Assignment 1

MA4033 - Time Series and Stochastic Processes

Time Series and Stochastic Processes

Git Repo: https://github.com/sajeevan16/Time-Series-and-Stochastic-Processes

160544C - V. Sajeevan.

10th March 2020

Dataset 1

(a) Augmented Dickey−Fuller Test for Stationary:

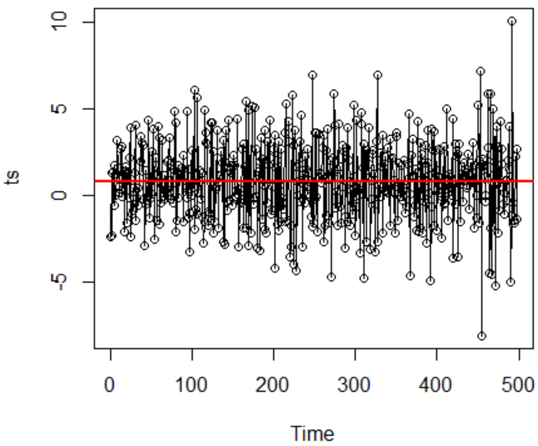
Alternative hypothesis: Stationary

Dickey-Fuller = -11.763, Lag order = 7, p-value = 0.01 (< 0.05)

There is no significant evidence for the first difference of the series to have a unit root under

95% confidence level. Therefore, we can assume that the first difference of the series is stationary.

(b) Visualizing the Time Series data:



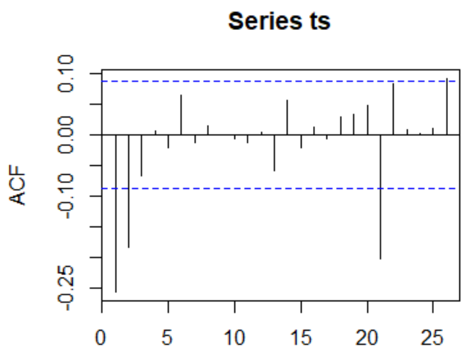
Time Series plot with mean line(red)

No visible trend with time.

Series has a non−zero mean.

Variability of the series remain constant.

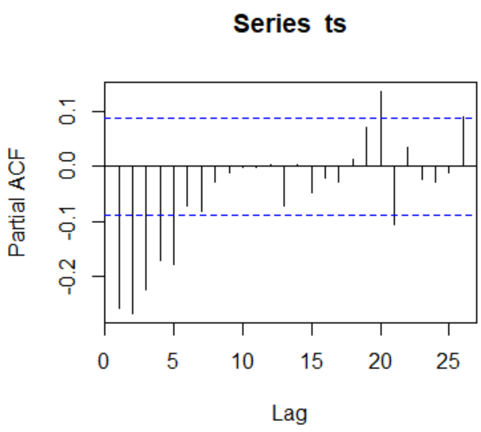
(c) Sample Auto−Correlation Function (ACF) of the first difference



Sample ACF Vs Time lags

Sample ACF is significant at lags 1,2,21 and 26 under simple standard error bound. We can neglect the 26th term due to random error. but cannot neglect 21^{st} term. ACF cuts of after lag 21. Therefore, ACF suggests that up to MA (21) seems to be a good model for the series.

(d) Sample Partial Auto−Correlation Function (PACF) of the first difference:



Sample PACF decreases gradually. PACF tails of with time lag. Therefore, PACF suggests that the model for the series contains MA terms rather than AR terms.

(e) Sample Extended Auto−Correlation Function (EACF) of the series:

```
> eacf(ts)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o o o o o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x x x o o o o o o o o  o  o  o
3 x x o o x o o o o o o  o  o  o
4 x x o x o o o o o o o  o  o  o
5 x x o x x o o o o o o  o  o  o
6 x x o x x x o o o o o  o  o  o
7 x o o o x o o o o o o  o  o  o
```

EACF suggest that MA (2) seems to be the simpler better model to represent the series.

(f) Model Selection:

ACF and EACF suggest that the more simpler model is MA(2) and PACF also supports this claim. Therefore, we will further analyse the MA(2) model.

