Problem 1 (12 credits)

HW3

Michael Brucek (mbrucek), Eduardo Chavez (echavezh), Kevin Grady (grady133), Danny Moncada (monca016), Tian Zhang (zhan7003)

March 25, 2020

```
suppressWarnings(suppressPackageStartupMessages({
  library(TSA)
  library(forecast)
  library(ggplot2)
  library(dplyr)
  library(tseries) #For the ADF test only
}))
```

Identify given stochastic processes

Please load the data from file `problem1.Rds`

```
problem1 <- readRDS("problem1.Rds") # Please do not change this line
```

Please note that at least some of the stochastic processes here are such that **`auto.arima` fails to detect the right model** here.

You should use the tools that we studied in class to detect the correct model manually. **Please report all the plots that you found necessary as well as your reasoning for choosing the appropriate model.**

Extra note:

- Please note that one of these time series will have a seasonality with the period of 4. As you are solving each problem, you should be able to discover which one it is and please make sure to mark the seasonal part of ARIMA rather than non-seasonal part of ARIMA for that case.
- For the processes that don't have the seasonal component **please put the zeros rather than NAs** into the seasonal ARIMA part. You can leave the seasonal period as NA.

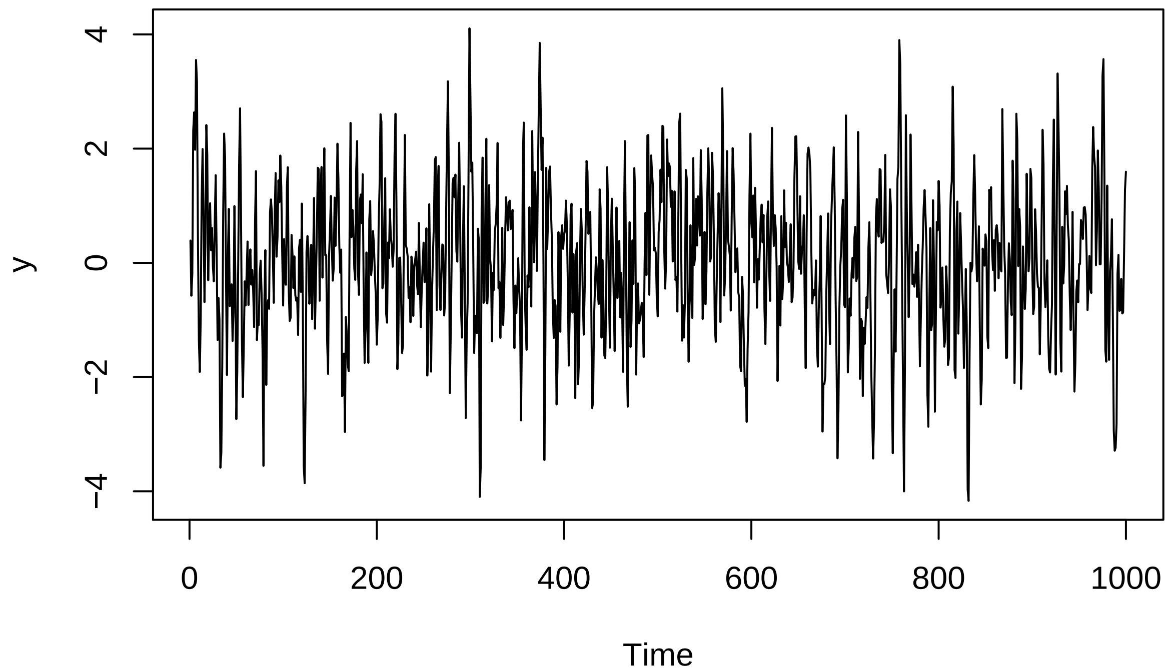
General Requirements

- Please do not change the path in `readRDS()`, your solutions will be automatically run by the bot and the bot will not have access to the folders that you have.
- Please review the resulting PDF and **make sure that all code fits into the page.** If you have lines of code that run outside of the page limits we will deduct points for incorrect formatting as it makes it unnecessarily hard to grade.
- If the true model is seasonal but you did not specify it as a seasonal model, this will be counted as an incorrect solution
- Please avoid using esoteric R packages. We have already discovered some that generate arima models incorrectly. Stick to tried and true packages: base R, `forecast`, `TSA`, `zoo`, `xts`.

Question 1 (3 credits)

Please look at the `problem1$sample_path1` and identify the ARIMA order of the underlying stochastic process

```
# Please do your analysis below
y = problem1$sample_path1
ts.plot(y)
```

