set) is said to be weak stationary if,

i. $E(Z_t) = \mu \qquad \forall\ t \in I$
ii. $E(Z_t^2) < \infty \qquad \forall\ t \in I$
iii. $cov(Z_t, Z_{t+k}) = \Upsilon\ (k) \qquad \forall\ t, k \in I$

➢ Strict stationary:

The time series $\{Z_t, t \in I\}$ is said to be strict stationary if the joint distribution of $(Z_{t1}, Z_{t2}, \dots, Z_{tn})$ is the same as that of $(Z_{t1+k}, Z_{t2+k}, \dots, Z_{tn+k})$ for any integer k .

Mean Stationary:

The different Box-Jenkins models are identified by the number of autoregressive parameters (*p*), the degree of differencing (*d*), and the number of moving average parameters (*q*). Any such model can be written using the notation ARIMA (*p,d,q)*.

➢ ACF:

ACF, autocorrelation coefficient measures the linear relationship between the time series observations separated by a lag of k time units. The sample autocorrelation coefficient (ACF) of lag *k* is computed for the (*n*-*k*) pairs is given by

$$\widehat{\rho_k} = \frac{\sum_{t=k+1}^{T}(Z_t-\bar{Z})(Z_{t-k}-\bar{Z})}{\sum_{t=1}^{T}(Z_t-\bar{Z})^2}$$

➢ PACF:

PACF, partial autocorrelation coefficient of lag k, denoted by, φ_{kk} , is a measure of the correlation between Z_t and Z_{t-k} after adjusting for the presence of $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$. This adjustment is done to see if there is any correlation between Z_t and Z_{t-k} beyond that induced by the correlation Z_t has with $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$.

$$\alpha(1) = \text{cor}(Z_t, Z_{t-1,})$$
$$\alpha(\text{k}) = \text{cor}[\, Z_{k+1} - P_{t,k}(Z_{k+1}),\ Z_{1} - P_{t,k}(Z_1)] \quad \text{for k} \geq 2$$

Where $P_{t,k}(x)$ denotes the projection of 'x' onto the space spanned by $x_2, x_3, \dots, x_k$.

➢ Choice of Model:

If the ACF trails off and the PACF shows spikes, then an autoregressive (AR) model with order 'p' equal to the number of significant PACF spikes is considered the best model. If the PACF trails off and the ACF shows spikes, the moving average (MA) model with order 'q' equal to the number of significant ACF spikes is the best model. If both the ACF and the PACF trail off then an autoregressive moving average (ARMA) model is used with proper choice of p and q. If the data had to be differenced for it to become stationary, then the ARIMA model is used.

➢ **Different Models:**

- AR(p) model: $Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-1} + \ldots + \alpha_p Z_{t-p} + \varepsilon_t$
- MA(q) model: $Z_t = \beta_0 \varepsilon_t + \beta_1 \varepsilon_t + \ldots + \beta_q \varepsilon_{t-q}$
- ARMA (p,q) model :

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \ldots + \alpha_p Z_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_q \varepsilon_{t-q}$$

$$\text{i.e., } \varphi(B)\, Z_t = \psi(B)\, \varepsilon_t$$

- ARIMA (p,d,q) model : $\varphi(B)(1-B)^d\, Y_t = \psi(B)\varepsilon_t$
- SARIMA(p,d,q)x(P,D,Q)s model :

$$\varphi_p(B)\Phi_P(B^s)(1-B^s)^D(1-B)^d\, Y_t = \psi_q(B)\Psi_Q(B^s)\varepsilon_t$$

[where $\varphi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p$, $\psi(z) = 1 + \beta_1 z + \beta_2 z^2 + \ldots + \beta_q z^q$]

where ε_t are independent random quantities

All these model parameters (situation wise) are chosen on the basis of **minimum AIC** or **minimum FPE** (final error prediction) criterion so that it becomes a parsimonious model.

➢ **Adequacy of fitted model:**

After an appropriate model is fitted on the stationary data, we can compare the predicted values with original observations. To check the adequacy of fitted model, the residuals (r_t) are calculated by $r_t = Zt - \widehat{Zt}$ and the **Ljung-Box test** is performed on the residual. The hypothesis of the test is

H_0: The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0)

vs ***H_1***: The data are not independently distributed; they exhibit serial correlation.

If the residuals are independently dist. (H_0 accepted), it signifies a proper fit. Hence a good model.

Variance stationary (Volatility):

Consider ARMA-residual as X_t for t^{th} time point.

- ✓ Firstly, **plot** of X_t series is drawn. If the plot shows large fluctuations followed by small fluctuations and then again followed by large and this continues, then there is an indication that variation evolves over time. That means the past variation affects present variation. To capture this heteroscedastic variation one should opt for ARCH or GARCH model.
- ✓ To confirm whether significant volatility is present or not, **Lagrange multiplier test** is performed on X_t series. The hypothesis of the test is,

$\boldsymbol{H_0}$: No heteroscedasticity is present vs $\boldsymbol{H_1}$: Heteroscedasticity is present

If the test confirms presence of heteroscedasticity, we opt for appropriate model.

➢ Different Models:

Let $X_t = \varepsilon_t * \sqrt{h_t}$, where $\boldsymbol{\varepsilon_t}$ is a strong white noise process.

Now h_t is modelled differently as follows,

- ARCH (p) : $h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 + \cdots + \alpha_p x_{t-p}^2$
- GARCH (p,q): $h_t = \alpha_0 + \alpha_1 x_{t-1}^2 + \alpha_2 x_{t-2}^2 + \cdots + \alpha_p x_{t-p}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots + \beta_q h_{t-q}$

All these model parameters (situation wise) are chosen on the basis of **minimum AIC** criterion so that it becomes a parsimonious model.

➢ Adequacy of fitted model:

After an appropriate model is fitted on the stationary data, comparison of predicted values with original observations is done. To check the adequacy of fitted model, the residuals (r_t) are calculated by $\boldsymbol{\widehat{\varepsilon_t} = \frac{X_t}{\sqrt{\widehat{h_t}}}}$ and the **Ljung-Box test** is performed on the residual. If the residuals are independently distributed (i.e. H_0 accepted), it signifies a proper fit, hence a good model.

Confidence Interval (CI):

Our first goal is to find an approximate distribution of ARMA-residual i.e. r_t series.

- ✓ Firstly, **QQ-plot** of r_t series is drawn. It may differ significantly from Normal distribution.
- ✓ For confirmation **Anscombe test** on kurtosis is performed. The hypothesis is,

$$\boldsymbol{H_0} : \text{Kurtosis} = 3 \quad \text{vs} \quad \boldsymbol{H_1} : \text{Kurtosis} > 3$$

If H_0 is rejected, it is assumed that the distribution of r_t series to be approximately " t distribution", otherwise "Normal distribution". The parameters of the distribution are estimated from r_t series.

Now we construct ***CI*** as,

$$\text{predicted mean (ARMA)} \pm t_{\frac{\alpha}{2}} * \text{predicted standard deviation (ARCH/GARCH)}$$

$$\left(\widehat{\boldsymbol{Zt}} - \boldsymbol{t}_{\frac{\alpha}{2}} * \sqrt{\widehat{\boldsymbol{h_t}}} \ , \ \widehat{\boldsymbol{Zt}} + \boldsymbol{t}_{\frac{\alpha}{2}} * \sqrt{\widehat{\boldsymbol{h_t}}} \right)$$

Where $t_{\frac{\alpha}{2}}$ is upper $(100*\frac{\alpha}{2})\%$ point of fitted t-dist or normal-dist.

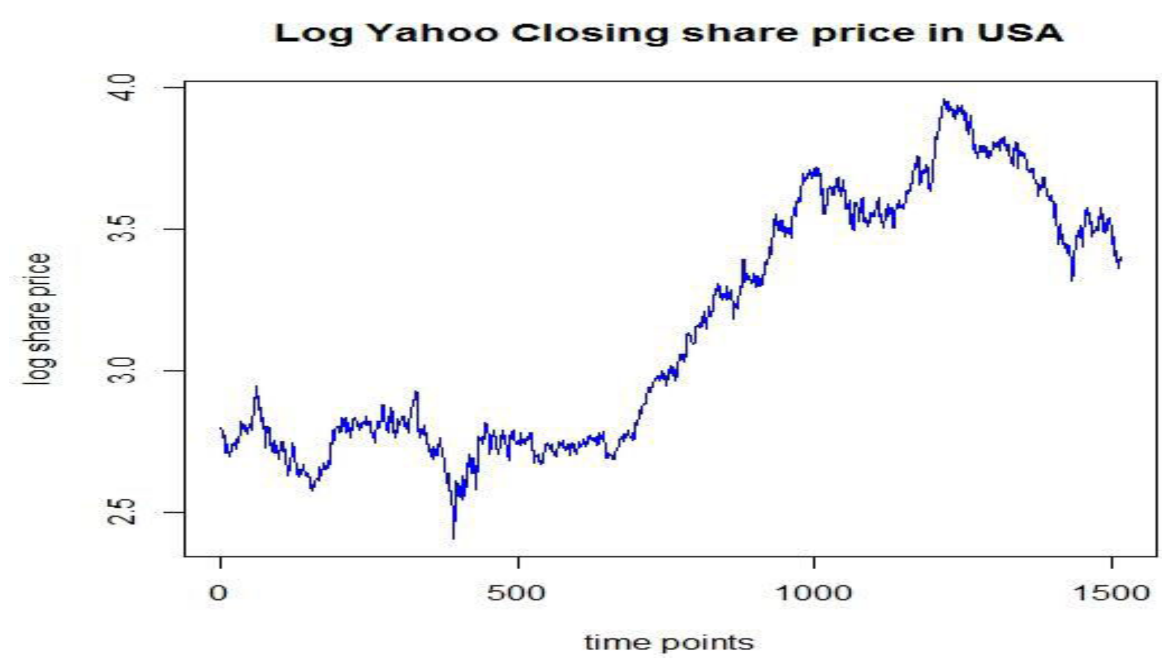
Forecast:

Mean is forecasted through fitted ARMA model and forecasted trend and corresponding seasonal index is added then. It gives required forecasted value on that time point. Then standard deviation is forecasted through fitted ARCH or GARCH model. Following the above procedure similarly one can forecast CI. Then both forecasted mean value and CI is gone through exponential transformation to compare with original values.

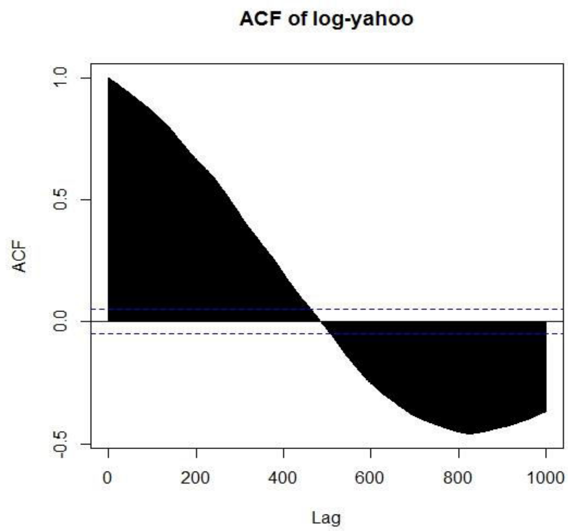
Following this general approach now we proceed for separate time series analysis of Yahoo share price data of USA and Japan.

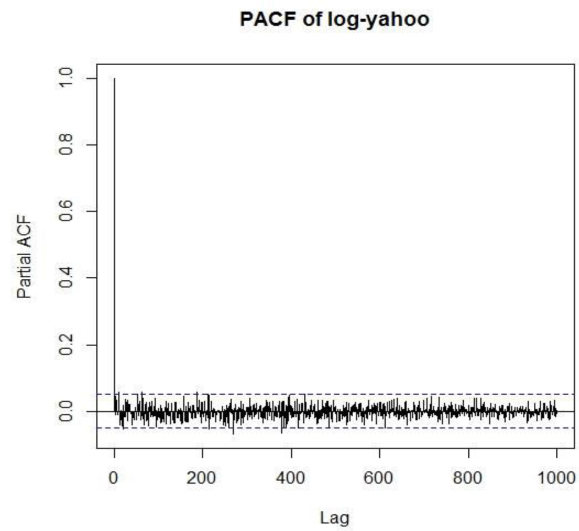
Time Series Analysis of Yahoo Share price of USA

The plots of original data with log transformation are ,



ACF, PACF plots of log transformed data are ,





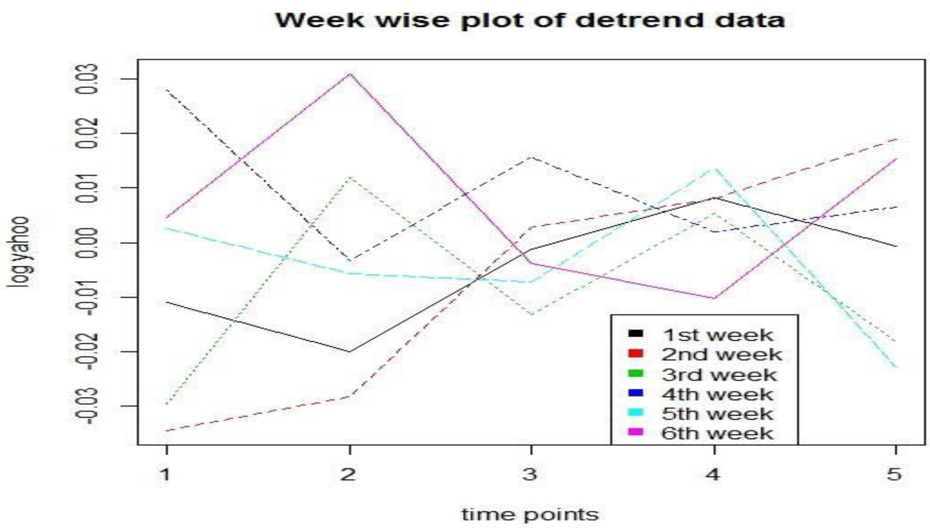
- ✓ The log transformed plot shows a very prominent increasing trend.
- ✓ The **Mann kendal test** result is tau = 0.609, 2-sided ***p-value =< 2.22e-16*** (i.e., H_0 $is\ rejected$). Hence the test also supports presence of significant trend .

Now following the very common way one can procced for detrending by considering $\mathbf{1^{st}}$ **order difference**. The plot of 1^{st} order difference is as follow,

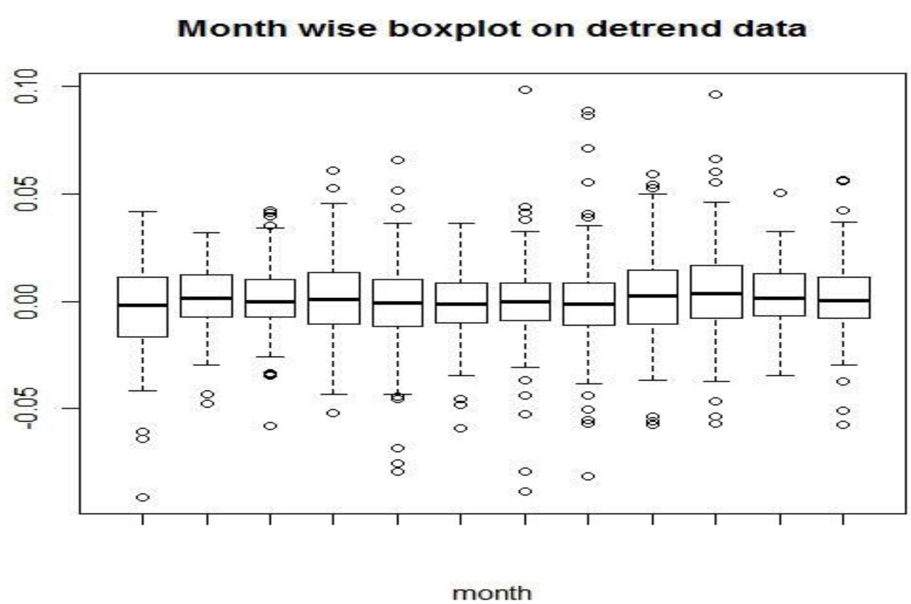
- ✓ The plot shows a random fluctuation about The line Y=0, hence the trend is considered to be removed properly.

✓ Moreover, Mann kendal test on 1^{st} order difference gives ***p-value =0.83552*** confirming absence of significant trend .

Now the week wise plot of first few detrended data is plotted, which shows absence of seasonal (weekly) fluctuation.