(g) Model Analysis – MA(2) :

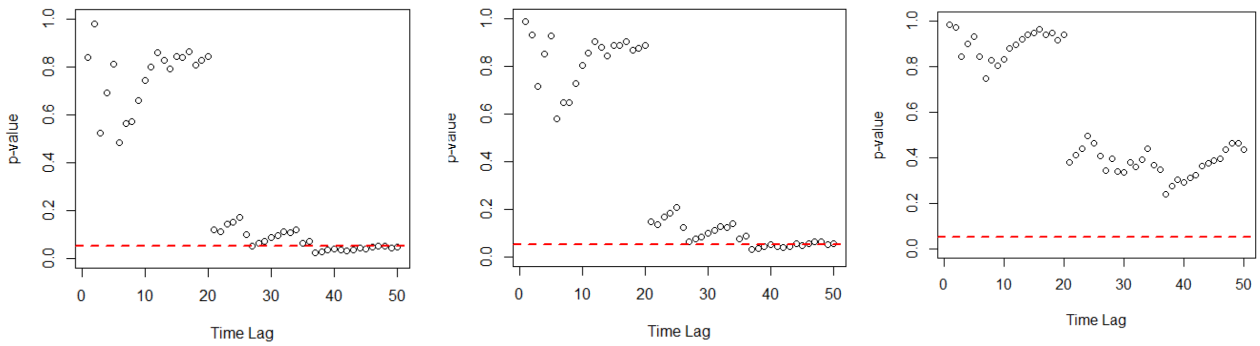
```
> model_fit

Call:
arima(x = ts, order = c(p, d, q))

Coefficients:
          ma1      ma2  intercept
      -0.4878  -0.2593     0.8390
s.e.   0.0420   0.0401     0.0234

sigma^2 estimated as 4.194:  log likelihood = -1063.97,  aic = 2133.93
> t_stat(model_fit)
              ma1       ma2 intercept
t.stat -11.62553 -6.462378  35.85954
p.val    0.00000  0.000000   0.00000
> AIC(model_fit)
[1] 2135.933
> BIC(model_fit)
[1] 2152.775
```

All of the coefficients of MA(2) model are significant under 95% confidence level.

Model Equation: $Y_t = e_t - 4878\ e_{t-1} - 0.2593e_{t-2}$

(h) Model Residual Analysis

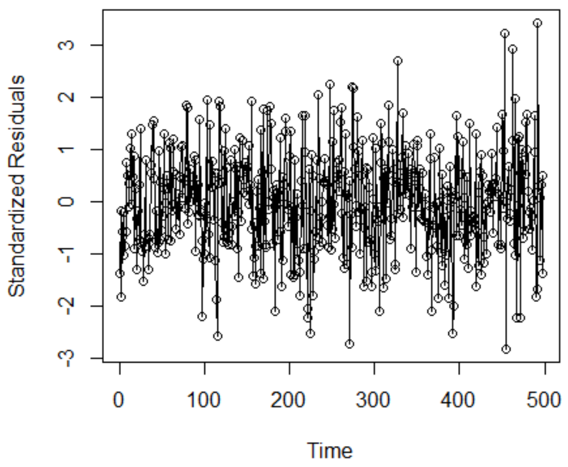
1. Modified Box−Pierce (Ljung−Box) Test MA(2) vs MA(3) vs ARIMA(1,3)



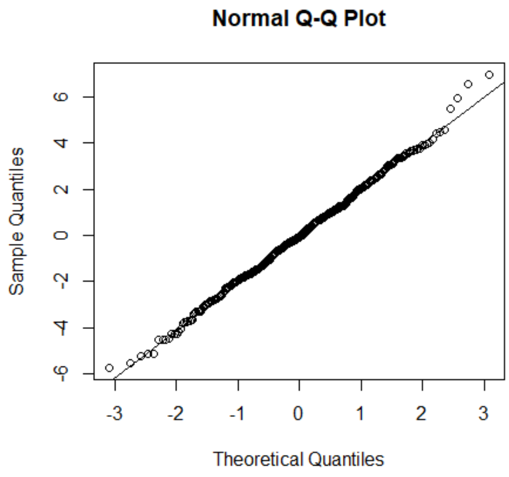
When considering the Modified Box−Pierce (Ljung−Box) Test for MA(2), MA(3), ARIMA(1,0,3), in the ARIMA (1,0,3) there is no sufficient evidence to reject the null hypothesis in different time lags under 95% confidence level. And P- values at different time lags are greater than 0.05.