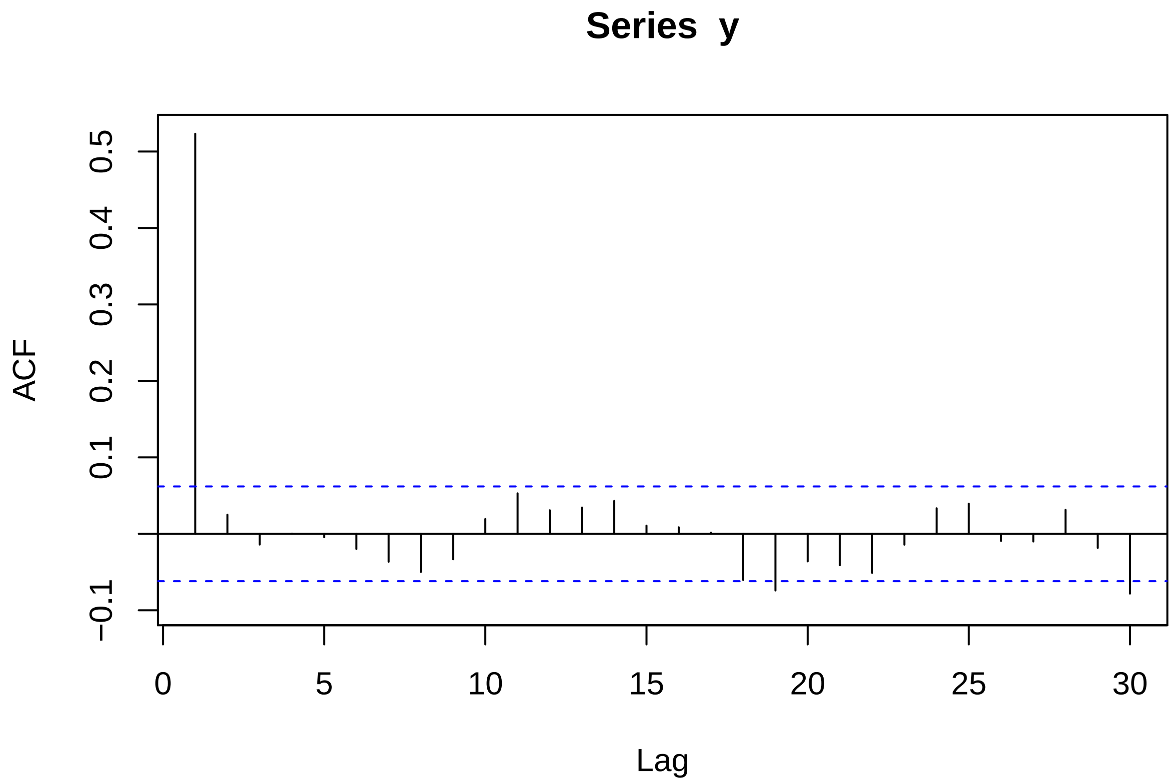



```
adf.test(y)
```

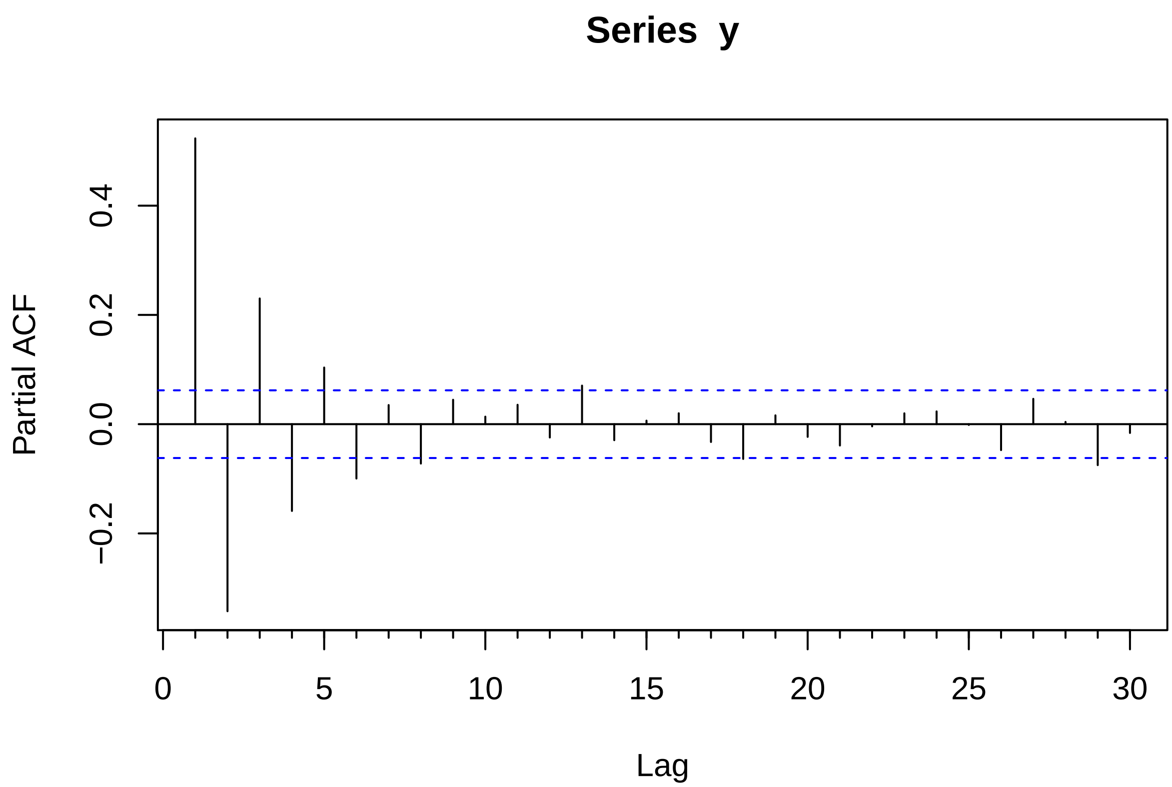
```
## Warning in adf.test(y): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -9.7278, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```
Acf(y)
```




```
Pacf(y)
```




```r
eacf(y)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o  o  o  o
## 1 x x 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 o o o o o o o  o  o  o
## 4 x x x o o o o o o o o  o  o  o
## 5 x x x x o o o o o o o  o  o  o
## 6 x x x x o x o o o o o  o  o  o
## 7 x x x x x o x o o o o  o  o  o
```

```r
auto.arima(y)
```

```
## Series: y
## ARIMA(0,0,2) with zero mean
##
## Coefficients:
##          ma1     ma2
##       0.8594  0.0741
## s.e.  0.0318  0.0324
##
## sigma^2 estimated as 0.9796:  log likelihood=-1408.14
## AIC=2822.29   AICc=2822.31   BIC=2837.01
```

```r
Arima(y,order=c(0,0,1))
```