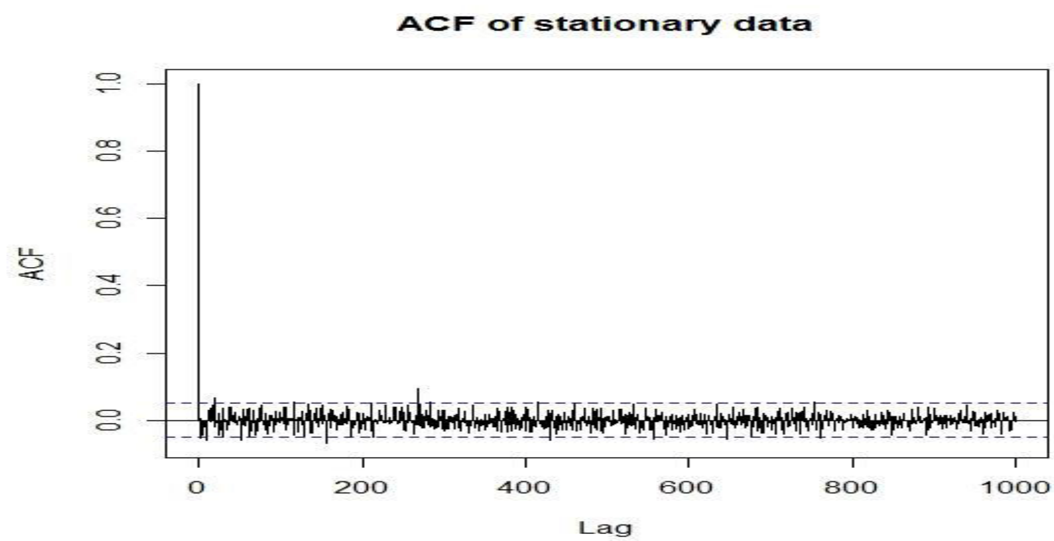


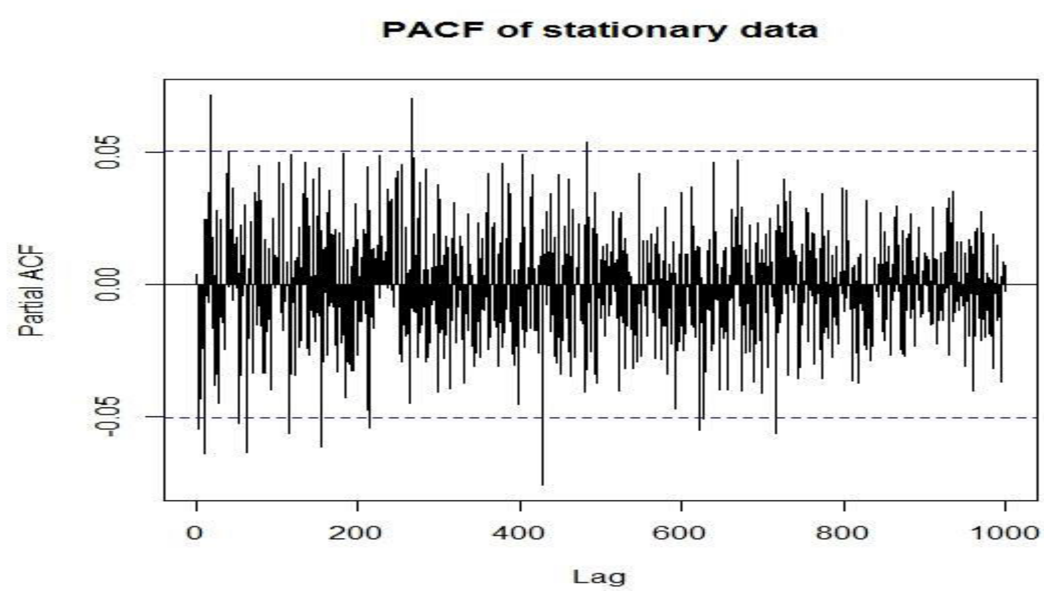
The month wise boxplot also says no significant seasonal (monthly) pattern is present . The plot is as follow,



✓ **Kruskal-Wallis test** : chi-squared = 14.6261, df = 11, ***p-value = 0.2003*** (i.e., H_0 *is accepted*).

Hence the detrended data is considered as stationary data (as it does not contain seasonal fluctuation). The ACF,PACF plots of stationary data are as follows,





✓ ACF tails off and PACF shows few significant spikes, hence AR(p) model with a large p value is appropriate.

- ✓ To keep the model parsimonius we opt for ARMA (p,q) model.Equivalently, we opt for ARIMA(p,d=1,q) model on the original log-transformed data.

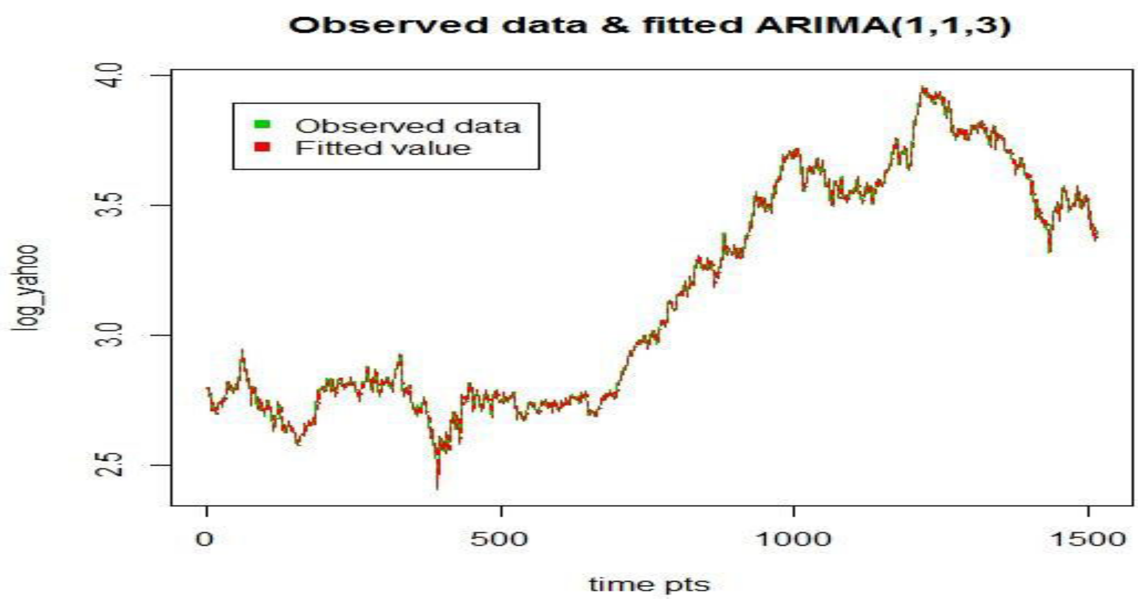
The model parameters chosen by minimum AIC criterion. The model is ***ARIMA(1,1,3)*** [on original log transformed data] as follow,

Yahoo share price of USA Y_{1t} at t^{th} time point,

$$\boldsymbol{Y_{1t} - Y_{1\overline{t-1}} = 0.6422 * \left(Y_{1\overline{t-1}} - Y_{1\overline{t-2}}\right) + \varepsilon_t - 0.6426 * \varepsilon_{t-1} - 0.0041 * \varepsilon_{t-2} - 0.0562 * \varepsilon_{t-3}}$$

With aic = -7633.88

The following plot shows fitted value with original value .

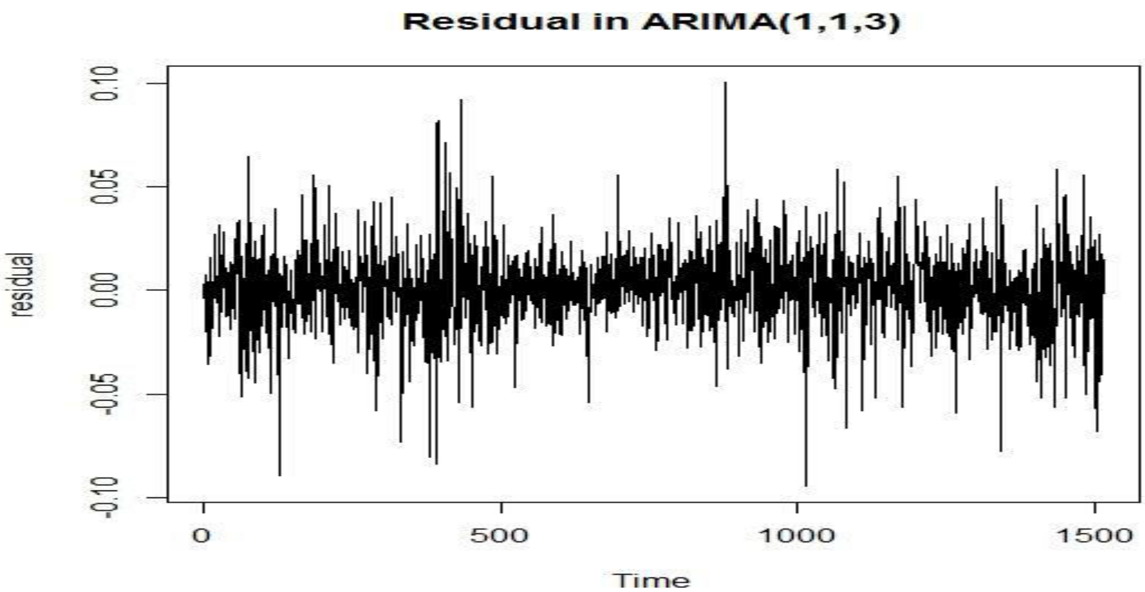


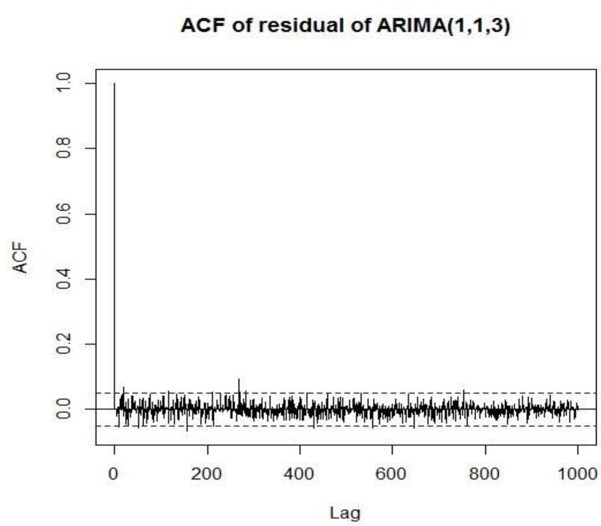
- ✓ Now the ARIMA(1,1,3) residuals are calculated and Parformance of **Ljung-Box test** on residuals is as follow,

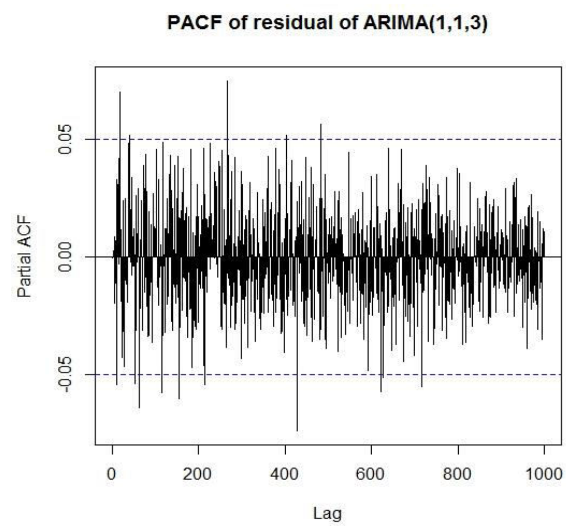
Lags	*Statistic*	*df*	*p-value*
5	0.3959217	5	0.9990010
10	5.4452888	10	0.8611389
15	11.1606360	15	0.7422577

The test accepts H_0, which tells ARIMA residuals are purely random. Hence a good fit.

The Residual plot with ACF and PACF plots are as follows ,







- ✓ The residual plot shows large fluctuations followed by small fluctuations and then again followed by large and this continues, then there is an indication that variation evolves over time.
- ✓ For confirmation, whether ARCH or GARCH model is suitable, the **Lagrange multiplier test** for ARCH(1) is as follow, Chi-squared = 15.6018, ***p-value = 7.818e-05*** which confirms atleast ARCH(1) is required to model the volatility present in the ARIMA residual .

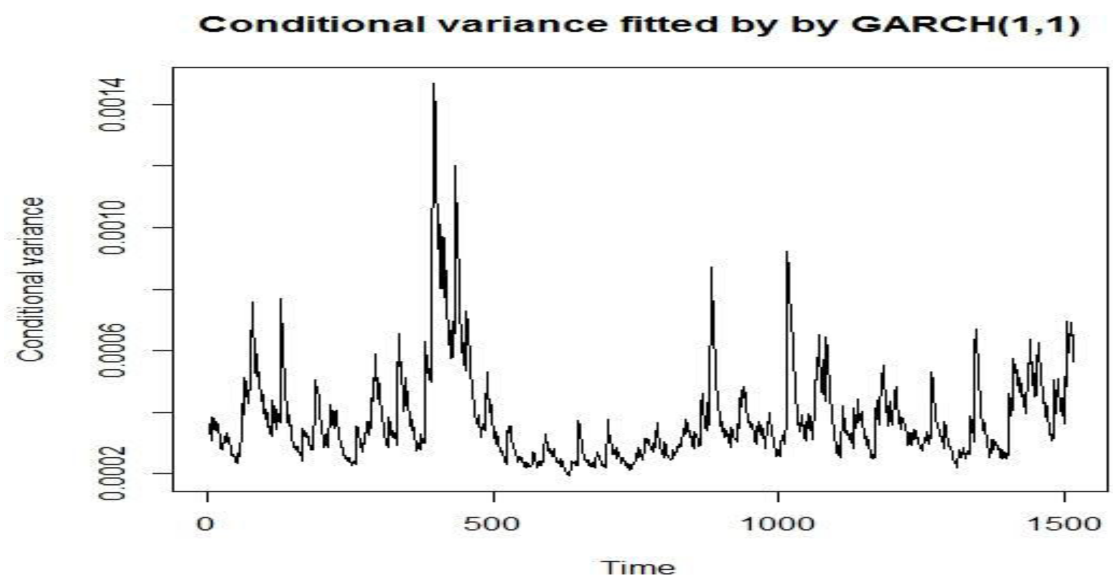
The optimum model parameter is chosen by minimum AIC criterion. The model is ***GARCH(1,1)***

ARIMA residual, $\boldsymbol{r_t = 7.213 * 10^{-4} + \varepsilon_t * \sqrt{\widehat{h_t}}}$

Where $\boldsymbol{\widehat{h_t} = 1.426 * 10^{-5} + 0.05264 * r_{t-1}^2 + 0.91 * h_{t-1}}$

- ✓ **Ljung-Box Test** on fitted GARCH(1,1) residual we find the model to be adqequate to capture volatility. The test result is , X-squared = 0.011, df = 1, ***p-value = 0.9164,*** hence a good fit.

The plot of conditional variance, i.e., $\widehat{\mathrm{h}_t}$ is as follow,

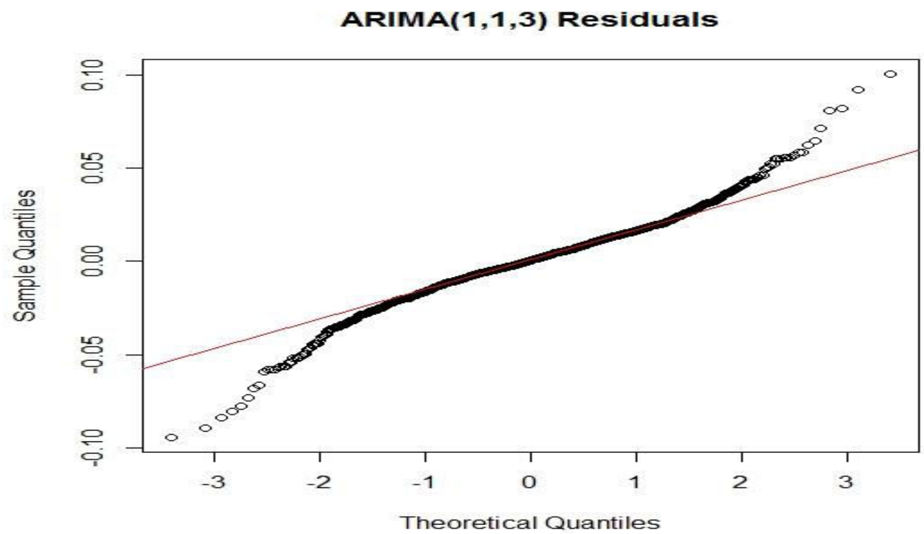


Now to find the C.I. of time series data, approximate $dist^n$ of ARIMA residual is found first .

Kurtosis $(\boldsymbol{r_t})$ **=5.867355** , which is higher than that of Normal distribution.