Therefore, we will further analyze the ARMA(1,3) model.

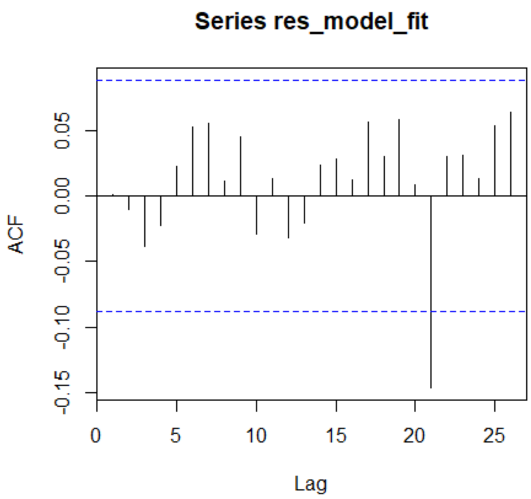
2. Plot of Residuals



3. Normality of Residuals



4. Autocorrelation of Residuals



No any significant ACF of residuals with time lag.

(h) Conclusion:

Model Equation: $Y_t = -0.9993\, Y_{t-1} + 0.8391 + e_t + 0.543e_{t-1} - 0.7426e_{t-2} - 0.2932e_{t-2}$

According to the above timeseries analysis, we can conclude that the ARMA(1,3) model is the smaller and better model for the given timeseries data.

Dataset 2

(a) Augmented Dickey−Fuller Test for Stationary:

Alternative hypothesis: Stationary

Dickey-Fuller = -3.4035, Lag order = 6, p-value = 0.05413 (> 0.05)

There is significant evidence for the series to have a unit root under 95% confidence level. Therefore, the series is not stationary.

Let's consider the first difference of the series:

Dickey-Fuller = -8.1346, Lag order = 6, p-value = 0.01 (< 0.05)

There is no significant evidence for the first difference of the series to have a unit root under

95% confidence level. Therefore, we can assume that the first difference of the series is stationary.

(b) Visualizing the Time Series data:

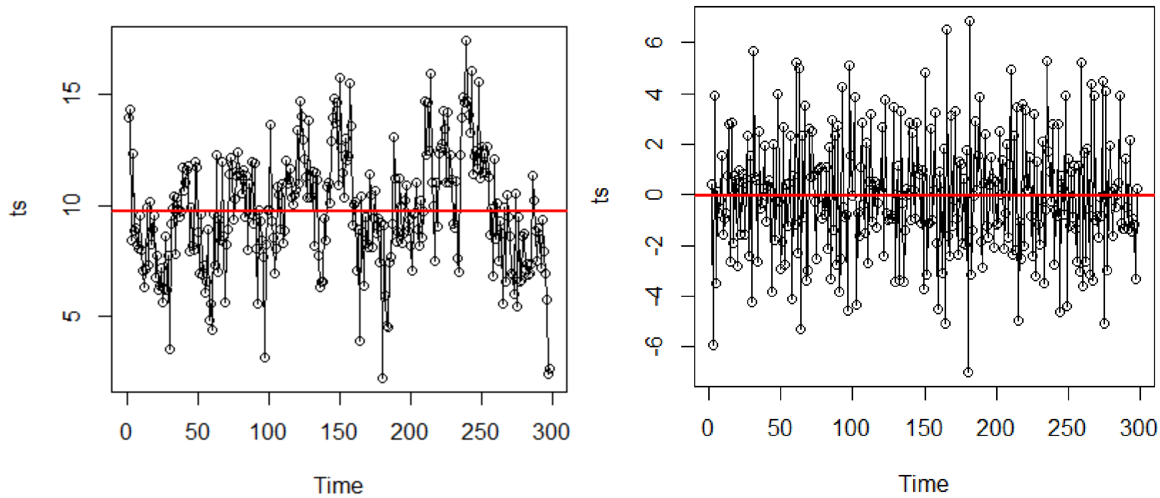


Figure 1: Time Series plot with mean line

Since the first difference of the series seems to be stationary, we will continue to analyze the first difference of the series further to fit an appropriate model.