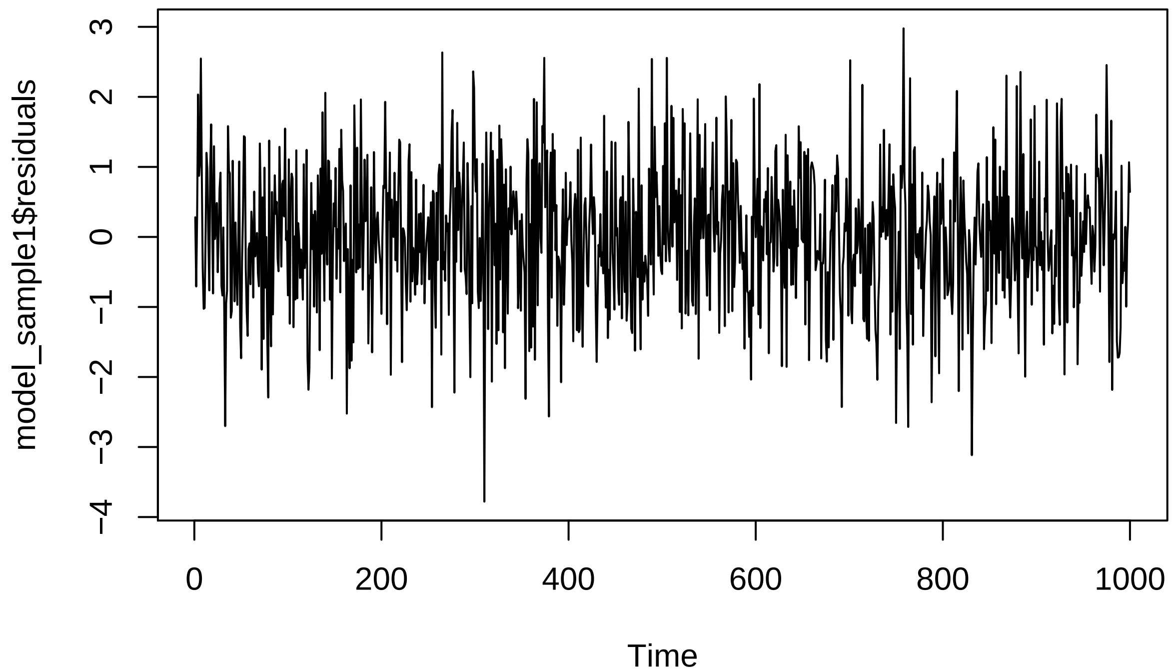
```
## Series: y
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1    mean
##       0.8012  0.0202
## s.e.  0.0180  0.0564
##
## sigma^2 estimated as 0.9846:  log likelihood=-1410.68
## AIC=2827.37   AICc=2827.39   BIC=2842.09
```

```r
Arima(y,order=c(0,0,2))
```

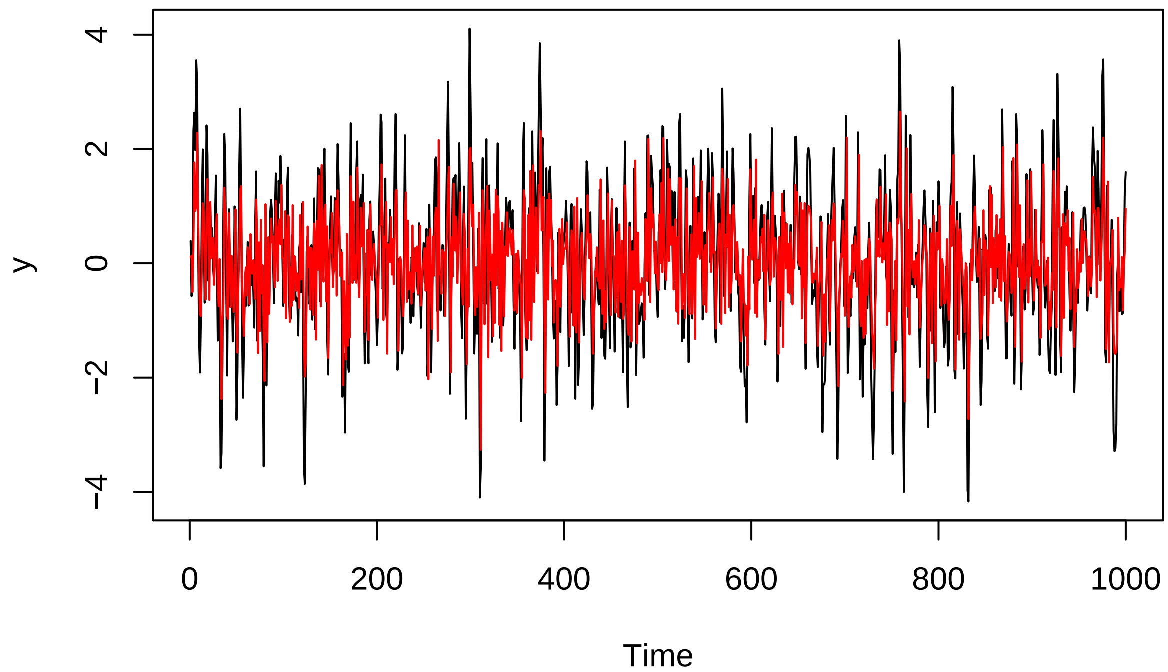
```
## Series: y
## ARIMA(0,0,2) with non-zero mean
##
## Coefficients:
##          ma1     ma2    mean
##       0.8593  0.0740  0.0204
## s.e.  0.0318  0.0324  0.0604
##
## sigma^2 estimated as 0.9805:  log likelihood=-1408.09
## AIC=2824.17   AICc=2824.21   BIC=2843.8
```

```r
model_sample1 = Arima(y, order=c(0,0,2))

plot(model_sample1$residuals)
```




```r
plot(y)
lines(y - model_sample1$residuals, col="red")
```




```r
# Please do your analysis above
```