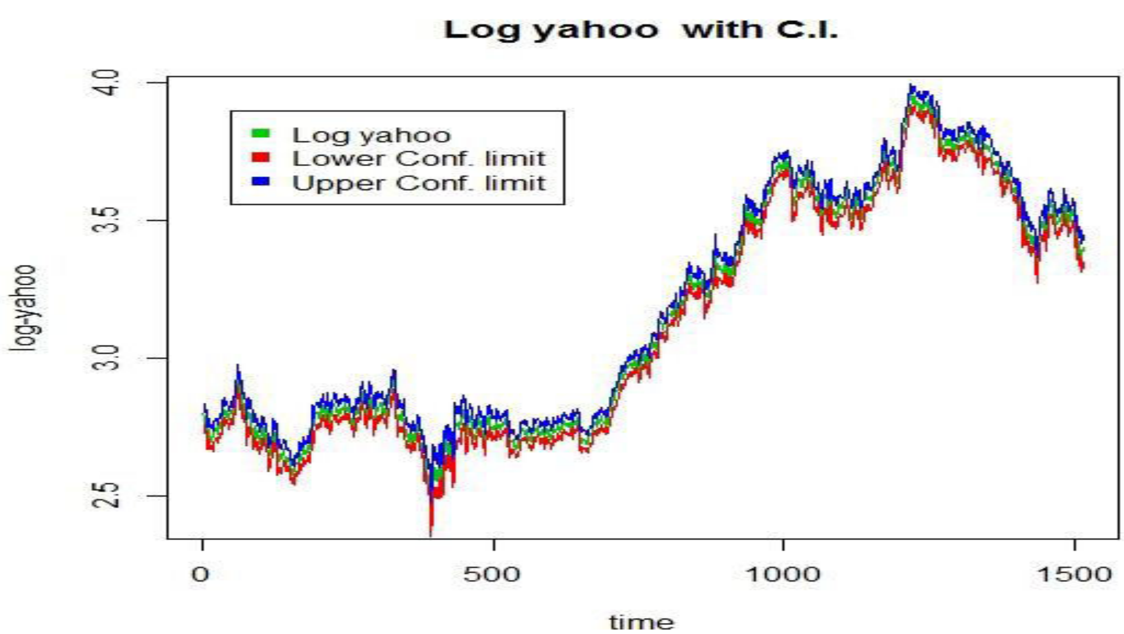
- ✓ **Anscome test** gives the result : kurt = 5.8674, z = 10.3475, ***p-value < 2.2e-16*** (H_0 is rejected). Hence kurtosis significantly differ from 3 .

✓ The **QQ-plot** also supports the fact obtained above



Having large frequency on tail observations, we fit '**t(ν)**' distribution. The value of ν is estimated from ARIMA residual as, **ν = 5** . Then upper upper 100*(α/2)% point of t(ν=5) distribution is, $t_{\alpha/2} = 2.015048$, for α=.05 .

Now the original log transformed data with confidence interval is plotted as follow,



Forecast:

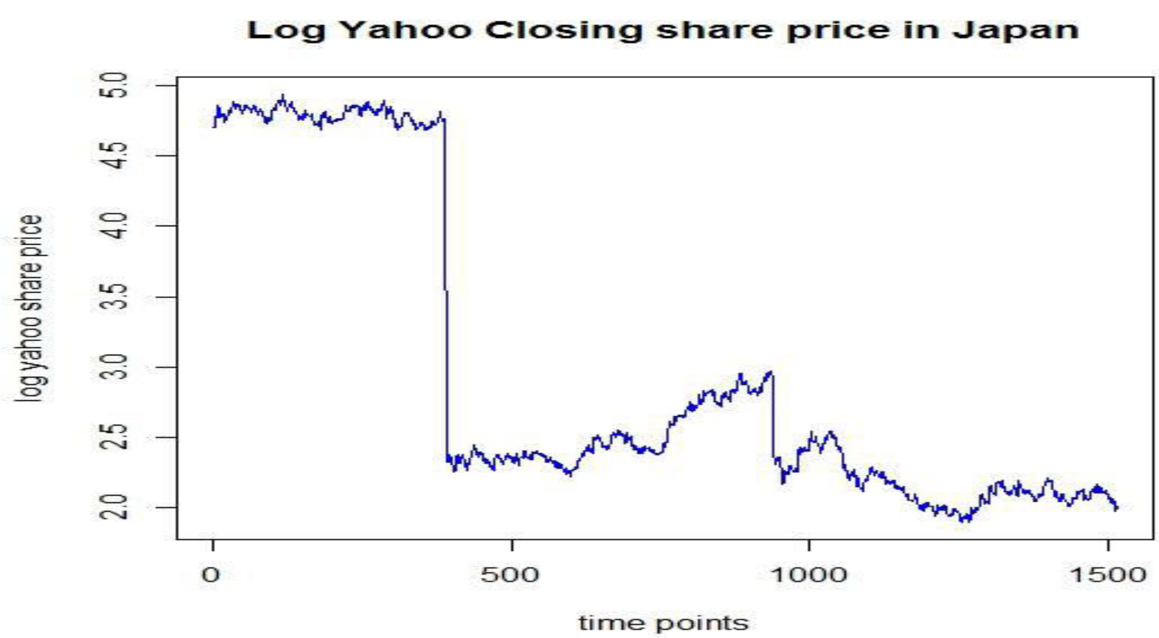
Proceeding through the steps described in Forecast method, the following table is obtained ,

Time point	*Observed value*	*Predicted value*	*95 % lower conf limit*	*95 % upper conf limit*
27-01-16	29.69	29.96680	28.60014	31.39876
28-01-16	28.75	29.95979	28.60064	31.38352
29-01-16	29.51	29.94493	28.59342	31.36032
01-02-16	29.57	29.93538	28.59103	31.34293
02-02-16	29.06	29.92924	28.59169	31.32937
03-02-16	27.68	29.92532	28.59425	31.31835
04-02-16	29.15	29.92278	28.59792	31.30901
05-02-16	27.97	29.92116	28.60228	31.30086
08-02-16	27.05	29.92012	28.60698	31.29352
09-02-16	26.82	29.91946	28.61187	31.28680

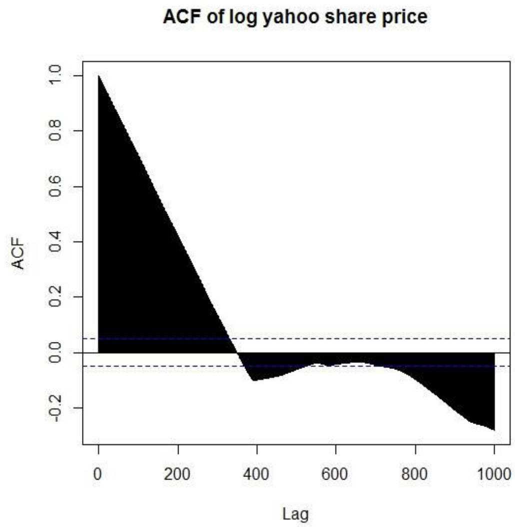
(all the values are in **dollar**)

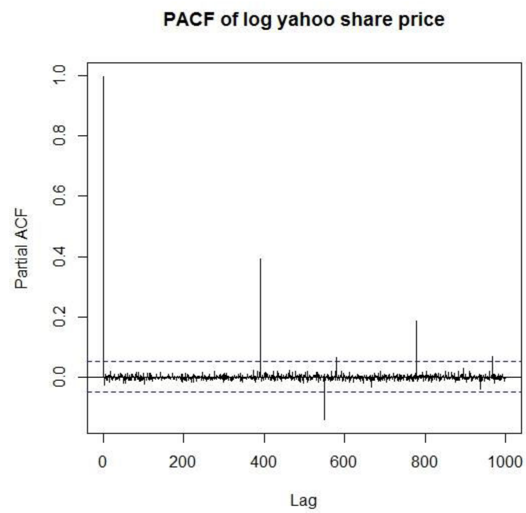
Time Series Analysis of Yahoo Share price of Japan

The plots of original data with log transformation are ,



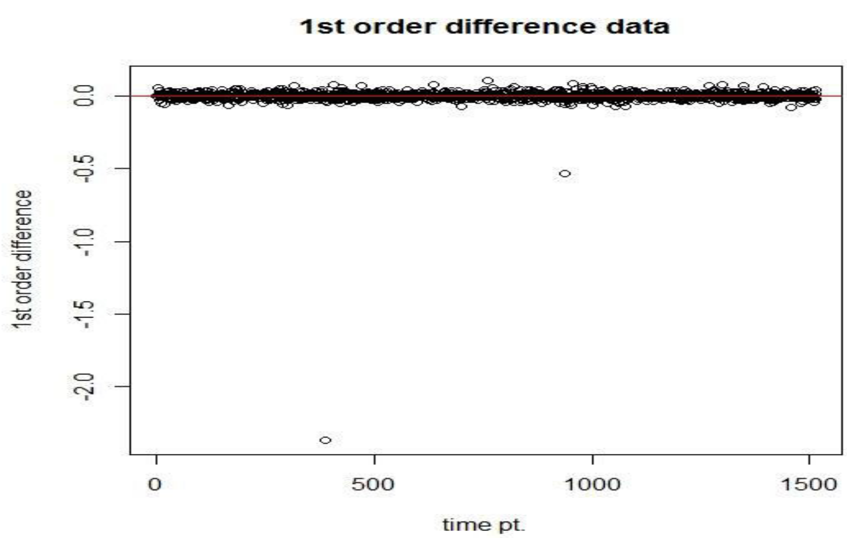
Now, the ACF, PACF plots of log transformed data are,





- ✓ The plot shows prominent decreasing trend with two sharp jumps at time points t=389 and t=937, i.e., at 03/08/2011 and 08/10/2013 respectively. The first sharp jump is assumed to be caused by the **effect of earthquake and tsunami** that hit Japan in 2011.
- ✓ To find whether monotonic trend is present or not **Mann kendal test** result is tau = -0.604, 2-sided ***p-value =< 2.22e-16*** (i.e., H_0 $is\ rejected$). Hence the test also supports presence of significant trend.

Now following the very common way one can procced for detrending by considering 1^{st} order difference. The plot of 1^{st} order difference is as follow,

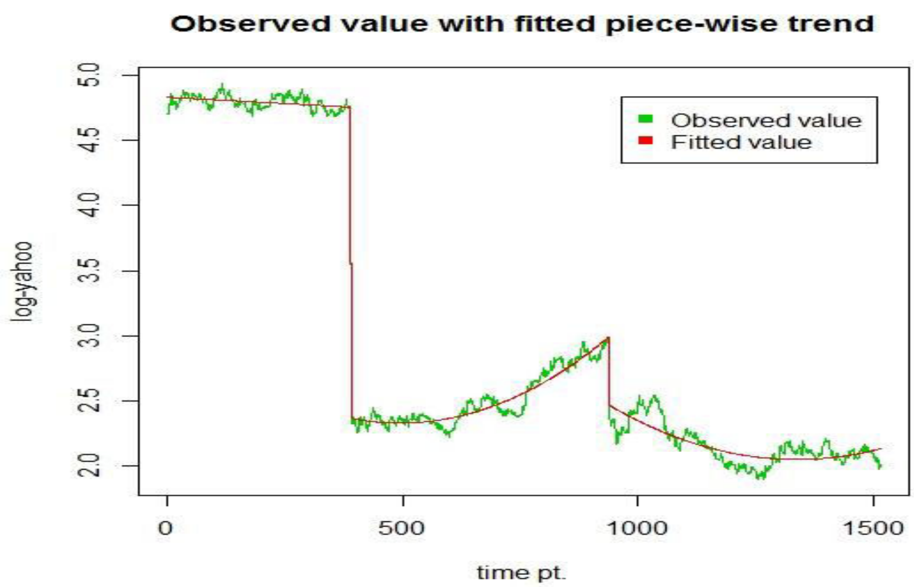


This plot shows **presence of two very big outliers**. Due to this reason, differencing is discarded.

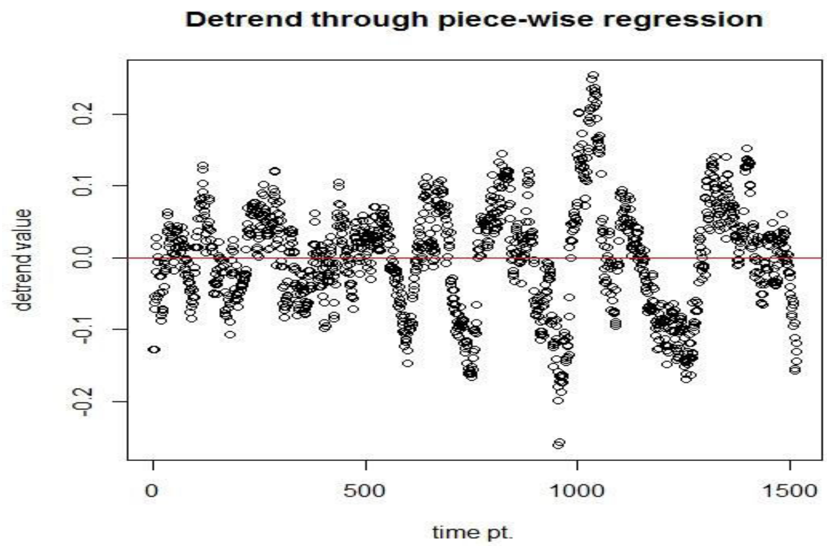
Appropriate **mathematical curve** is fitted to detrend the data. Here ***piece wise*** linear and quadratic ***splines*** are fitted. For time point 1:389 linear trend is fitted, and for time point 390:937 and 938:1515 quadratic trends are fitted. These choices are subjective, based on the plot of data. The coefficients for trend curves are as follows,

Time point	***intercept***	***time***	***time^2***
1:389	4.83345	-0.0001942	----
390:937	3.154	-3.337*10^{-3}	3.372*10^{-6}
938:1515	6.744	-7.028*10^{-3}	2.630 *10^{-6}

Plot of fitted trend with original log transformed data is,

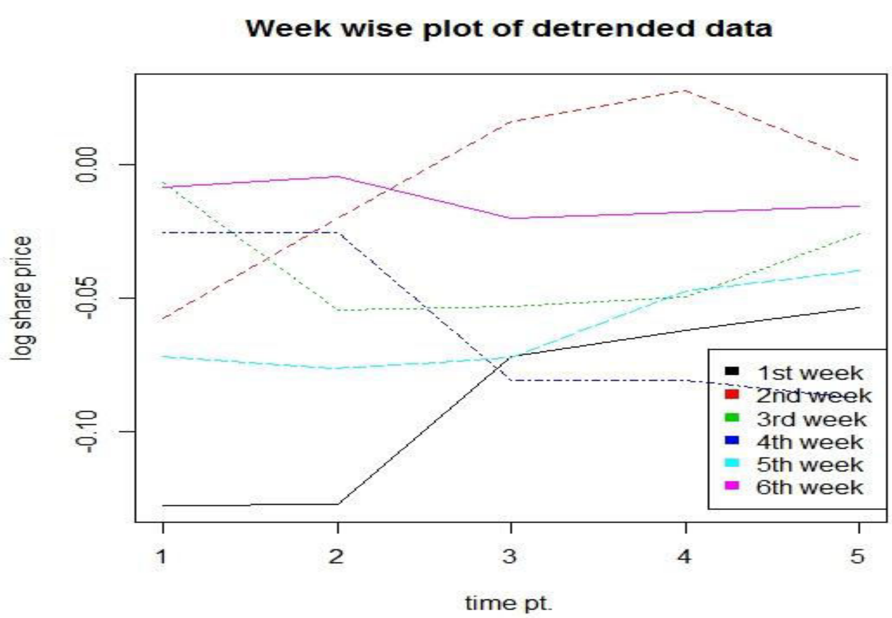


The detrended values are obtained by subtracting fitted values from observed values. The plot of detrended data is shown below.