(c) Sample Auto−Correlation Function (ACF) of the first difference:

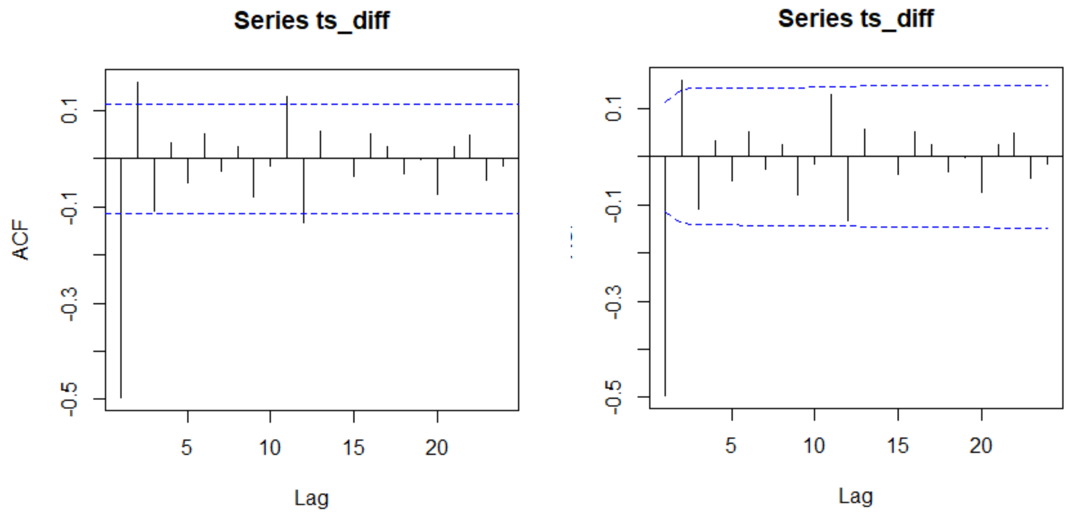


Figure 2: Sample ACF of first difference Vs Time lags

we can also neglect the ACF terms at lags 11 and 12 due to random error.

Therefore, ACF suggests that MA (2) seems to be a good model for the series.

(d) Sample Partial Auto−Correlation Function (PACF) of the first difference:

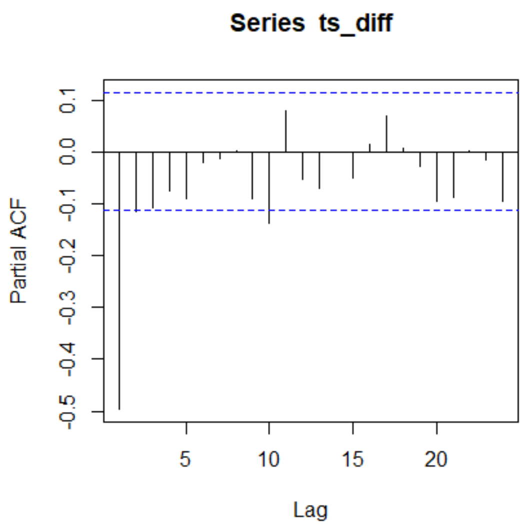


Figure 3: Sample PACF of first difference Vs Time lags

Sample PACF is significant at lags 1,2 and 10. Neglect lag 10 due to the random error. PACF is significant up to the 1st term and cuts after lag 1. Therefore, PACF suggests that AR (1) seems to be a good model for the series.

(e) Sample Extended Auto−Correlation Function (EACF) of the first difference:

```
> eacf(ts_diff)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o o o o o o x  x  o  o
1 x x o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x x o x o o o o o o o  o  o  o
4 x x o x o o o o o o o  o  o  o
5 x x o x o o o o o o o  o  o  o
6 x x o x x o o o o o o  o  o  o
7 x o o o x o o o o o o  o  o  o
> 
```

EACF suggest that ARMA (1,2) seems to be the better model to represent the first difference of the series.