Please describe your reasoning below:

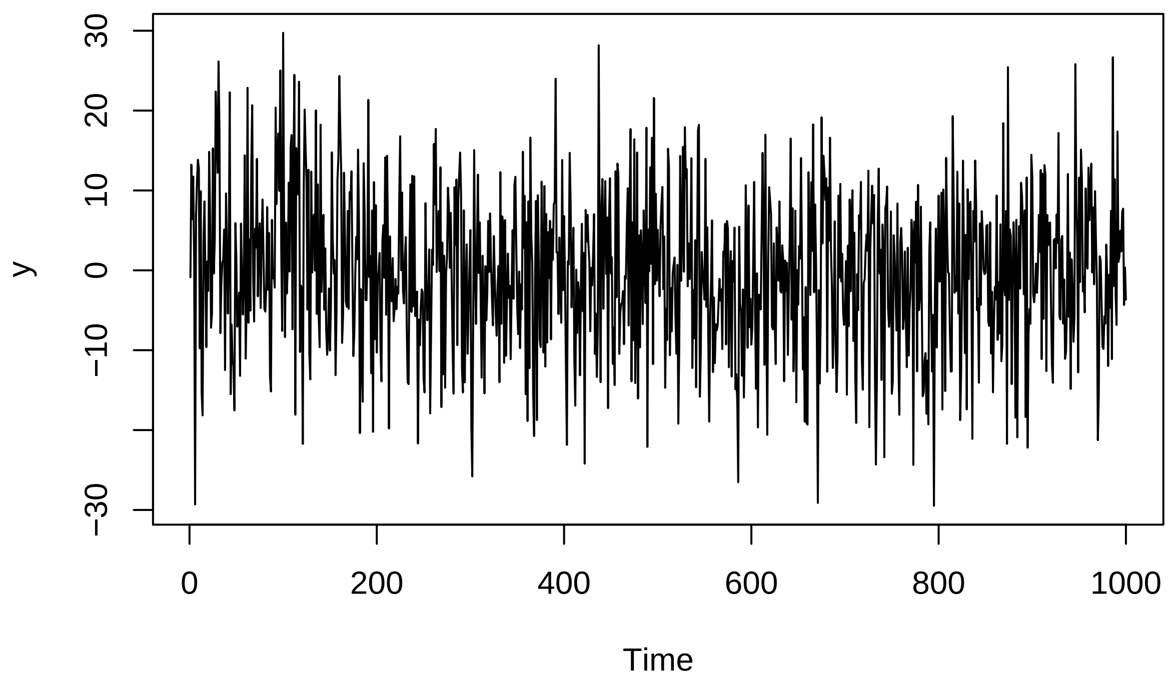
- The initial time series plot appears to have consistent mean of 0, and the ADF test confirms stationarity (p-value = 0.01).
- The ACF & EACF plots show a MA(1) trend.
- The auto.arima output of this sample path gives us a model of ARIMA(0,0,1).
- When we compare the AIC & BIC scores between our ARIMA(0,0,1) and ARIMA(0,0,2) models, we have a lower scores for the MA(2) model (AIC=2824.17).
- When we plot the residuals of the MA(2) model, they fit well with the first sample path.
- Our final conclusion is that this is an MA(2) model.

```r
# Please specify the estimated ARIMA(p,d,q) orders below
Q1_p <- 0
Q1_d <- 0
Q1_q <- 2
Q1_Sp <- NA # [Optional] Seasonal AR() order
Q1_Sd <- NA # [Optional] Seasonal I() order
Q1_Sq <- NA # [Optional] Seasonal MA() order
Q1_S <- NA # [Optional] seasonal period
```

Question 2 (3 credits)

Please look at the `problem1$sample_path2` and identify the ARIMA order of the underlying stochastic process

```r
# Please do your analysis below
y = problem1$sample_path2
ts.plot(y)
```