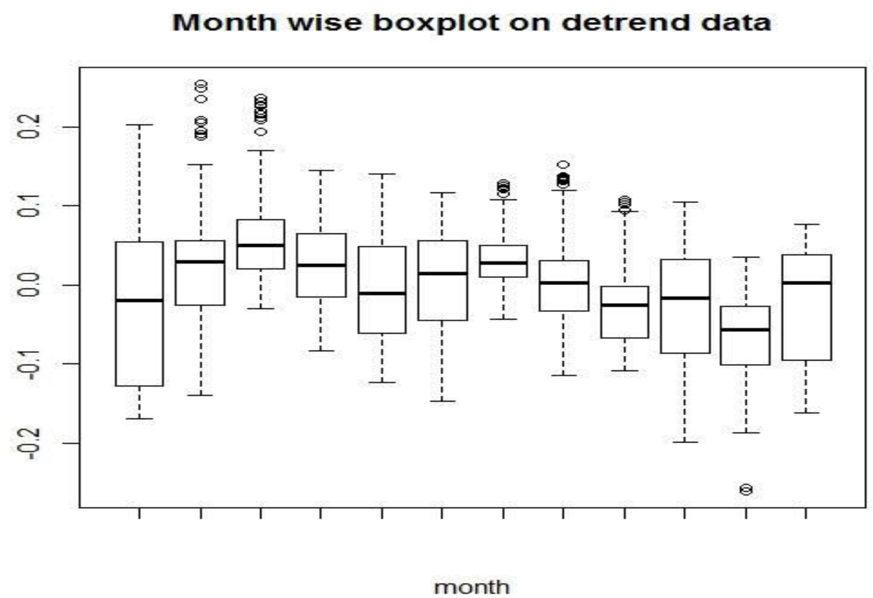


- ✓ The plot shows more or less a random fluctuation about The line Y=0, hence the trend is considered to be removed.
- ✓ Moreover, **Mann kendal test** on detrended data gives tau = -0.00171, 2-sided ***p-value =0.92061*** (i.e., $H_0 is\ accepted$) confirming absence of significant trend.

Now the week wise plot of first few detrended data is plotted, which shows absence of seasonal (weekly) fluctuation



But month wise boxplot says significant seasonal (monthly) pattern is present. The plot is as follow,

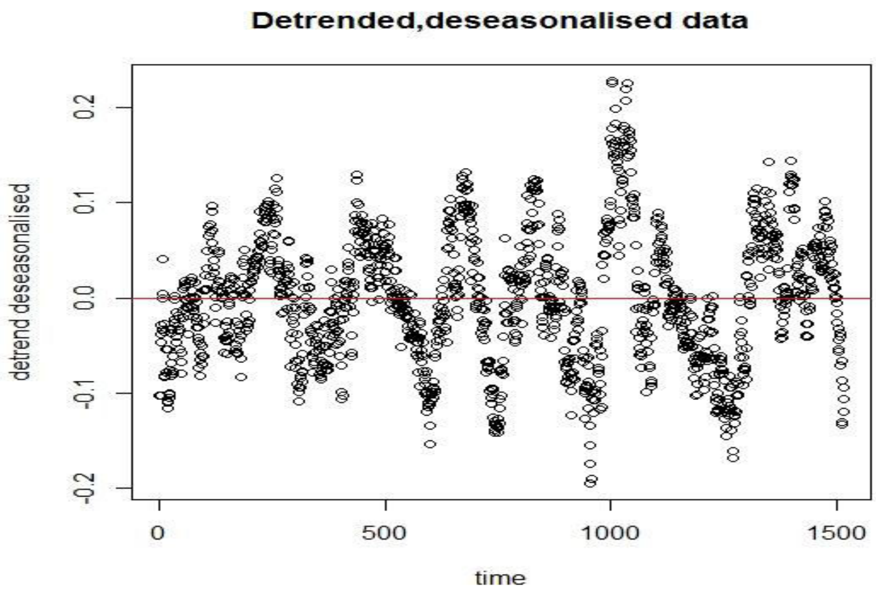


Kruskal-Wallis test : chi-squared = 291.0448, df = 11, ***p-value < 2.2e-16*** (i.e., H_0 $is\ rejected$), which implies significant monthly seasonality is present .
[***Note:*** this test is performed assuming month-wise variance not to differ significantly]

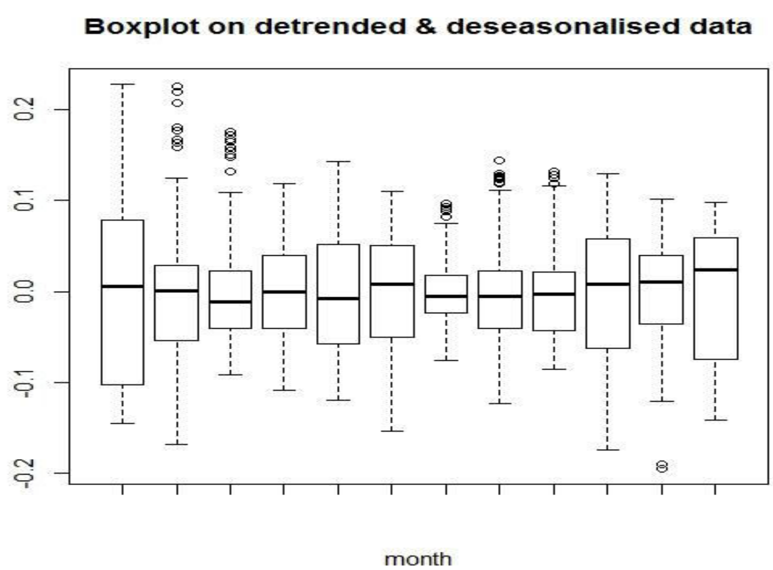
Hence the next step is to remove monthly seasonality. This is done by method of average. The **seasonal indices** are calculated as follow,

Month	***Seasonal indices***	***Month***	***Seasonal indices***
January	-0.024448540	July	0.032761126
February	0.028520220	August	0.008119021
March	0.061377242	September	-0.023294866
April	0.025797036	October	-0.024831424
May	-0.002915900	November	-0.066095090
June	0.005856755	December	-0.020845580

Now the detrended data is deseasonalised by subtracting corresponding seasonal indices. The plot is as follows,

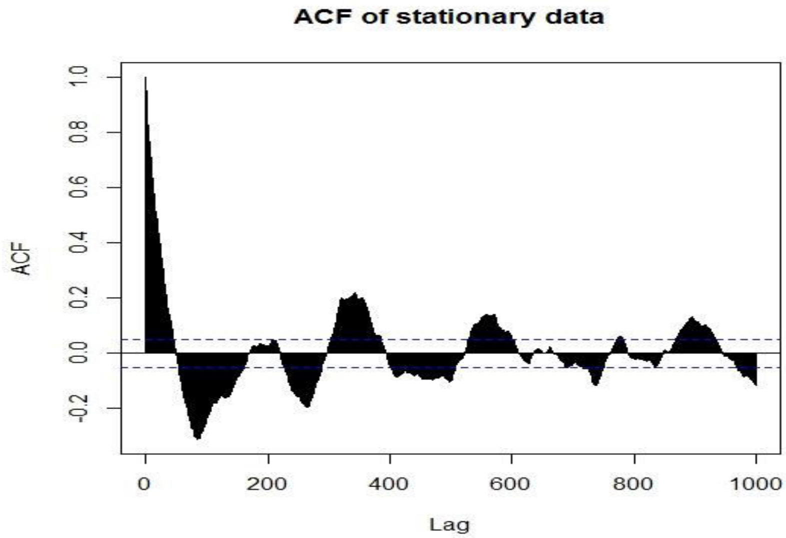


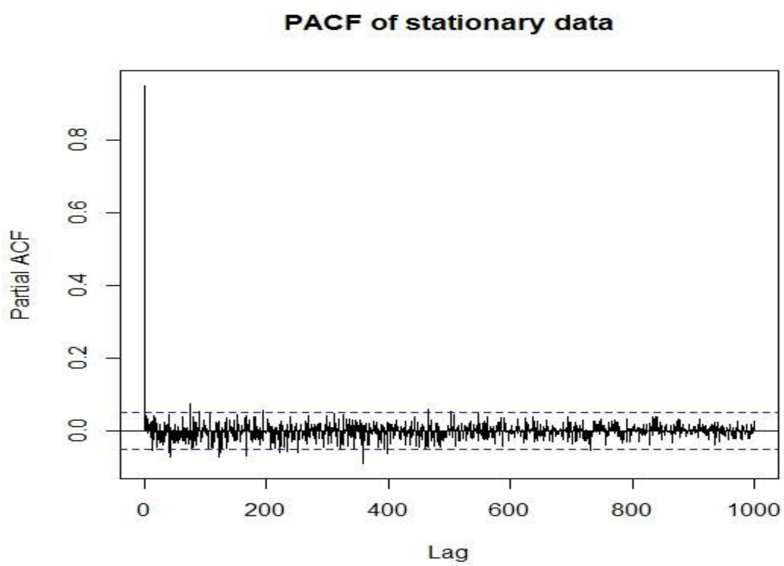
We draw boxplot and perform Kruskul wallis test on this deseasonalised data for rechecking of whether seasonality is removed or not. Both supports removal of seasonality.



✓ **Kruskal-Wallis test** : chi-squared = 3.1963, ***p-value = 0.9879*** (i.e. H_0 *is accepted*). Hence we can consider this detrended and deseasonalised data as stationary data.

The ACF,PACF plots of stationary data are as follows,





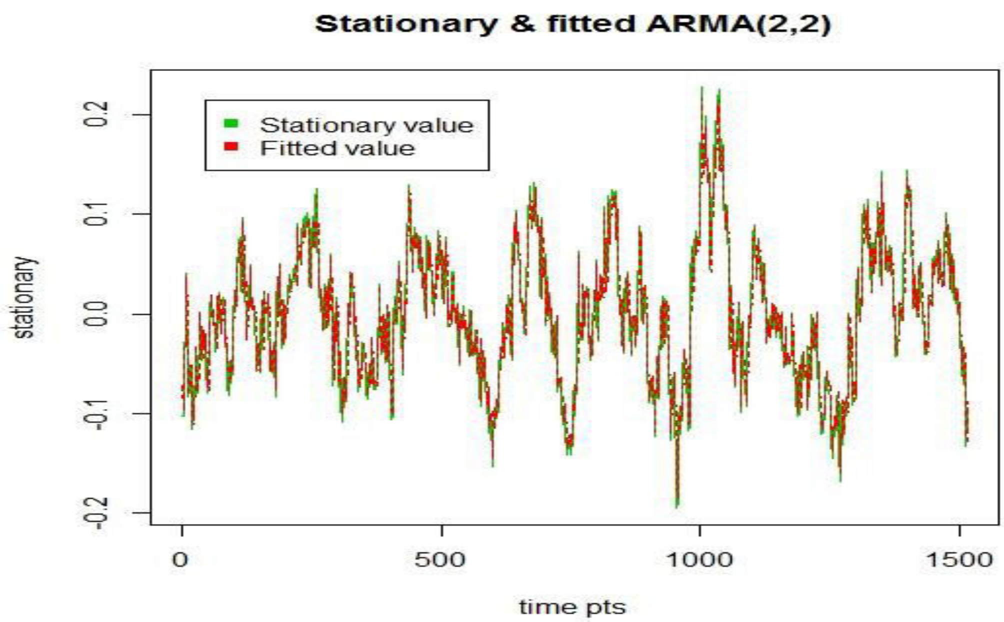
- ✓ PACF tails off and ACF shows few significant spikes, hence MA(q) model with a large q value is appropriate.
- ✓ To keep the model parsimonius we opt for ARMA (p,q) model on stationary data.

The model parameters chosen by $\boldsymbol{min^m}$ **AIC criterion**. The model is ***ARMA(2,2)*** [on stationary data] as follow , detrended and deseasonalised Yahoo share price of Japan Z_{2t} at t^{th} time point,

$$\boldsymbol{Z_{1t} = -0.0005 - 0.0356 * Z_{1\overline{t-1}} + 0.9437 * Z_{1\overline{t-2}} + \varepsilon_t + 0.9809 * \varepsilon_{t-1} - 0.0191 * \varepsilon_{t-2}}$$

With aic = -7458.5

The following plot shows fitted value with stationary value .

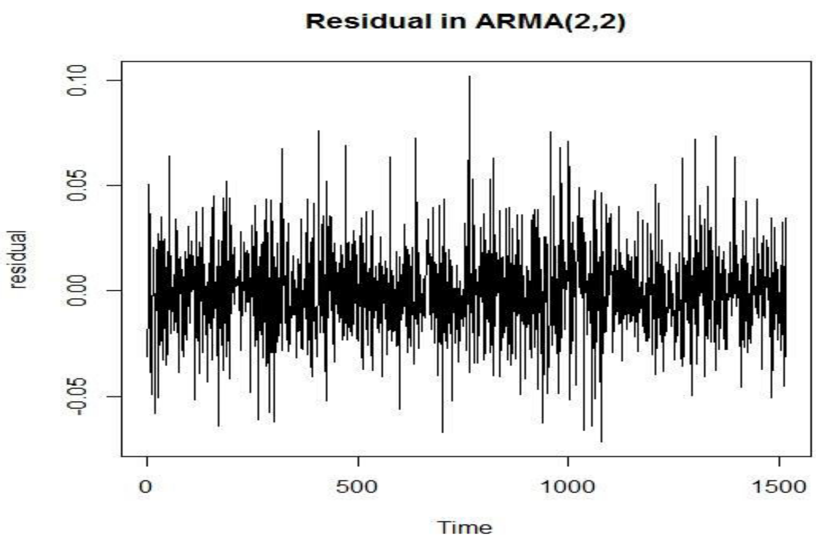


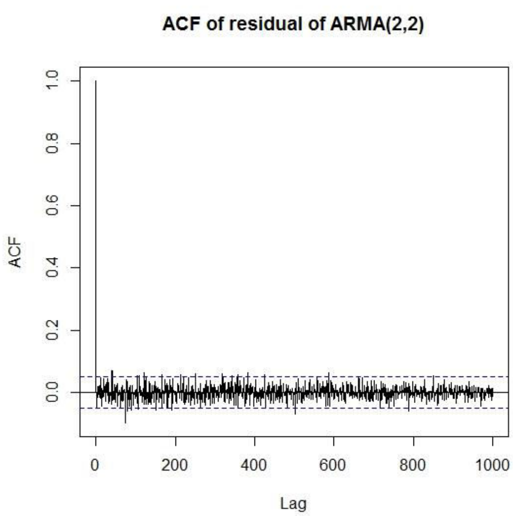
Now the ARMA(2,2) residuals are calculated and Parformance of **Ljung-Box test** on residuals is as follow

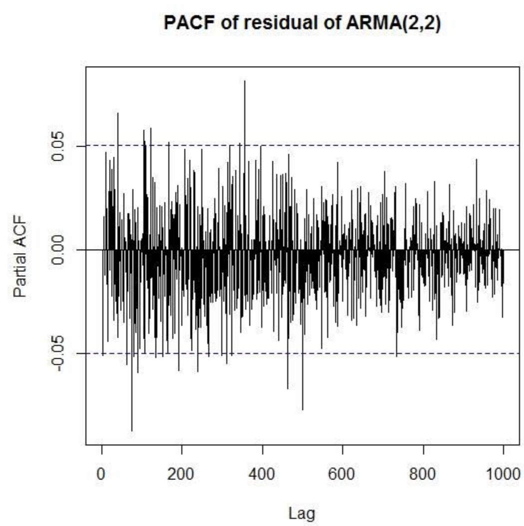
Lags	***Statistic***	***df***	***p-value***
5	5.552310	5	0.3246753
10	7.725947	10	0.6573427
15	15.566587	15	0.4175824
20	21.002731	20	0.4085914

The test accepts H_0, which tells ARMA residuals are purely random. Hence a good fit.

The residual plot and ACF, PACF plot of residuals are,







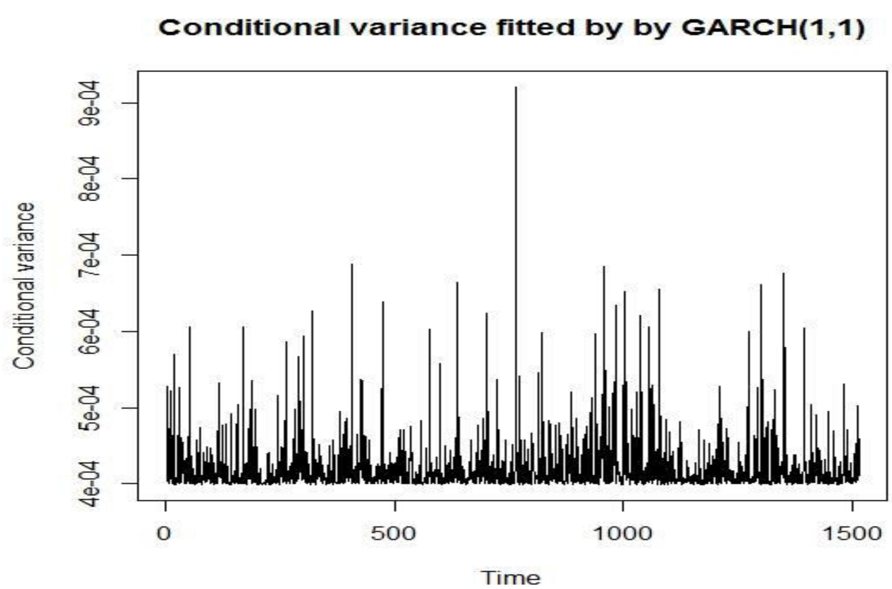
- ✓ The residual plot shows large fluctuations followed by small fluctuations and then again followed by large and this continues, then there is an indication that variation evolves over time.
- ✓ For confirmation, whether ARCH or GARCH model is suitable, the **Lagrange multiplier test** for ARCH(1) is as follow, Chi-squared = 2.105, df = 1, ***p-value = 0.04468****, w*hich confirms atleast ARCH(1) is required to model the volatility present in the ARIMA residual .