(f) Model Selection:

ACF suggest IMA (1, 2) model whereas PACF suggest ARI(1, 1) model for the series. But according to EACF, both IMA (1, 2) and ARI (1, 1) are not significant. And EACF suggests that ARIMA (1, 1, 2) is the better simpler model.

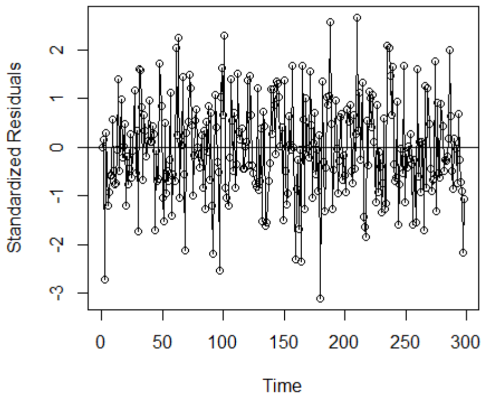
(g) Model Analysis – ARIMA(1,1,2) :

```
> model_fit

Call:
arima(x = ts, order = c(p, d, q))

Coefficients:
          ar1     ma1      ma2
      -0.7304  0.1328  -0.3251
s.e.   0.2052  0.2171   0.1501

sigma^2 estimated as 4.243:  log likelihood = -636.25,  aic = 1278.51
> t_stat(model_fit)
             ar1      ma1       ma2
t.stat -3.560249 0.611821 -2.165683
p.val   0.000371 0.540656  0.030335
> AIC (model_fit)
[1] 1280.507
> BIC(model_fit)
[1] 1295.282
> res_model_fit <- residuals(model_fit)
```

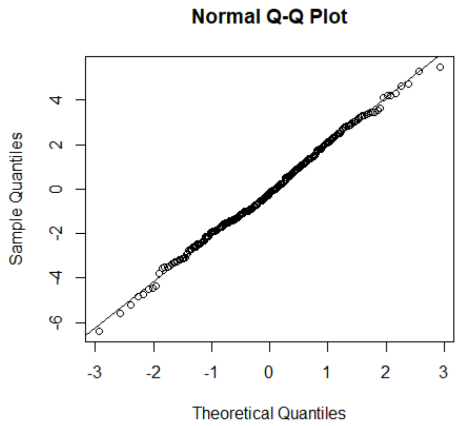
Model Equation: $W_t = Y_t - Y_{t-1} = -0.7304\,8\,(Y_{t-1} - Y_{t-2}) + e_t + 0.1328e_{t-1} - 0.3251e_{t-2}$

(h) Model Residual Analysis

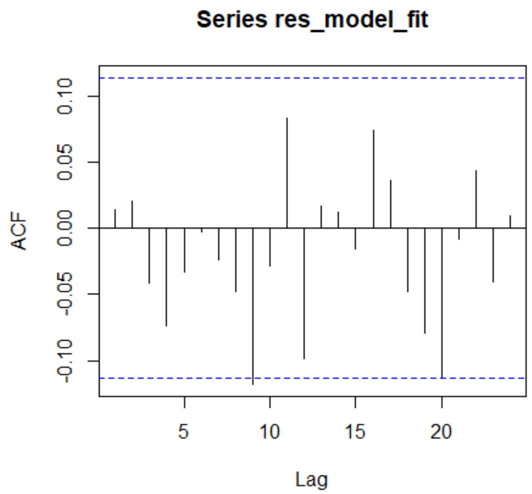
1. Plot of Residuals:



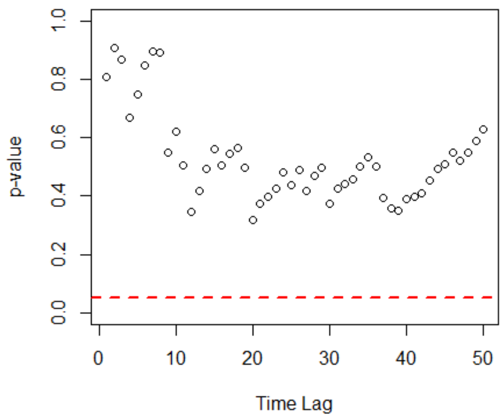
2. Normality of Residuals



3. Auto−Correlation of Residuals



4. Modified Box−Pierce (Ljung−Box) Test



(I) Conclusion

According to the above timeseries analysis, we can conclude that the ARIMA (1, 1, 2) model is the smaller and better model for the given timeseries data.