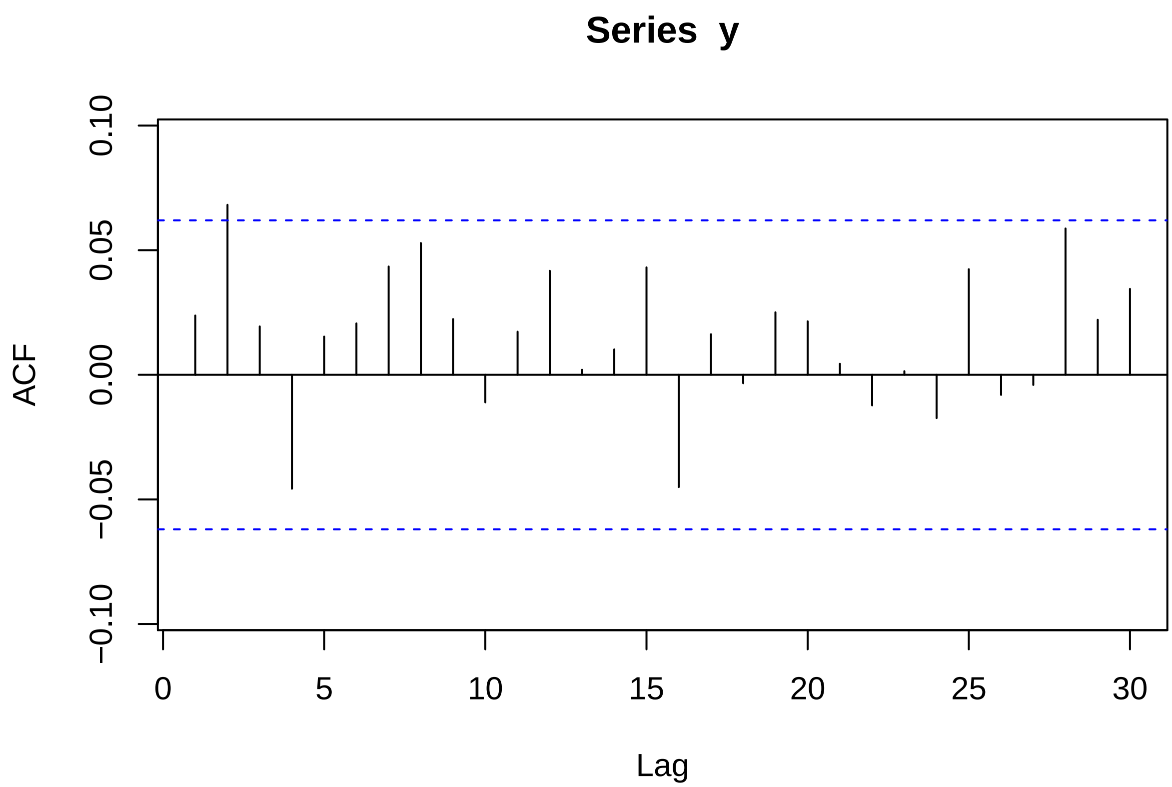



```r
adf.test(y)
```

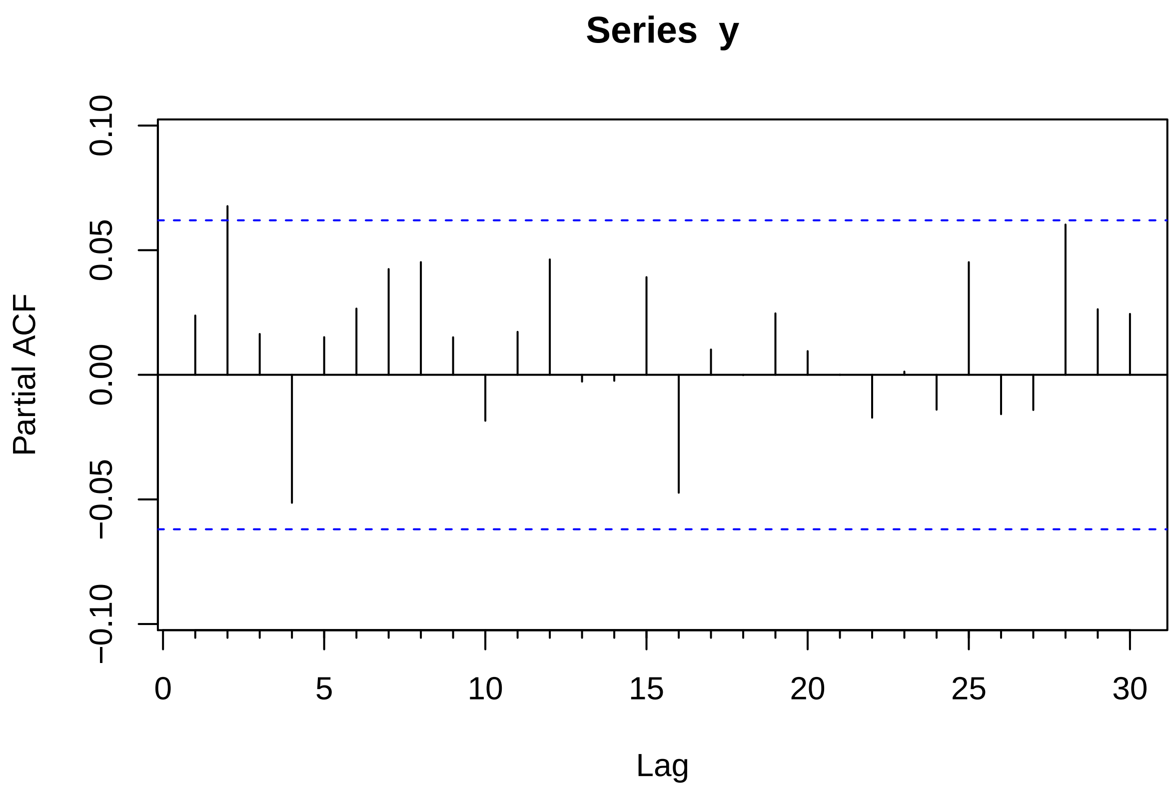
```
## Warning in adf.test(y): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -9.3444, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```r
Acf(y)
```




```
Pacf(y)
```