The optimum model parameter is chosen by min^m AIC criterion. The model is ***GARCH(1,1)***

ARMA residual,	$r_t = -2.835 * 10^{-6} + \varepsilon_t * \sqrt{\widehat{h_t}}$
Where	$\widehat{h_t} = 1.755 * 10^{-5} + 0.01459 * r_{t-1}^2 + 0.9435 * h_{t-1}$

✓ **Ljung-Box Test** on fitted GARCH (1,1) residual we find the model to be adqequate to capture volatility. The test resuli is , X-squared = 0.1372, ***p-value = 0.7111*** $(H_0\ is\ accepted)$, hence a good fit .

The plot of conditional variance, i.e., $\widehat{h_t}$ is as follow,

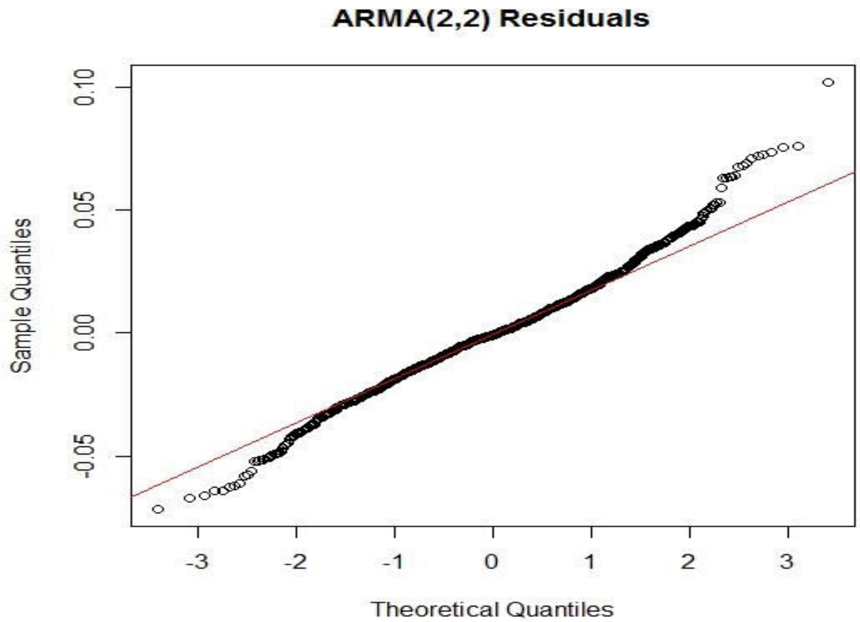


Now to find the C.I. of time series data, appropriate distribution of ARMA residual is found first .

Kurtosis $(r_t) =$ **4.35881**, which is higher than expected kurtosis of Normal $dist^n$.

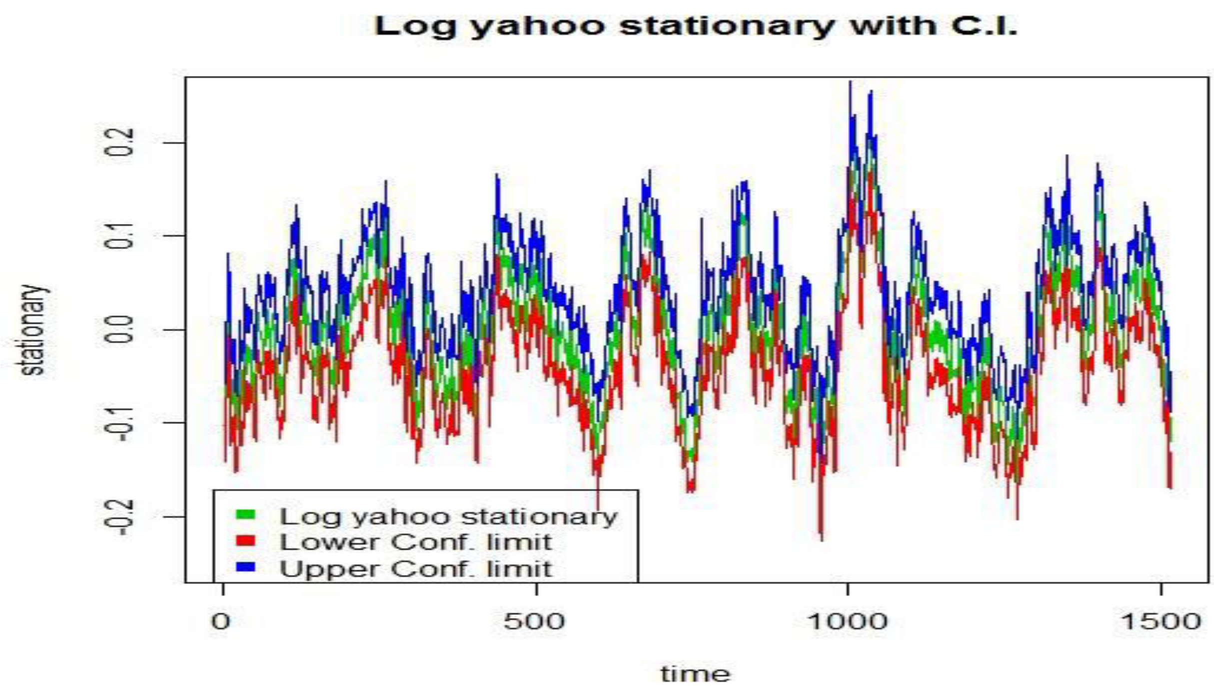
✓ **Anscome test** gives the result : kurt = 4.3588, z = 6.8133, ***p-value = 4.768e-12*** (H_0 is rejected). Hence kurtosis significantly differ from 3 .

✓ The **QQ-plot** also supports the fact obtained above .

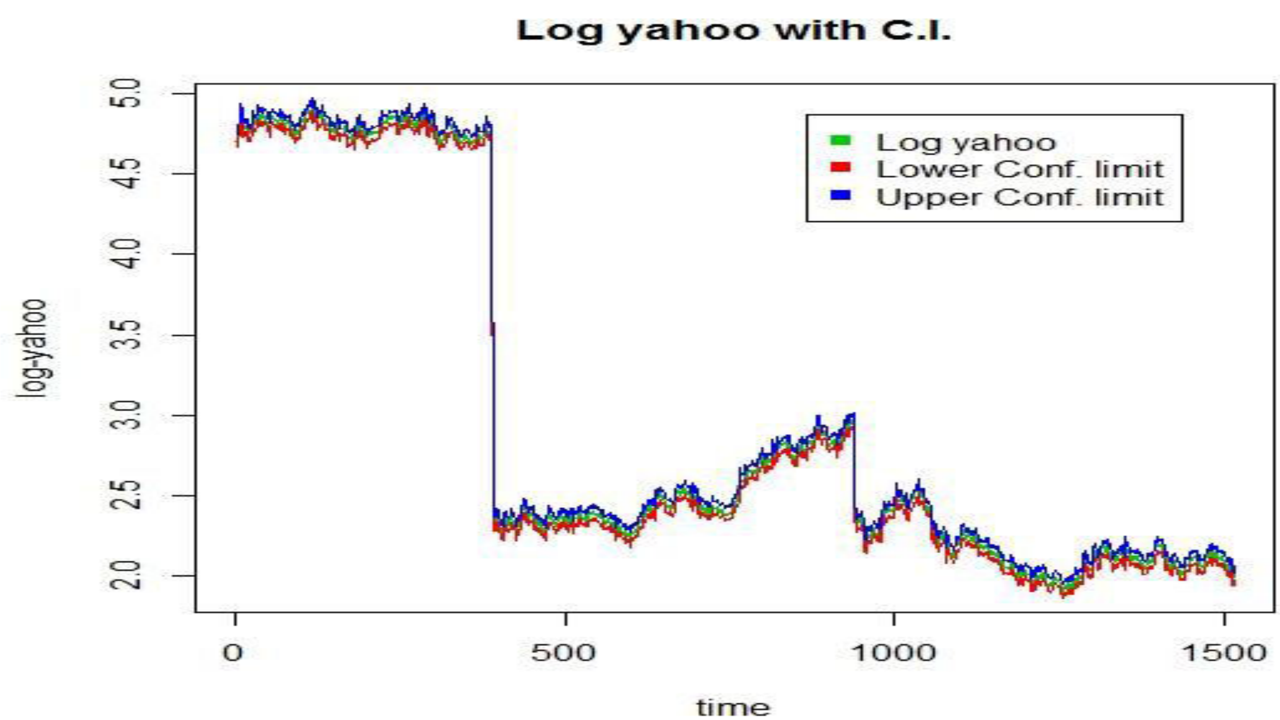


Having large frequency on tail observations , we fit '**t(ν)**' distribution. The value of ν is estimated from ARMA residual as , **ν = 5** . Then upper upper 100*(α/2)% point of t(ν=5) distribution is , $t_{\alpha/2} = 2.015048$, for α=.05.

Now the original log transformed stationary data with confidence interval is plotted as follow,



Now adding the fitted trend and corresponding seasonal indices with conf. limits we get conf. limits for original log transformed data.



Forecast:

Proceeding through the steps described in Forecast method the following table is obtained,

Time point	*Observed value*	*Predicted value*	*95 % lower conf limit*	*95 % upper conf limit*
27-01-16	7.44	7.462995	7.156204	7.782939
28-01-16	7.47	7.483270	7.175869	7.803841
29-01-16	7.58	7.544727	7.235017	7.867696
01-02-16	7.59	7.973986	7.646873	8.315092
02-02-16	7.74	8.035886	7.706445	8.379410
03-02-16	7.92	8.052431	7.722515	8.396442
04-02-16	7.74	8.111622	7.779476	8.457948
05-02-16	7.62	8.126072	7.793523	8.472811
08-02-16	7.78	8.182732	7.848046	8.531692
09-02-16	7.48	8.195284	7.860259	8.544589

(the values are in **yen**)

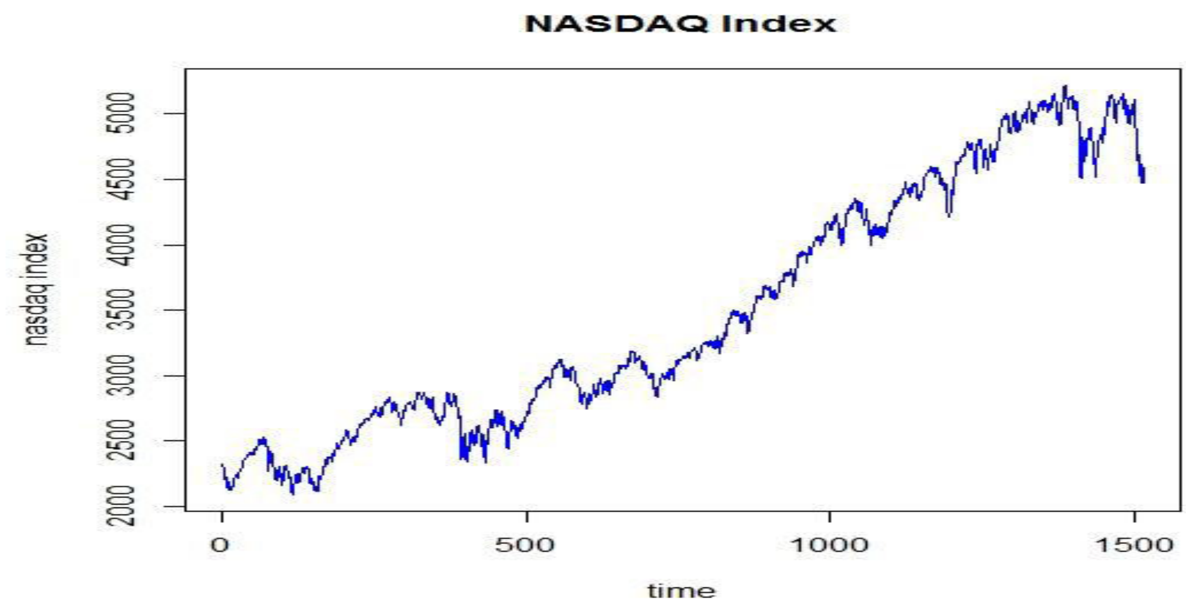
Remark:

The time series analysis on yahoo share price of both USA and Japan, provides quite a satisfactory prediction and forecast.

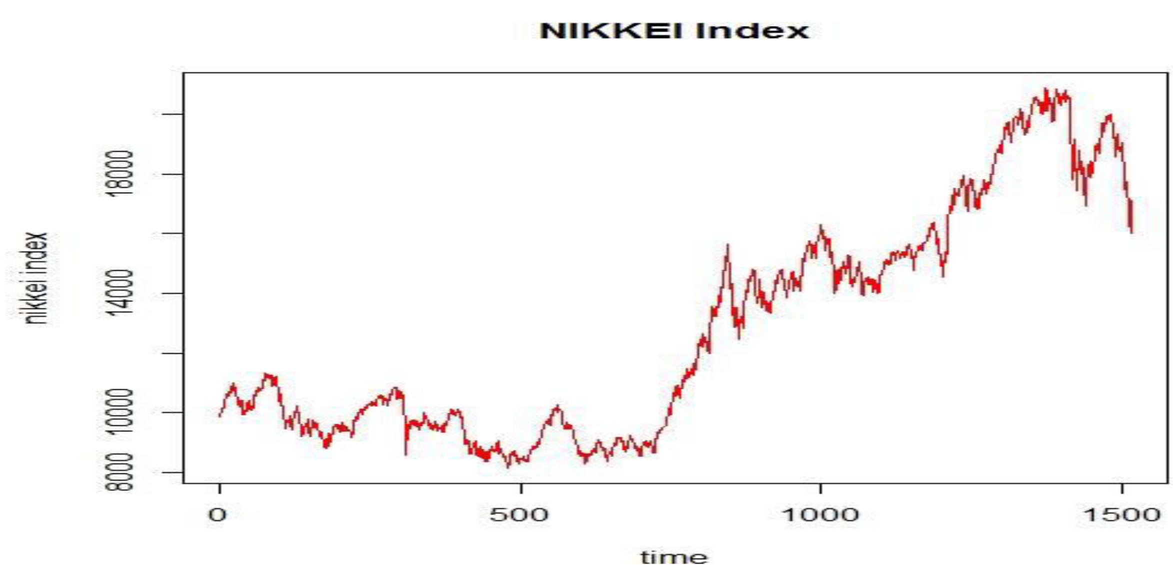
REGRESSION ANALYSIS

Description of exogenous variables:

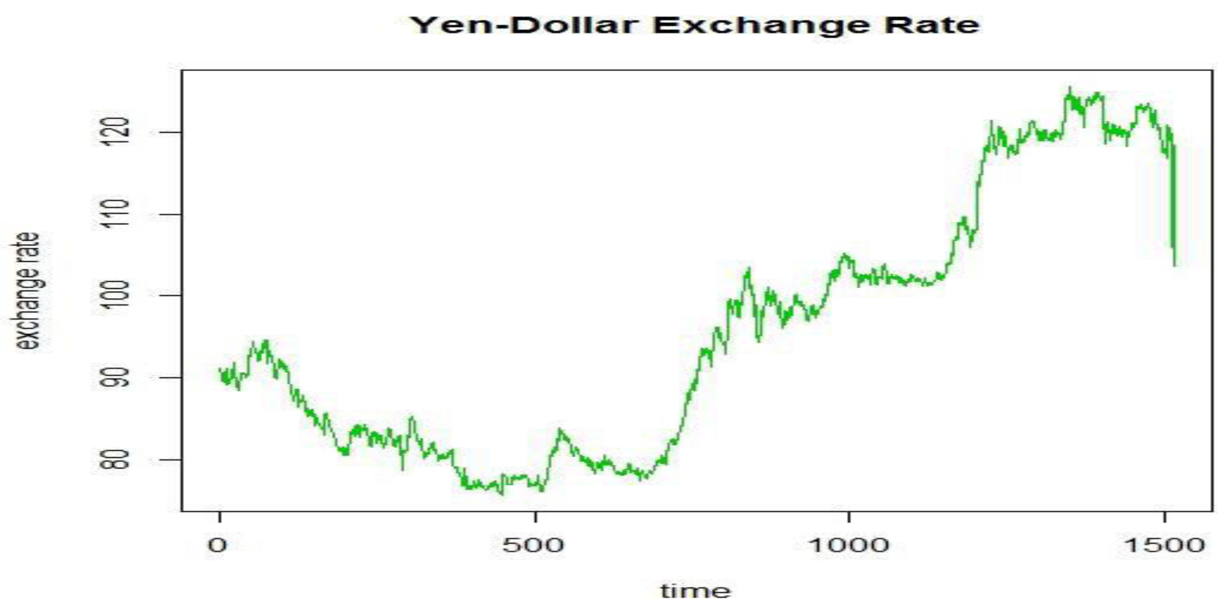
❖ **NASDAQ** (Nasdaq) is an American stock exchange. It is the second-largest exchange in the world by market capitalization.