Dataset 3

(a) Augmented Dickey–Fuller Test for Stationary:

Alternative hypothesis: Stationary

Dickey-Fuller = -3.429, Lag order = 6, p-value = 0.04992 (<≈ 0.05)

It's very nearly equal to There is significant evidence for the series to have a unit root under 95% confidence level. Therefore, the series is not stationary.

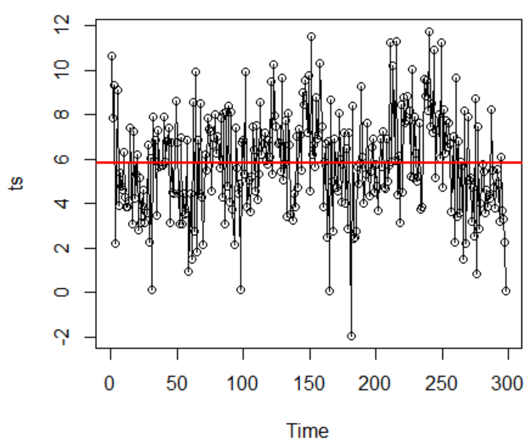
Let's consider the first difference of the series:

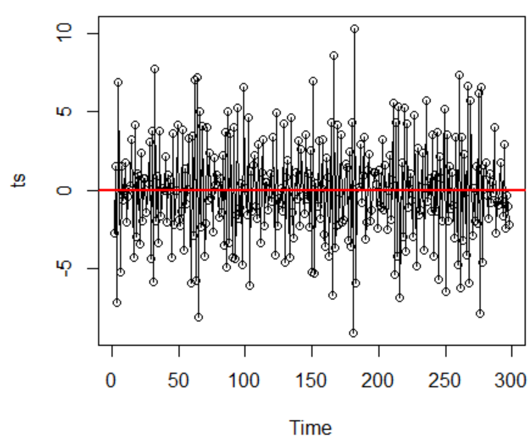
Dickey-Fuller = -9.2501, Lag order = 6, p-value = 0.01 (< 0.05)

There is no significant evidence for the first difference of the series to have a unit root under

95% confidence level. Therefore, we can assume that the first difference of the series is stationary.