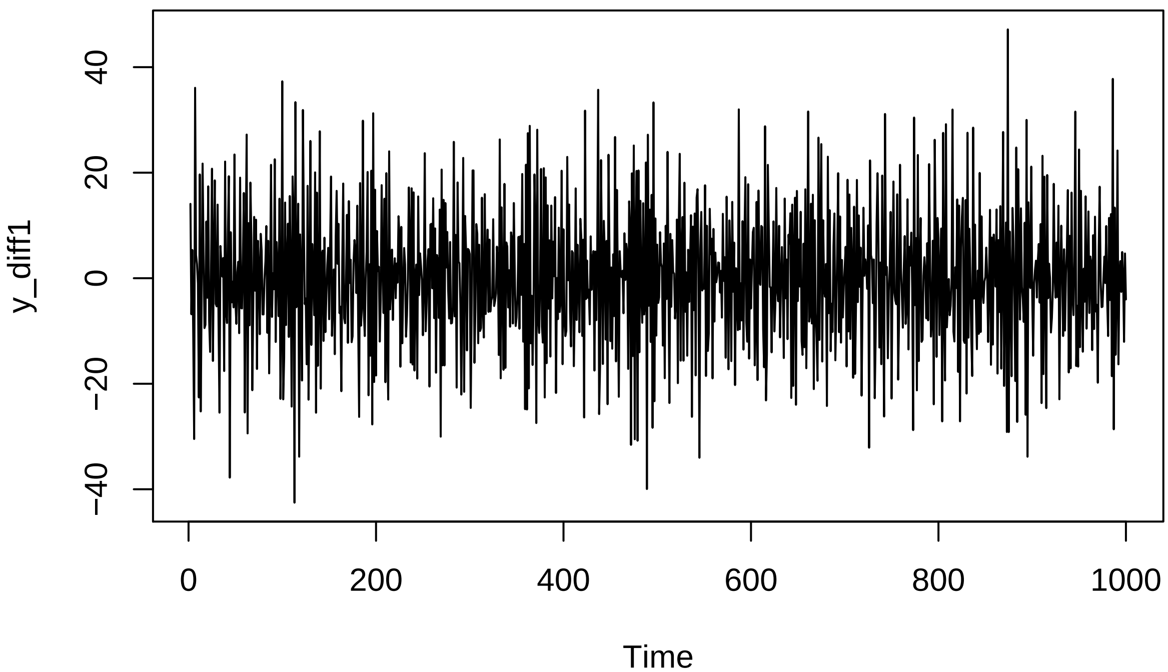

```r
eacf(y)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o 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 o o o o o o o o o  o  o  o
## 3 x x o o 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 o x x o o o o o o o  o  o  o
## 6 x x o x o x o o o o o  o  o  o
## 7 x x x o o x o o o o o  o  o  o
```

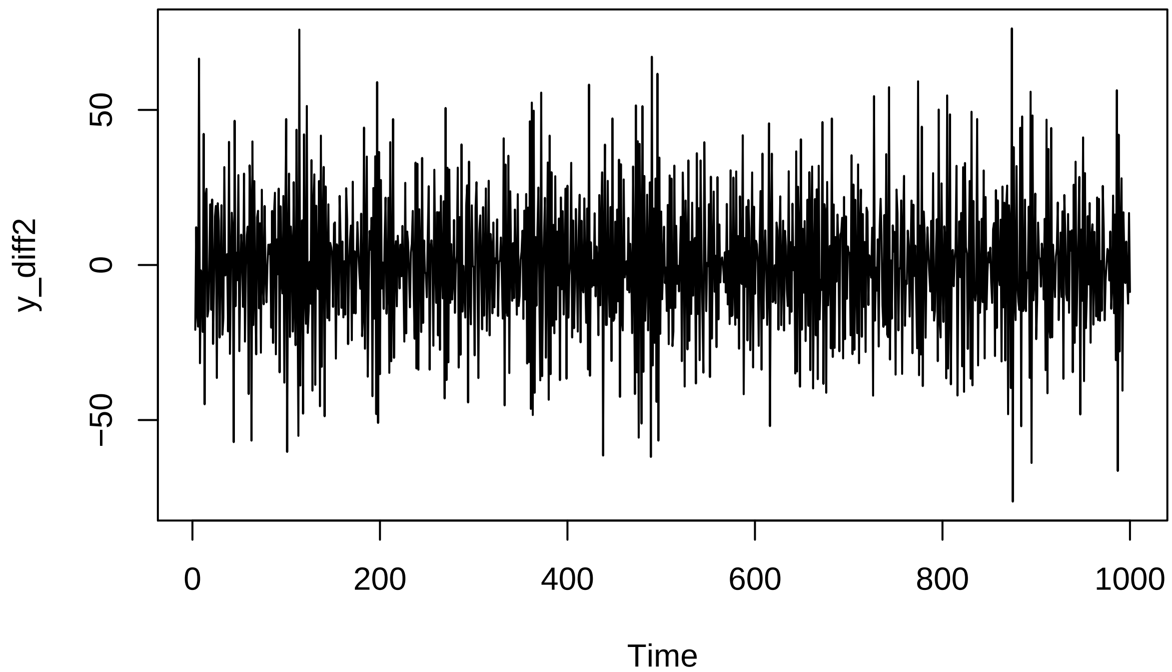
```r
auto.arima(y)
```

```
## Series: y
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.9866
## s.e.   0.0061
##
## sigma^2 estimated as 97.86:  log likelihood=-3708.29
## AIC=7420.59   AICc=7420.6   BIC=7430.4
```

```r
y_diff1=diff(y,differences = 1)
y_diff2=diff(y_diff1, differences = 1)
ts.plot(y_diff1)
```




```r
ts.plot(y_diff2)
```




```r
Arima(y,order=c(0,1,1))
```