❖ **Nikkei** Stock Average is a stock market index for the Tokyo stock exchange.



❖ **Yen-dollar exchange ratio** gives the amount of Yen which is equal to one dollar. There are some missing observations in this variable. The missing values have been estimated through **MICE** (Multiple Imputation by Chained Equations) algorithm. MICE is carried out considering the other two exogenous variables Nasdaq and Nikkei indices.



The exogenous variables are considered one lag before corresponding the endogenous variables. Considering the two endogenous variables, yahoo share price of USA and Japan, it is a multivariate time series. Hence a VAR model approach is appropriate. **Log transformation of all the variables** have been considered, as

- ✓ the transformation stabilizes the data
- ✓ it converts the multiplicative model (that is generally used to model Economical time series) into simpler additive model.

VAR model:

A VAR model describes the evolution of a set of k variables (called endogenous variables) over the same sample period (t = 1, ..., T) as a linear function of only their past values. The variables are collected in a k × 1 vector Y_t, which has as the i th element, Y_{it}, the observation at time "t" of the i th variable. A *p-th order VAR*, denoted **VAR(*p*)**, is

$$\boxed{\boldsymbol{Y_t = c + A_1\, Y_{t-1} + A_2\, Y_{t-2} + \ \dots . + A_p\, Y_{t-p} + \ e_t}}$$

where the *l*-periods back observation Y_{t-1} is called the *l*-th *lag* of *Y*, *c* is a $k \times 1$ vector of constants (intercepts), A_i is a time-invariant $k \times k$ matrix and e_t is a $k \times 1$ vector of error terms satisfying

1. $\boldsymbol{E(e_t) = 0}$ — every error term has mean zero;
2. $\boldsymbol{E(e_t\, e_t') = \Omega}$ — the contemporaneous covariance matrix of error terms is Ω (a $k \times k$ positive-semi definite matrix);
3. $\boldsymbol{E(e_t\, e_{t-k}') = 0}$ for any non-zero k — there is no correlation across time; in particular, no serial correlation in individual error terms.

A *p* th-order VAR is also called a VAR with *p* lags. The process of choosing the maximum lag *p* in the VAR model requires special attention because inference is dependent on correctness of the selected lag order. The order of VAR model is considered by **minimum AIC** criterion.

Here we observe ***VAR(1)*** model. The general form of VAR(1) in two endogenous variables Y_1, Y_2 with three exogenous variables X_1, X_2, X_3 can be written in matrix form as,

$$\boxed{\boldsymbol{\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Y_{1\overline{t-1}} \\ Y_{2\overline{t-1}} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}}}$$

Considering trend also, we get **fitted VAR(1)** model as,

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \begin{bmatrix} 0.28 \\ -0.24 \end{bmatrix} + \begin{bmatrix} 0.17 * 10^{-4} \\ -4.83 * 10^{-5} \end{bmatrix} * t + \begin{bmatrix} 1.01 & -0.7 * 10^{-3} \\ 4.6 * 10^{-3} & 0.98 \end{bmatrix} \begin{bmatrix} Y_{1\overline{t-1}} \\ Y_{2\overline{t-1}} \end{bmatrix}$$
$$+ \begin{bmatrix} -0.04 & 0.35 * 10^{-2} & -0.14 * 10^{-2} \\ 2.52 * 10^{-2} & -6.49 * 10^{-2} & 0.15 \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ \cdots \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

With Correlation matrix of residuals:

$$\begin{array}{cc} & \begin{array}{cc} Y_1 & Y_2 \end{array} \\ \begin{array}{c} Y_1 \\ Y_2 \end{array} & \begin{bmatrix} 1.0000 & 0.1531 \\ 0.1531 & 1.0000 \end{bmatrix} \end{array}$$

➢ Adequacy of fitted model:

A time series X is said to **Granger-cause** Y if it can be shown, usually through a series of t-tests and F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y. The **Granger causality test** is a statistical hypothesis test for determining whether one time series is useful in forecasting another.
Here the test is performed on the fitted model and obtained result is,

- ✓ ***H0***: Y_2(yahoo japan) do not Granger-cause Y_1(yahoo usa)
  F-Test = 0.3532, df1 = 1, df2 = 3014, ***p-value = 0.5524***

- ✓ ***H0***: Y_1(yahoo usa) do not Granger-cause Y_2(yahoo japan)
  F-Test = 0.1051, df1 = 1, df2 = 3014, ***p-value = 0.7458***

- ✓ ***H0***: No instantaneous causality between: Y_1 and Y_2
  Chi-squared = 34.5869, df = 1, ***p-value = 4.076e-09***

➢ Conclusion:

From the above tests we can conclude that, none of Y_1 and Y_2 are granger cause to each other. But there is a significant ***instantaneous causality*** between Y_1 and Y_2. Hence we can opt for ***Simultaneous Equations Model*** (SEM).

Simultaneous Equations Model (SEM):

➢ Structural and reduced form:

Suppose there are *m* regression equations of the form,

$$\boldsymbol{y_{it} = y'_{-i,t}\,\gamma_i + x'_{it}\,\beta_i + u_{it} \qquad i = 1(1)m}$$

- ✓ i is the equation number, and $t = 1, \ldots, T$ is the observation index.
- ✓ x_{it} is the $k_i \times 1$ vector of exogenous variables,
- ✓ y_{it} is the dependent variable,
- ✓ $y_{-i,t}$ is the $n_i \times 1$ vector of all other endogenous variables which enter the i^{th} equation on the right-hand side,
- ✓ u_{it} are the error terms.
- ✓ The regression coefficients β_i and γ_i are of dimensions $k_i \times 1$ and $n_i \times 1$ correspondingly.

Vertically stacking the *T* observations corresponding to the i^{th} equation, we can write each equation in vector form as

$$\boldsymbol{y_i = Y_{-i}\gamma_i + X_i\beta_i + u_i \qquad i = 1(1)m}$$

Finally, we can move all endogenous variables to the left-hand side and write the *m* equations jointly in vector form as,

$$\boldsymbol{Y\Gamma = XB + U}$$

This representation is known as the *"**structural form**"*

Post-multiplying the structural equation by Γ^{-1}, the system can be written in the *"**reduced form**"* as

$$\boldsymbol{Y = XB\Gamma^{-1} + U\Gamma^{-1} = X\Pi + V}$$

➢ Assumptions:

- ✓ Firstly, rank **(X)=k** , both in finite samples and in the limit as $T \longrightarrow \infty$. Matrix $\boldsymbol{\Gamma}$ is also assumed to be **non-degenerate**.
- ✓ Secondly, error terms are assumed to be serially independent and identically distributed. In particular, this implies that $\mathbf{E}[\boldsymbol{U}] = \mathbf{0}$, and $\mathbf{E}[\boldsymbol{U}'\boldsymbol{U}] = \boldsymbol{T\Sigma}$ **.**
- ✓ Lastly, the **identification conditions** require that the number of unknowns in this system of equations should not exceed the number of equations.
  a) More specifically, the ***order condition*** requires for each equation $\boldsymbol{k_i + n_i \le k}$**,** which can be phrased as "the number of excluded exogenous variables is greater or equal to the number of included endogenous variables".
  b) The ***rank condition*** of identifiability is that $\boldsymbol{rank\ (\Pi_{io}) = n_i}$, where Π_{io} is a $(k - k_i)\times n_i$ matrix which is obtained from Π by crossing out those columns which correspond to the excluded endogenous variables, and those rows which correspond to the included exogenous variables.

➢ Two-stages least squares (2SLS):

The simplest and the most common estimation method for the simultaneous equations model is the so-called **two-stage least square** method. It is an equation-by-equation technique, where the endogenous regressors on the right-hand side of each equation are being instrumented with the regressors *X* from all other equations. The method is called "two-stage" because it conducts estimation in two steps,

- **Step 1:** Regress Y_{-i} on X and obtain the predicted values, $\hat{Y}_{-i}$
- **Step 2:** Estimate γ_i, β_i by the ordinary least squares regression of y_i on $\hat{Y}_{-i}$ and X_i.

If the ith equation in the model is written as

$$\boldsymbol{y_i} = (\boldsymbol{Y_{-i}} \quad \boldsymbol{X_i})\begin{pmatrix}\boldsymbol{\gamma_i}\\ \boldsymbol{\beta_i}\end{pmatrix} + \boldsymbol{u_i} \equiv \boldsymbol{Z_i\delta_i} + \boldsymbol{u_i}$$

- ✓ Z_i is a $T\times(n_i + k_i)$ matrix of both endogenous and exogenous regressors in the i^{th} equation,
- ✓ δ_i is an $(n_i + k_i)$-dimensional vector of regression coefficients,

then the **2SLS estimator** of δ_i will be given by

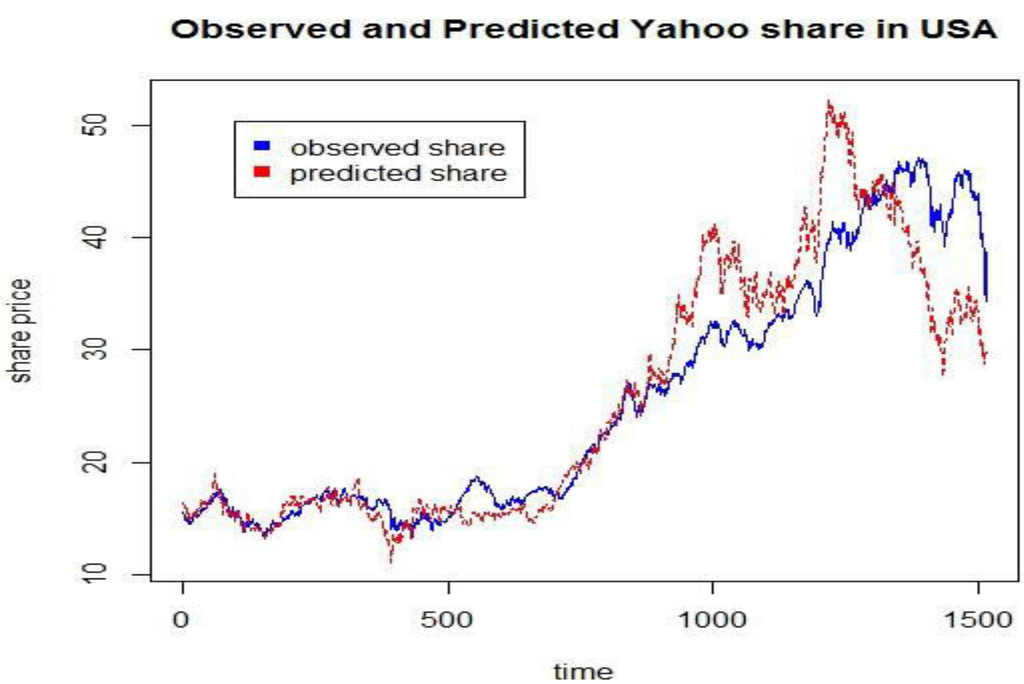
$$\boxed{\hat{\boldsymbol{\delta}}_i = \left(\hat{\boldsymbol{Z}}_i' \hat{\boldsymbol{Z}}_i\right)^{-1} \hat{\boldsymbol{Z}}_i' \boldsymbol{y}_i = (\boldsymbol{Z}_i' \boldsymbol{P} \boldsymbol{Z}_i)^{-1} \boldsymbol{Z}_i' \boldsymbol{P} \boldsymbol{y}_i}$$

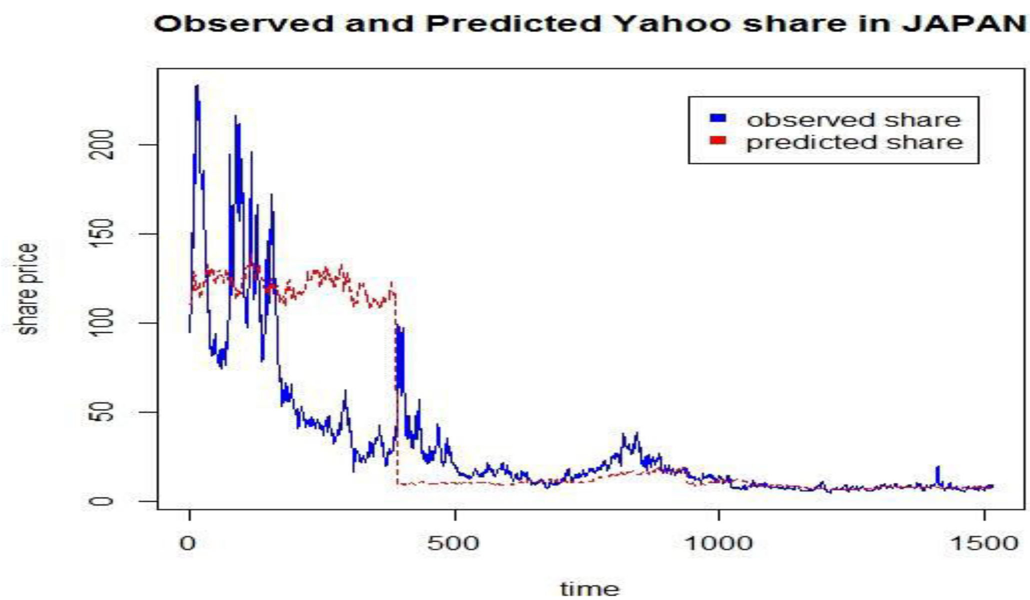
where $P = X(X'X)^{-1}X'$ is the projection matrix onto the linear space spanned by the exogenous regressors X.

Finally we get the following **reduced form** on **log transformed data** by applying **2SLS** method,

$$\boxed{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} -9.669 \\ 20.989 \end{bmatrix} + \begin{bmatrix} 0.862 & 0.290 & 0.677 \\ -6.612 & 4.066 & -0.574 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}}$$

Then reverse exponential transformation is done to get the predicted values of original series. The plots of predicted values of Y_1 and Y_2 through 2SLS are as follow,





➢ Forecast:

Forecast about later 10 days is as follow,

Time point	*Predicted* Y_1	*Observed* Y_1	*Predicted* Y_2	*Observed* Y_2
27-01-16	38.354600	29.69	6.921515	7.44
28-01-16	37.609248	28.75	7.671636	7.47
29-01-16	37.86620	29.51	7.36738	7.58
01-02-16	39.639950	29.57	7.341739	7.59
02-02-16	39.836134	29.06	7.658916	7.74
03-02-16	38.92950	27.68	9.04666	7.92
04-02-16	38.251513	29.15	9.093485	7.74
05-02-16	37.948725	27.97	8.798173	7.62
08-02-16	36.84478	27.05	10.56277	7.78
09-02-16	36.12601	26.82	12.63097	7.48

Remark:

The main reason of predicted values to differ is insufficient covariates. Share price actually depends on so many factors and their mutual interactions. Consideration of more covariates should improve the predictions.