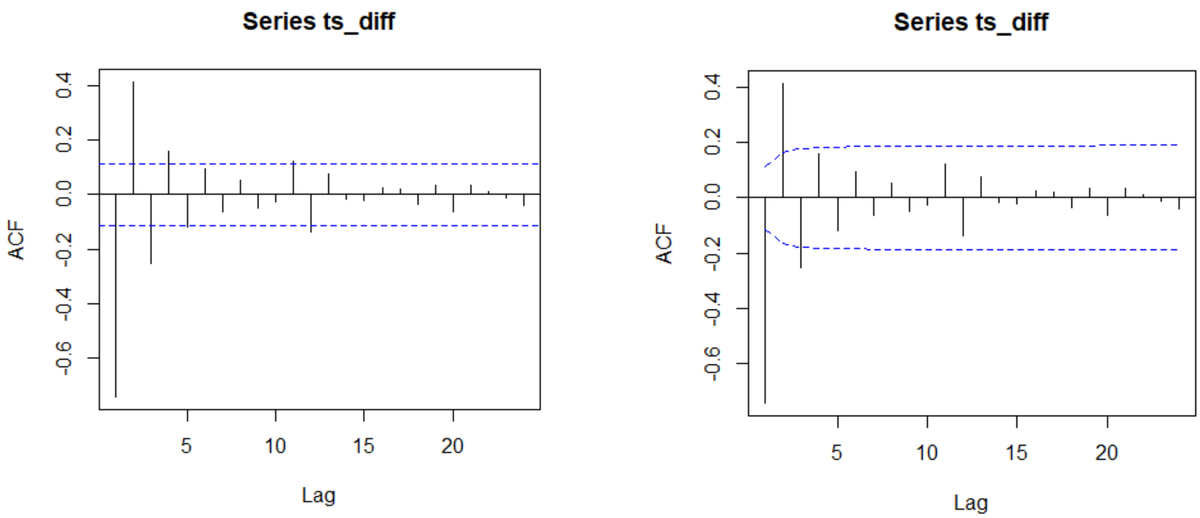
(b) Visualizing the Time Series data





In the first difference of the series No visible trend with time Variability of the series remain constant i.e., does not show any increasing or decreasing pattern.

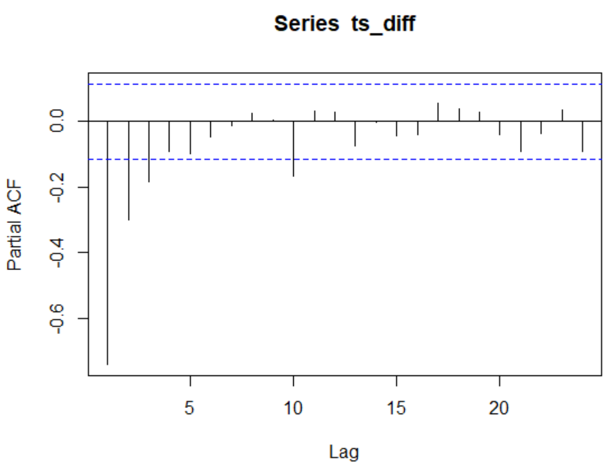
(c) Sample Auto−Correlation Function (ACF) of the first difference:



Sample ACF of first difference Vs Time lags

Sample ACF is significant at lags 1,2,3,4,5,11 and 12 under simple standard error bound. But under more sophisticated standard error bound, we can neglect the ACF terms at lags 11 and 12. Further we can also neglect the ACF terms at lags 4 and 5 due to random error. Only the first, second and third terms are significant and ACF cuts of after lag 3.

(d) Sample Partial Auto−Correlation Function (PACF) of the first difference:



Sample PACF is significant at lags 1,2,3 and 10. We can neglect the 10th term due to random error

PACF is significant up to 3rd term and cuts of after lag 3.

(e) Sample Extended Auto−Correlation Function (EACF) of the first difference:

```
> eacf(ts_diff)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x o o o o o x  x  o  o
1 x o o o o o o o o o x  x  o  o
2 x x o o o o o o o o o  o  o  o
3 x x x x o o o o o o o  o  o  o
4 x x x x o o o o o o o  o  o  o
5 x o x x o o o o o o o  o  o  o
6 x o x o x o o o x o o  o  o  o
7 x o o o x o o o x o o  o  o  o
>
```

EACF suggest that ARMA(1,2) seems to be the better model to represent the first difference of the series.

(f) Model Selection:

ACF suggest IMA (1, 3) model whereas PACF suggest ARI(1, 3) model for the series. But according to EACF, both IMA (1, 3) and ARI (1, 3) are not significant. And EACF suggests that ARIMA (1, 1, 2) is the better simpler model.

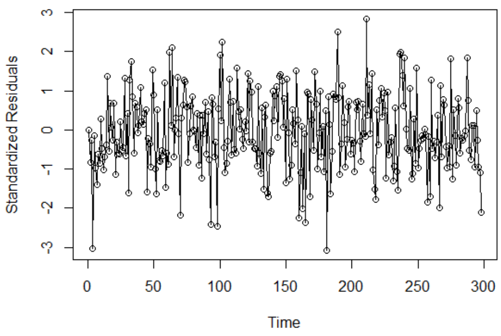
(g) Model Analysis – ARIMA(1,1,2)

```
> model_fit

Call:
arima(x = ts, order = c(p, d, q))

Coefficients:
          ar1      ma1      ma2
      -0.6651  -0.4334  -0.2173
s.e.   0.1013   0.1212   0.1229

sigma^2 estimated as 4.215:  log likelihood = -635.76,  aic = 1277.52
> t_stat(model_fit)
             ar1       ma1       ma2
t.stat -6.565105 -3.574763 -1.768484
p.val   0.000000  0.000351  0.076980
> AIC(model_fit)
[1] 1279.522
> BIC(model_fit)
[1] 1294.297
```