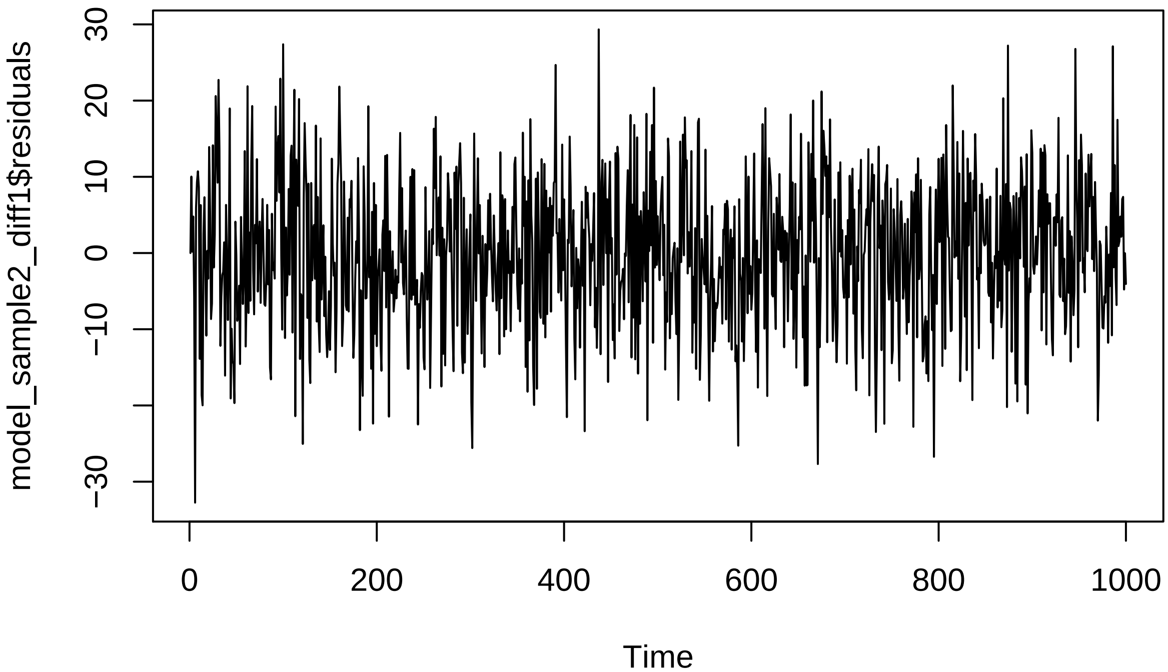
```
## Series: y
## ARIMA(0,1,1)
##
## Coefficients:
##           ma1
##       -0.9866
## s.e.   0.0061
##
## sigma^2 estimated as 97.86:  log likelihood=-3708.29
## AIC=7420.59   AICc=7420.6   BIC=7430.4
```

```r
Arima(y,order=c(0,0,1))
```