References:

1. Introduction to Modern Time Series Analysis - Kirchgassner, Wolters, Hassler.
2. Applied Econometric Time Series - Walter Enders.
3. Econometric methods – Johnston, Dinardo.
4. wikipedia

Acknowledgements:

*I am highly indebted to Prof. **Sugata Sen Roy,** Prof. **Manisha Pal** and Prof. **Uttam Bandyaopadhyay** for their guidance, support and constant supervision in completion of the project.*

*I also want to offer my sincere gratitude to our department, **Department of Statistics, University of Calcutta** for giving me the opportunity to do this project.*

*I would also like to thank my **parents** and **friends** who helped me a lot in finalizing this project within the limited time frame.*

Snapshot of data: [Appendix : I]

Date	*usa*	*japan*	*Date_1*	*nasdaq*	*nikkei*	*exc ratio*	*time*
20-01-10	16.38	110.6	19-01-10	2320.4	9862.82	91.23	1
21-01-10	16.2	110.6	20-01-10	2291.25	10107.87	90.25	2
22-01-10	15.88	116.9	21-01-10	2265.7	10105.68	90.07	3
25-01-10	15.86	117.9999	22-01-10	2205.29	10083.48	90.12	4
26-01-10	15.99	118.9999	25-01-10	2210.8	10177.41	89.63	5
27-01-10	15.98	118.4999	26-01-10	2203.73	10163.8	89.41	6
28-01-10	15.44	123	27-01-10	2221.41	10142.05	90.08	7
29-01-10	15.01	127.5	28-01-10	2179	10183.47	90.38	8
01-02-10	15.05	128.95	29-01-10	2147.35	10378.03	90.8	9
02-02-10	15.17	125.6001	01-02-10	2171.2	10536.92	90.39	10
03-02-10	15.46	124.55	02-02-10	2190.06	10494.71	91.11	11
04-02-10	15.01	118.7499	03-02-10	2190.91	10634.23	89.11	12
05-02-10	15.19	118.85	04-02-10	2125.43	10638.06	89.25	13
08-02-10	14.99	119.2499	05-02-10	2141.12	10546.44	89.32	14
09-02-10	15.07	122.1	08-02-10	2126.05	10654.79	89.45	15
10-02-10	14.8	122.1	09-02-10	2150.87	10681.83	89.88	16
11-02-10	15.22	122.1	10-02-10	2147.87	10731.45	89.82	17
12-02-10	15.17	115.5	11-02-10	2177.41	10681.66	90	18
16-02-10	15.41	115.5	12-02-10	2183.53	10798.32	90.46	19
17-02-10	15.44	114.7	16-02-10	2214.19	10879.14	90.79	20
18-02-10	15.54	116.45	17-02-10	2226.29	10735.03	91.43	21
19-02-10	15.58	115.95	18-02-10	2241.71	10907.68	91.94	22
22-02-10	15.49	116.3501	19-02-10	2243.87	10982.1	91.11	23
23-02-10	15.38	119.32	22-02-10	2242.03	10855.08	90.04	24
24-02-10	15.59	120.1999	23-02-10	2213.44	10764.9	90.06	25
25-02-10	15.24	123.9799	24-02-10	2235.9	10737.52	88.85	26
26-02-10	15.31	124.45	25-02-10	2234.22	10868.41	88.84	27
01-03-10	15.79	122.5	26-02-10	2238.26	10590.55	89.28	28
02-03-10	15.73	122.75	01-03-10	2273.57	10512.69	88.95	29
03-03-10	15.57	123	02-03-10	2280.79	10325.28	88.43	30
04-03-10	15.81	124.9001	03-03-10	2280.68	10252.08	89.02	31
05-03-10	16.06	129.8001	04-03-10	2292.31	10414.29	90.32	32
08-03-10	16.52	133	05-03-10	2326.35	10198.04	90.3	33
09-03-10	16.53	132.5	08-03-10	2332.21	10205.02	89.89	34
10-03-10	16.79	129.2499	09-03-10	2340.68	10371.09	90.62	35
11-03-10	16.53	127.4301	10-03-10	2358.95	10404.33	90.6	36
12-03-10	16.32	130.6	11-03-10	2368.46	10355.98	90.46	37
15-03-10	16.46	126.5201	12-03-10	2367.66	10057.09	90.45	38
16-03-10	16.36	126.5	15-03-10	2362.21	9951.82	90.43	39
17-03-10	16.5	130.1	16-03-10	2378.01	9932.9	90.42	40
18-03-10	16.56	130.1	17-03-10	2389.09	9963.99	90.46	41
19-03-10	16.44	128.0101	18-03-10	2391.28	10092.19	90.5	42
22-03-10	16.34	128.65	19-03-10	2374.41	10013.3	90.02	43
23-03-10	16.03	128.15	22-03-10	2395.4	10034.25	90.24	44
24-03-10	16.09	129.65	23-03-10	2415.24	10306.83	91.89	45
25-03-10	16.32	125.4001	24-03-10	2398.76	10335.69	92.69	46
26-03-10	16.54	122.33	25-03-10	2397.41	10123.58	92.56	47
29-03-10	16.56	123.7299	26-03-10	2395.13	10400.47	92.65	48
30-03-10	16.61	124.9801	29-03-10	2404.36	10352.1	92.89	49
31-03-10	16.53	122.14	30-03-10	2410.69	10198.83	93.4	50
01-04-10	16.29	126	31-03-10	2397.96	10101.96	NA	51
05-04-10	16.51	125	01-04-10	2402.58	10126.03	94.51	52

Date	*usa*	*japan*	*Date_1*	*nasdaq*	*nikkei*	*exc ratio*	*time*
10-02-15	43.07	7.12	09-02-15	4726.01	17335.85	118.84	1274
11-02-15	42.96	7.13	10-02-15	4787.64	17678.74	118.68	1275
12-02-15	43.93	7.15	11-02-15	4801.18	17504.62	118.88	1276
13-02-15	44.42	7.27	12-02-15	4857.61	17648.5	119.13	1277
17-02-15	43.53	7.27	13-02-15	4893.84	17711.93	118.88	1278
18-02-15	43.65	7.37	17-02-15	4899.27	17652.68	119.36	1279
19-02-15	44.37	7.35	18-02-15	4906.36	17652.68	119.72	1280
20-02-15	44.11	7.27	19-02-15	4924.7	17979.72	120.06	1281
23-02-15	43.53	7.26	20-02-15	4955.97	17913.36	119.47	1282
24-02-15	43.38	7.55	23-02-15	4960.97	18004.77	119.76	1283
25-02-15	44.43	7.73	24-02-15	4968.12	17987.09	120.22	1284
26-02-15	44.45	8	25-02-15	4967.14	18199.17	120.93	1285
27-02-15	44.28	8.04	26-02-15	4987.89	18264.79	121.17	1286
02-03-15	44.11	8.01	27-02-15	4963.53	18332.3	121.2	1287
03-03-15	42.62	7.92	02-03-15	5008.1	18466.92	121.5	1288
04-03-15	43.99	7.86	03-03-15	4979.9	18603.48	121.28	1289
05-03-15	44.16	8.1	04-03-15	4967.14	18585.2	121.17	1290
06-03-15	43.44	7.7	05-03-15	4982.81	18785.79	121.3	1291
09-03-15	42.98	7.79	06-03-15	4927.37	18797.94	121.28	1292
10-03-15	42.68	7.65	09-03-15	4942.44	18826.88	120.92	1293
11-03-15	42.5	7.56	10-03-15	4859.79	18815.16	120.9	1294
12-03-15	42.95	7.71	11-03-15	4849.94	18703.6	120.28	1295
13-03-15	42.87	7.68	12-03-15	4893.29	18751.84	119.74	1296
16-03-15	43.51	7.69	13-03-15	4871.76	18971	119.9	1297
17-03-15	43.79	7.6	16-03-15	4929.51	18790.55	119.37	1298
18-03-15	44.67	7.85	17-03-15	4937.43	18665.11	119.01	1299
19-03-15	44.98	8.46	18-03-15	4982.83	18723.52	119.15	1300
20-03-15	45.04	8.47	19-03-15	4992.38	18991.11	120.11	1301
23-03-15	44.72	8.67	20-03-15	5026.42	19254.25	119.96	1302
24-03-15	44.42	8.65	23-03-15	5010.97	19246.06	119.62	1303
25-03-15	44.2	8.59	24-03-15	4994.73	19437	119.74	1304
26-03-15	44.47	8.28	25-03-15	4876.52	19544.48	118.96	1305
27-03-15	45.1	8.21	26-03-15	4863.36	19476.56	119.05	1306
30-03-15	44.95	8.2	27-03-15	4891.22	19560.22	120.36	1307
31-03-15	44.44	8.24	30-03-15	4947.44	19560.22	119.96	1308
01-04-15	44.13	8.1	31-03-15	4900.88	19713.45	120.32	1309
02-04-15	44.15	8.07	01-04-15	4880.23	19746.2	120.29	1310
06-04-15	43.67	8.38	02-04-15	4886.94	19471.12	120.32	1311
07-04-15	43.61	8.42	06-04-15	4917.32	19285.63	119.26	1312
08-04-15	45.17	8.74	07-04-15	4910.23	19411.4	119.23	1313
09-04-15	45.63	8.9	08-04-15	4950.82	19206.99	119.18	1314
10-04-15	45.18	8.84	09-04-15	4974.56	19034.84	119.02	1315
13-04-15	44.77	8.8	10-04-15	4995.98	19312.79	119.27	1316
14-04-15	45.53	8.84	13-04-15	4988.25	19435.08	119.45	1317
15-04-15	45.73	8.79	14-04-15	4977.29	19397.98	119.9	1318
16-04-15	45.78	8.7	15-04-15	5011.02	19640.54	119.69	1319
17-04-15	44.45	8.61	16-04-15	5007.79	19789.81	118.98	1320
20-04-15	44.66	8.56	17-04-15	4931.81	19937.72	119.12	1321
21-04-15	44.49	8.67	20-04-15	4994.6	19907.63	118.8	1322
22-04-15	43.98	8.93	21-04-15	5014.1	19905.46	118.83	1323
23-04-15	43.7	8.53	22-04-15	5035.17	19908.68	119.86	1324
24-04-15	44.52	8.48	23-04-15	5056.06	19869.76	120.21	1325
27-04-15	44.36	8.56	24-04-15	5092.08	19885.77	120.25	1326
28-04-15	44.34	8.38	27-04-15	5060.25	19652.88	119.87	1327

R – codes: [Appendix : II]

❖ code for Yahoo time series analysis of USA:

```
z=read.csv(file="D:/My Documents/Project_ Yahoo/a_yahoo usa _final.csv",header = T,sep=",")
cl=z$Close
plot(cl,main="Yahoo Closing share price in USA",xlab="time points",ylab="share
price",type="l",col=3)
lcl=log(cl)
plot(lcl,main="Log Yahoo Closing share price in USA",xlab="time points",ylab="log share
price",type="l",col=4)
y=lcl  ## study variable
acf(y,lag.max=1000,main="ACF of log-yahoo")
pacf(y,lag.max=1000,main="PACF of log-yahoo")
 ##------------ check of trend ------------##
library(Kendall)
MannKendall(y)     ## monotonic trend present
d1=diff(y)
plot(d1,main="1st order difference data",xlab="time pt.",ylab="1st order difference")
abline(h=0,col=2)
MannKendall(d1)    ## monotonic trend absent  ## hence we can treat 1st order diff data as detrended
data.
##------------ check of seasonality ---------------##
m1=c(d1[1:5])
m2=c(d1[6:10])
m3=c(d1[11:15])
m4=c(d1[16:20])
m5=c(d1[21:25])
m6=c(d1[26:30])
m=cbind(m1,m2,m3,m4,m5,m6)
matplot(m,type="l",main="Week wise plot of detrend data",ylab="log share price",xlab="time
points",col=c(1,2,3,4,5,6))
legend(locator(1),col=c(1,2,3,4,5,6),legend=c("1st week","2nd week","3rd week","4th week","5th
week","6th week"),pch=15)
```

```
d2=d1  ## data is sorted along month.
jan=c(d2[1:8],d2[242:261],d2[494:513],d2[744:764],d2[996:1016],d2[1248:1267],d2[1500:1515])
feb=c(d2[9:27],d2[262:280],d2[514:533],d2[765:783],d2[1017:1035],d2[1268:1286])
mar=c(d2[28:50],d2[281:303],d2[534:554],d2[784:803],d2[1036:1056],d2[1287:1308])
apr=c(d2[51:71],d2[304:323],d2[555:575],d2[804:825],d2[1057:1077],d2[1309:1329])
may=c(d2[72:91],d2[324:344],d2[576:597],d2[826:847],d2[1078:1098],d2[1330:1349])
june=c(d2[92:113],d2[345:366],d2[598:618],d2[848:867],d2[1099:1118],d2[1350:1371])
july=c(d2[114:134],d2[367:386],d2[619:639],d2[868:889],d2[1119:1141],d2[1372:1393])
aug=c(d2[135:156],d2[387:409],d2[640:662],d2[890:911],d2[1142:1162],d2[1394:1414])
sept=c(d2[157:177],d2[410:430],d2[663:681],d2[912:931],d2[1163:1184],d2[1415:1435])
oct=c(d2[178:198],d2[431:451],d2[682:702],d2[932:954],d2[1185:1206],d2[1436:1457])
nov=c(d2[199:219],d2[452:472],d2[703:723],d2[955:974],d2[1207:1225],d2[1458:1477])
dec=c(d2[220:241],d2[473:493],d2[724:743],d2[975:995],d2[1226:1247],d2[1478:1499])
boxplot(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec,main="Month wise boxplot on detrend
data",xlab="month")
kruskal.test(list(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec))
acf(d1,lag.max=1000,main="ACF of stationary data")
pacf(d1,lag.max=1000,main="PACF of stationary data")

 ##-------------------- MEAN stationary---------------##
library(forecast)
for(i in 0:4)
{ for(k in 0:4)
{ b1=arima(y,order=c(i,1,k))
print(i)
print(k)
print(AIC(b1))
}}   ## on min AIC, ARIMA(3,1,1) is selected
f1=arima(y,order=c(1,1,3))
f2=fitted(f1)  ##predicted value
g=cbind(y,f2)
matplot(g,type="l",col=c(3,2),main="Observed data & fitted
ARIMA(1,1,3)",ylab="log_yahoo",xlab="time pts")
legend(locator(1), legend =c("Observed data","Fitted value"),col=c(3,2), pch=15)
resid=y-f2  ## residual of ARIMA(1,1,3)
plot(resid,main="Residual in ARIMA(1,1,3)",ylab="residual",col=1)
acf(resid,main="ACF of residual of ARIMA(1,1,3)",lag.max=1000)
pacf(resid,main="PACF of residual of ARIMA(1,1,3)",lag.max=1000)
library(portes)
```

```
portest(resid,test="LjungBox")  ## white noise is absent in ARIMA(3,1,1) residual
library(FinTS)  ##  test for arch/garch   ##
ArchTest(resid,lags=1)
##----------------- VARIANCE stationary --------------##
library(tseries)
for(i in 0:3)
{ for(j in 1:3)
{ ff=garch(resid,order=c(i,j),mse="uncond",trace=FALSE)
print(i)
print(j)
print(AIC(ff))
}}  ## hence on basis of min aic garch(1,1)
 f=garch(resid,order=c(1,1),mse="uncond",trace=FALSE)
print(AIC(f))
library(fGarch)
fit=garchFit(~ garch(1,1),data=resid)
fit
 ##---------- Generate plot of Log Price, 95% Upper and Lower limit -------##
f2=fitted.values(f1)  ## fitted mean by ARIMA(3,1,1)
ht=f$fit[,1]^2  ## fitted conditional standard deviation by GARCH(1,1)
plot(ht,main='Conditional variance fitted by by GARCH(1,1)',ylab="Conditional variance",col=1)
library(moments)
kurtosis(resid)
qqnorm(resid,main='ARIMA(1,1,3) Residuals')
qqline(resid,col=2) ## plot shows distn. of ARIMA(3,1,1) residual differs from Normal distn.
anscombe.test(resid, alternative = "greater")  ## Anscombe-Glynn test of kurtosis for normal samples
anscombe.test(resid, alternative = "less")
kur=kurtosis(resid)
nu=round((6/kur)+4)
nu   ## parameter of 't'-distn ( to be fitted in ARMA residual)
b=qt(.95, df=nu, ncp=0, lower.tail = TRUE, log.p = FALSE)  ## upper 95% value of t(nu=6) distn.
low=f2-b*sqrt(ht)  ## lower conf. limit
high=f2+b*sqrt(ht) ## upper conf. limit
plot(y,type="l",main='Log yahoo  with C.I.',xlab="time",ylab="log-yahoo",col=3)
lines(low,col=2)
lines(high,col=4)
legend(locator(1), legend =c("Log yahoo","Lower Conf. limit","Upper Conf. limit"),col=c(3,2,4),
pch=15)
 ##---------- FORECAST----------------##
```

```
f3=forecast(f1,h=10)  ## mean forecast by ARIMA(3,1,1)   ## pr(say)
p1=predict(fit, n.ahead = 10,plot=T) ## std. devtn forecast by GARCH(1,1)
pr=c(3.400090,3.399856,3.399360,3.399041,3.398836,3.398705,3.398620,3.398566,3.398531,3.39850
9 )
pred=exp(pr)    ## predicted log-yahoo
lcll=pr-b*p1[,3] ## predicted lower conf limit
ucl=pr+b*p1[,3] ## predicted upper conf limit
lcll=exp(lcll)
ucl=exp(ucl)
```