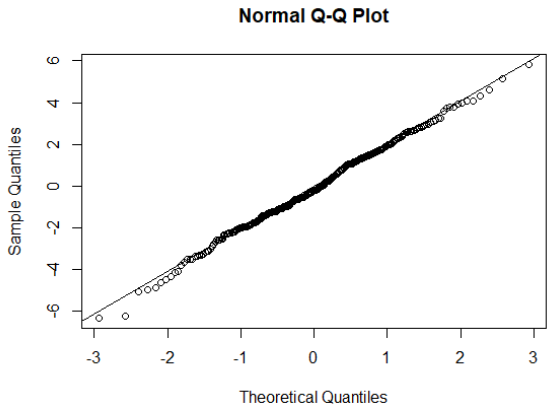
Model Equation: $W_t = Y_t - Y_{t-1} = -0.6651\,(\mathrm{Y}_{t-1} - \mathrm{Y}_{t-2}) + e_t - 0.1212e_{t-1} - 1229e_{t-2}$

(h) Model Residual Analysis

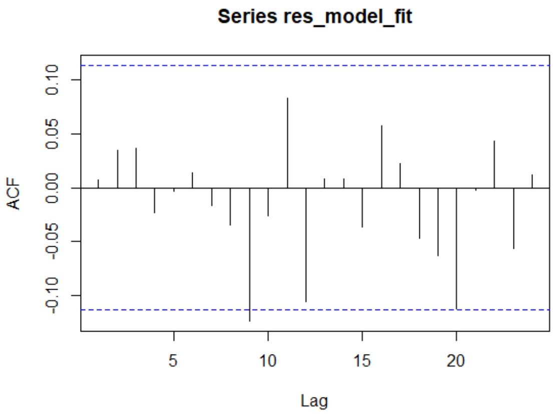
1. Plot of Residuals:



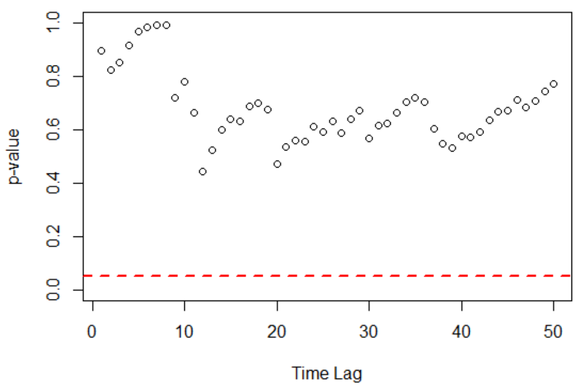
2. Normality of Residuals



3. Auto−Correlation of Residuals



4. Modified Box−Pierce (Ljung−Box) Test



(I) Conclusion

According to the above timeseries analysis, we can conclude that the ARIMA (1, 1, 2) model is the smaller and better model for the given timeseries data.

Summarized Results

Dataset 1

Model	ARMA(1,3)
Model Equation	$Y_t = -0.9993\, Y_{t-1} + 0.8391 + e_t + 0.543 e_{t-1} - 0.7426 e_{t-2} - 0.2932 e_{t-2}$
Log likelihood	-1058.64
BIC	2154.547

Dataset 2

Model	ARIMA (1, 1, 2)
Model Equation	$W_t = Y_t - Y_{t-1} = -0.7304\,8\, (\mathrm{Y}_{t-1} - \mathrm{Y}_{t-2}) + e_t + 0.1328 e_{t-1} - 0.3251 e_{t-2}$
Log likelihood	-636.25
BIC	1295.282

Dataset 3

Model	ARIMA (1, 1, 2)
Model Equation	$W_t = Y_t - Y_{t-1} = -0.6651\, (\mathrm{Y}_{t-1} - \mathrm{Y}_{t-2}) + e_t - 0.1212 e_{t-1} - 1229 e_{t-2}$
Log likelihood	-635.76
BIC	1294.297