❖ <u>code for Yahoo time series analysis of Japan</u>:

```
z=read.csv(file="D:/My Documents/Project_ Yahoo/a_yahoo japan _final.csv",header = T,sep=",")
cl=z$Close
plot(cl,main="Yahoo Closing share price in Japan",xlab="time points",ylab="yahoo share
price",type="l",col=3)
lcl=log(cl)
plot(lcl,main="Log Yahoo share price in Japan",xlab="time points",ylab="log yahoo share
price",type="l",col=4)
y=lcl    ## study variable
acf(y,lag.max=1000,main="ACF of log yahoo share price")
pacf(y,lag.max=1000,main="PACF of log yahoo share price")
 ##------------ check of trend ------------##
library(Kendall)
MannKendall(y)     ## monotonic trend present
d1=diff(y)
plot(d1,main="1st order difference data",xlab="time pt.",ylab="1st order difference") ## outlier present
abline(h=0,col=2)  ## outlier present in the plot .hence we opt for piece-wise regression to fit the trend.
g1=cbind(y[1:389],z$time.pt[1:389])
g2=cbind(y[390:937],z$time.pt[390:937])
g3=cbind(y[938:1515],z$time.pt[938:1515])
lm1=lm(g1[,1]~g1[,2])
lm2=lm(g2[,1]~g2[,2]+I(g2[,2]^2))
```

```
lm3=lm(g3[,1]~g3[,2]+I(g3[,2]^2))
l=c(predict(lm1),predict(lm2),predict(lm3))
plot(y,type="l",main="Observed value with fitted piece-wise trend",xlab="time pt.",ylab="log-
yahoo",col=3)
lines(l,col=2)
legend(locator(1), legend =c("Observed value","Fitted value"),col=c(3,2), pch=15)
d2=y-l  ## detrended value
plot(d2,main="Detrend through piece-wise regression",xlab="time pt.",ylab="detrend value")
abline(h=0,col=2)
MannKendall(d2)  ## monotonic trend absent
 ##------------ check of seasonality ---------------##
k1=c(d2[1:5])
k2=c(d2[6:10])
k3=c(d2[11:15])
k4=c(d2[16:20])
k5=c(d2[21:25])
k6=c(d2[26:30])
k=cbind(k1,k2,k3,k4,k5,k6)
matplot(k,type="l",main="Week wise plot of detrended data",ylab="log share price",xlab="time
pt.",col=c(1,2,3,4,5,6))
legend(locator(1),col=c(1,2,3,4,5,6),legend=c("1st week","2nd week","3rd week","4th week","5th
week","6th week"),pch=15)
jan=c(d2[1:8],d2[242:261],d2[494:513],d2[744:764],d2[996:1016],d2[1248:1267],d2[1500:1515])
feb=c(d2[9:27],d2[262:280],d2[514:533],d2[765:783],d2[1017:1035],d2[1268:1286])
mar=c(d2[28:50],d2[281:303],d2[534:554],d2[784:803],d2[1036:1056],d2[1287:1308])
apr=c(d2[51:71],d2[304:323],d2[555:575],d2[804:825],d2[1057:1077],d2[1309:1329])
may=c(d2[72:91],d2[324:344],d2[576:597],d2[826:847],d2[1078:1098],d2[1330:1349])
june=c(d2[92:113],d2[345:366],d2[598:618],d2[848:867],d2[1099:1118],d2[1350:1371])
july=c(d2[114:134],d2[367:386],d2[619:639],d2[868:889],d2[1119:1141],d2[1372:1393])
aug=c(d2[135:156],d2[387:409],d2[640:662],d2[890:911],d2[1142:1162],d2[1394:1414])
sept=c(d2[157:177],d2[410:430],d2[663:681],d2[912:931],d2[1163:1184],d2[1415:1435])
oct=c(d2[178:198],d2[431:451],d2[682:702],d2[932:954],d2[1185:1206],d2[1436:1457])
nov=c(d2[199:219],d2[452:472],d2[703:723],d2[955:974],d2[1207:1225],d2[1458:1477])
dec=c(d2[220:241],d2[473:493],d2[724:743],d2[975:995],d2[1226:1247],d2[1478:1499])
boxplot(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec,main="Month wise boxplot on detrend
data",xlab="month")
kruskal.test(list(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec))
##   removing seasonality   ##
s1=mean(jan)
```

```
s2=mean(feb)
s3=mean(mar)
s4=mean(apr)
s5=mean(may)
s6=mean(june)
s7=mean(july)
s8=mean(aug)
s9=mean(sept)
s10=mean(oct)
s11=mean(nov)
#s12=mean(dec)
s12=0-sum(s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11)
detrend=d2
d2=rep(0,1515)
d2[1:8]=s1
d2[242:261]=s1
d2[494:513]=s1
d2[744:764]=s1
d2[996:1016]=s1
d2[1248:1267]=s1
d2[1500:1515]=s1
d2[9:27]=s2
d2[262:280]=s2
d2[514:533]=s2
d2[765:783]=s2
d2[1017:1035]=s2
d2[1268:1286]=s2
d2[28:50]=s3
d2[281:303]=s3
d2[534:554]=s3
d2[784:803]=s3
d2[1036:1056]=s3
d2[1287:1308]=s3
d2[51:71]=s4
d2[304:323]=s4
d2[555:575]=s4
d2[804:825]=s4
d2[1057:1077]=s4
d2[1309:1329]=s4
```

```
d2[72:91]=s5
d2[324:344]=s5
d2[576:597]=s5
d2[826:847]=s5
d2[1078:1098]=s5
d2[1330:1349]=s5
d2[92:113]=s6
d2[345:366]=s6
d2[598:618]=s6
d2[848:867]=s6
d2[1099:1118]=s6
d2[1350:1371]=s6
d2[114:134]=s7
d2[367:386]=s7
d2[619:639]=s7
d2[868:889]=s7
d2[1119:1141]=s7
d2[1372:1393]=s7
d2[135:156]=s8
d2[387:409]=s8
d2[640:662]=s8
d2[890:911]=s8
d2[1142:1162]=s8
d2[1394:1414]=s8
d2[157:177]=s9
d2[410:430]=s9
d2[663:681]=s9
d2[912:931]=s9
d2[1163:1184]=s9
d2[1415:1435]=s9
d2[178:198]=s10
d2[431:451]=s10
d2[682:702]=s10
d2[932:954]=s10
d2[1185:1206]=s10
d2[1436:1457]=s10
d2[199:219]=s11
d2[452:472]=s11
d2[703:723]=s11
```

```
d2[955:974]=s11
d2[1207:1225]=s11
d2[1458:1477]=s11
d2[220:241]=s12
d2[473:493]=s12
d2[724:743]=s12
d2[975:995]=s12
d2[1226:1247]=s12
d2[1478:1499]=s12
seas_ind=d2
deseason=detrend-seas_ind
plot(deseason,main="Detrended,deseasonalised data",xlab="time",ylab="detrend deseasonalised")
abline(h=0,col=2)
d2=deseason
d2  ## detrended, deseasonalised data

jan=c(d2[1:8],d2[242:261],d2[494:513],d2[744:764],d2[996:1016],d2[1248:1267],d2[1500:1515])
feb=c(d2[9:27],d2[262:280],d2[514:533],d2[765:783],d2[1017:1035],d2[1268:1286])
mar=c(d2[28:50],d2[281:303],d2[534:554],d2[784:803],d2[1036:1056],d2[1287:1308])
apr=c(d2[51:71],d2[304:323],d2[555:575],d2[804:825],d2[1057:1077],d2[1309:1329])
may=c(d2[72:91],d2[324:344],d2[576:597],d2[826:847],d2[1078:1098],d2[1330:1349])
june=c(d2[92:113],d2[345:366],d2[598:618],d2[848:867],d2[1099:1118],d2[1350:1371])
july=c(d2[114:134],d2[367:386],d2[619:639],d2[868:889],d2[1119:1141],d2[1372:1393])
aug=c(d2[135:156],d2[387:409],d2[640:662],d2[890:911],d2[1142:1162],d2[1394:1414])
sept=c(d2[157:177],d2[410:430],d2[663:681],d2[912:931],d2[1163:1184],d2[1415:1435])
oct=c(d2[178:198],d2[431:451],d2[682:702],d2[932:954],d2[1185:1206],d2[1436:1457])
nov=c(d2[199:219],d2[452:472],d2[703:723],d2[955:974],d2[1207:1225],d2[1458:1477])
dec=c(d2[220:241],d2[473:493],d2[724:743],d2[975:995],d2[1226:1247],d2[1478:1499])
boxplot(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec,main="Month wise boxplot on detrend
data",xlab="month")
kruskal.test(list(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec))
boxplot(jan,feb,mar,apr,may,june,july,aug,sept,oct,nov,dec,main="Boxplot on detrended &
deseasonalised data",xlab="month")
acf(d2,lag.max=1000,main="ACF of stationary data")
pacf(d2,lag.max=1000,main="PACF of stationary data")
 ##---------------- MEAN stationary---------------##
library(forecast)
for(i in 0:4)
{ for(k in 0:4)
```

```
{ b2=arima(d2,order=c(i,0,k))
print(i)
print(k)
print(AIC(b2))
}}   ## on min AIC, ARIMA(2,0,4) is selected.
f1=arima(d2,order=c(2,0,2))
f2=fitted(f1)  ##predicted value
g=cbind(d2,f2)
matplot(g,type="l",col=c(3,2),main="Stationary & fitted ARMA(2,2)",ylab="stationary",xlab="time
pts")
legend(locator(1), legend =c("Stationary value","Fitted value"),col=c(3,2), pch=15)
resid=d2-f2  ## residual of ARMA(2,2)
plot(resid,main="Residual in ARMA(2,2)",ylab="residual")
acf(resid,main="ACF of residual of ARMA(2,2)",lag.max=1000)
pacf(resid,main="PACF of residual of ARMA(2,2)",lag.max=1000)
library(portes)
portest(resid,test="LjungBox")  ## white noise is absent in ARMA(2,2) residual
    ##  test for arch/garch   ##
library(FinTS)
ArchTest(resid,lags=1)
 ##------------ VARIANCE stationary --------------##
library(tseries)
for(i in 0:3)
{ for(j in 1:3)
{ ff=garch(resid,order=c(i,j),mse="uncond",trace=FALSE)
print(i)
print(j)
print(AIC(ff))
}}  ## hence on basis of min aic garch(1,1)
f=garch(resid,order=c(1,1),mse="uncond",trace=FALSE)   #garch(0,2)=arch(2)
summary(f)
library(fGarch)
fit=garchFit(~ garch(1,1),data=resid)
summary(fit)
 ##---------- Generate plot of Log Price, 95% Upper and Lower limit -------##
f2=fitted.values(f1)  ## fitted mean by ARMA(2,2)
ht=f$fit[,1]^2  ## fitted conditional standard deviation by GARCH(1,1)
plot(ht,main='Conditional variance fitted by by GARCH(1,1)',ylab="Conditional variance")
library(moments)
```

```
kurtosis(resid)
qqnorm(resid,main='ARMA(2,2) Residuals')
qqline(resid,col=2)  ## plot shows distn. of ARMA(2,1) residual differs from Normal distn.
anscombe.test(resid, alternative = "greater")  ## Anscombe-Glynn test of kurtosis for normal samples
anscombe.test(resid, alternative = "less")

kur=kurtosis(resid)
nu=round((6/kur)+4)
nu   ## parameter of 't'-distn ( to be fitted in ARMA residual)
b=qt(.95, df=nu, ncp=0, lower.tail = TRUE, log.p = FALSE)  ## upper 95% value of t(nu=6) distn.
low=f2-b*sqrt(ht)  ## lower conf. limit
high=f2+b*sqrt(ht) ## upper conf. limit
plot(d2,type="l",main='Log yahoo stationary with C.I.',xlab="time",ylab="stationary",ylim=c(-
.25,.25),col=3)
lines(low,col=2)
lines(high,col=4)
legend(locator(1), legend =c("Log yahoo stationary","Lower Conf. limit","Upper Conf.
limit"),col=c(3,2,4), pch=15)
plot(y,type="l",main='Log yahoo with C.I.',xlab="time",ylab="log-yahoo",col=3)
lines(low+l+seas_ind,col=2)
lines(high+l+seas_ind,col=4)
legend(locator(1), legend =c("Log yahoo","Lower Conf. limit","Upper Conf. limit"),col=c(3,2,4),
pch=15)
 ##---------------- FORECAST----------------##
p1=predict(fit, n.ahead = 10,plot=T) ## std. devtn forecast by GARCH(1,1)
f3=forecast(f1,h=10)  ## mean forecast by ARMA(2,2)   ## pr_1(say)
pr_1=c(-0.09902606,-0.09726107,-0.09003538,-0.08862722,-0.08185825,-0.08077056,-0.07442120,-
0.07362101,-0.06765739,-0.06711477) #  predicted stationary
w=coef(lm3)
seq=c(1516:1525)
pr_2=w[1]+w[2]*seq+w[3]*seq^2  ## predicted trend
pr_3=c(rep(s1,3),rep(s2,7)) ## seasonal indices accdn time pt.
pr_4=pr_1+pr_2+pr_3
pred=exp(pr_4)    ## predicted log-yahoo
lcll=pr_4-b*p1[,3] ## predicted lower conf limit
ucl=pr_4+b*p1[,3] ## predicted upper conf limit
lcll=exp(lcll)
ucl=exp(ucl)
```

❖ code for Regression Analysis:

```
z_u=read.csv(file="D:/My Documents/Project_ Yahoo/a_yahoo usa _final.csv",header = T,sep=",")
cl_u=z_u$Close
lcl_u=log(cl_u)
y_u=lcl_u   ## y_usa
z_j=read.csv(file="D:/My Documents/Project_ Yahoo/a_yahoo japan _final.csv",header = T,sep=",")
cl_j=z_j$Close
lcl_j=log(cl_j)
y_j=lcl_j    ## y_jap
y1=cbind(y_j,y_u)
zx=read.csv(file="D:/My Documents/Project_ Yahoo/exogeneous.csv",header = T,sep=",")
nas=zx$nasdaq    ## exogeneous variables
nik=zx$nikkei
exc=zx$exchange
mis=cbind(nas,nik,exc)  ## MICE algo to estimate missing values in exchange ratio.
library(mice)
mc=mice(mis)
mce=complete(mc)
exc1=mce[,3]
plot(nas,main="NASDAQ Index",xlab="time",ylab="nasdaq index",type="l",col=4)
plot(nik,main="NIKKEI Index",xlab="time",ylab="nikkei index",type="l",col=2)
plot(exc1,main="Yen-Dollar Exchange Rate",xlab="time",ylab="exchange rate",type="l",col=3)
log_nas=log(nas)
log_nik=log(nik)
log_exc=log(exc1)
d11=cbind(log_nas,log_nik,log_exc)
library(vars)    ## fit suitable VAR model and causality test
VARselect(cbind(y_u,y_j),exogen =d11, lag.max=10, type="both")$selection
vz1=VAR(cbind(y_u,y_j),p=1,type = "both", exogen =d11, lag.max =NULL,ic = "AIC")
summary(vz1)
causality(vz1,cause="y_j")
causality(vz1,cause="y_u")   ## y_j and y_u are instantaneous causal
   ## -------- Simultaneous equation ----------- ##
fit_y_j=predict(lm(y_j~log(nik)+log(nas)+log(exc1)))
fit_y_u=predict(lm(y_u~log(nik)+log(nas)+log(exc1)))
```

```
cbind(fit_y_j,y_j,fit_y_u,y_u)
l_j_1=lm(y_j~fit_y_u+log(nik)+log(exc1))
l_u_1=lm(y_u~fit_y_j+log(nas)+log(exc1))
c2=coef(l_j_1) #japan
c1=coef(l_u_1) #USA
f_y_j=predict(lm(y_j~fit_y_u+log(nik)+log(exc1)))
f_y_u=predict(lm(y_u~fit_y_j+log(nas)+log(exc1)))
matplot(cbind(exp(f_y_j),exp(y_j)),main="Observed and Predicted Yahoo share in
JAPAN",xlab="time",ylab="share price",col=c(4,2),type="l")
legend(locator(1),col=c(4,2),legend=c("observed share","predicted share"),pch=15)
matplot(cbind(exp(f_y_u),exp(y_u)),main="Observed and Predicted Yahoo share in
USA",xlab="time",ylab="share price",col=c(4,2),type="l")
legend(locator(1),col=c(4,2),legend=c("observed share","predicted share"),pch=15)
# structural form    # a*y=d+b*x , x=(log(nas),log(nik),log(exc))
a=matrix(c(1,-c1[2],-c2[2],1),nrow=2,byrow=T)
d=matrix(c(c1[1],c2[1]),nrow=2)
b=matrix(c(c1[3],0,c1[4],0,c2[3],c2[4]),nrow=2,byrow=T)
# reduced form   # y=d1+b1*x
d1=solve(a)%*%d
b1=solve(a)%*%b
for(j in 1:1515)     ## for cross check
{
w=d1+b1%*%d11[j,]
print(w)
}
   ##---- prediction--------##
data=read.csv(file="D:/My Documents/Project_ Yahoo/prediction.csv",header = T,sep=",")
data
x=cbind(log(data$nasdaq),log(data$nikkei),log(data$exchange))
x=t(x)
for(j in 1:10)
{
y=d1+b1%*%x[,j]
y=exp(y)
print(j)
print(y)
}
cbind(data$Japan,data$USA)
```