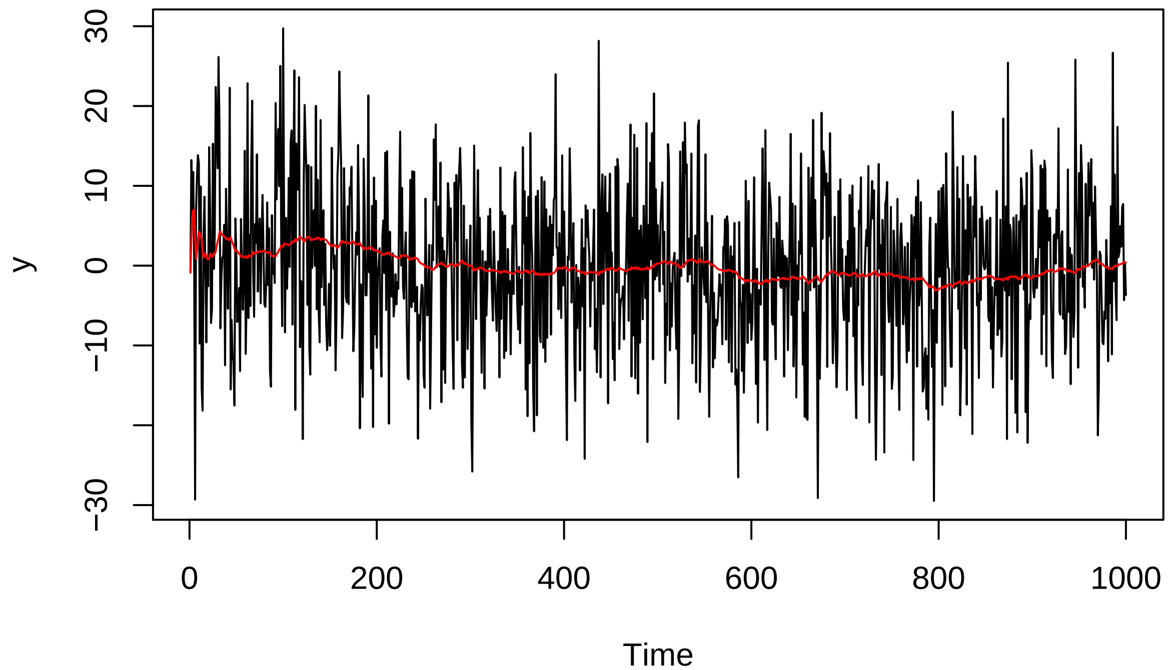
```
## Series: y
## ARIMA(0,0,1) with non-zero mean
##
## Coefficients:
##          ma1     mean
##       0.0209  -0.2626
## s.e.  0.0297   0.3200
##
## sigma^2 estimated as 98.47:  log likelihood=-3712.8
## AIC=7431.61   AICc=7431.63   BIC=7446.33
```

```r
model_sample2_diff1 = Arima(y,order=c(0,1,1))

plot(model_sample2_diff1$residuals)
```




```r
plot(y)
lines(y - model_sample2_diff1$residuals, col="red")
```



Please describe your reasoning below:

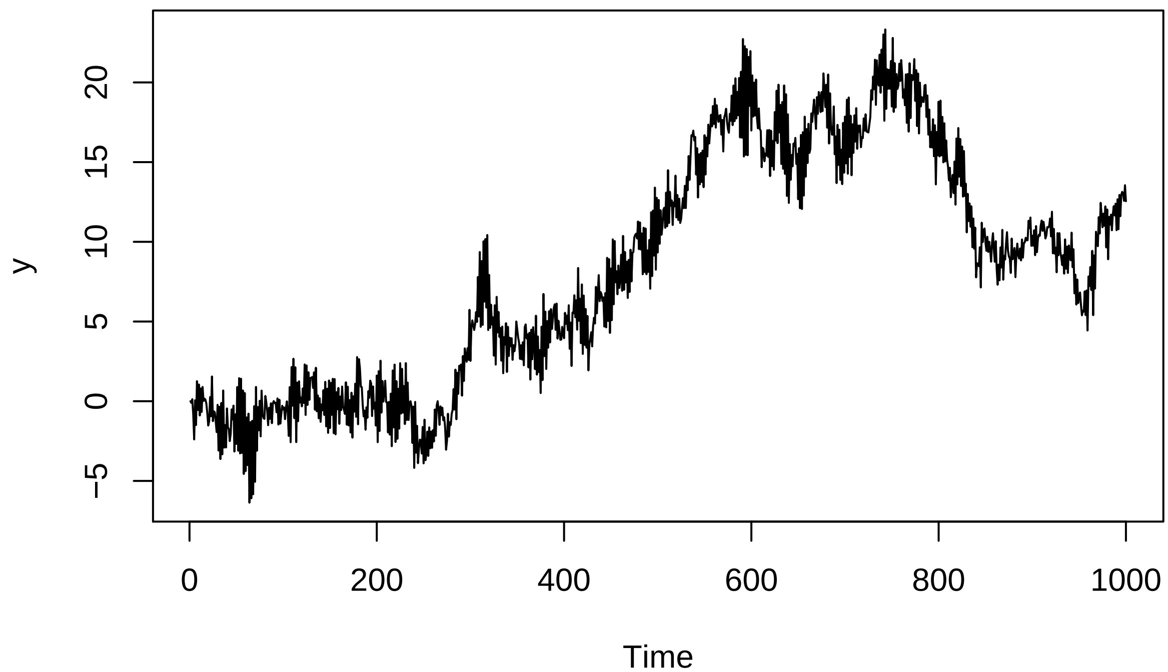
- The initial time series plot appears to have consistent mean of 0, and the ADF test confirms stationarity (p-value = 0.01).
- The PACF plot shows that the sample path may be a MA(1) model behavior, with a lag of 1.
- The EACF plot shows that the sample path may be following a ARMA(0,0) model, which is evidence that differencing may be needed. However, the next possibility in the EACF plot is an ARIMA(0,1,1) model.
- The auto.arima output for this sample path gives us a model of ARIMA(0,1,1).
- When we compare the AIC & BIC scores between our ARIMA(0,1,1) and ARIMA(0,0,1) models, we have a lower scores for the ARMA(1,1) model (AIC=7420.59).
- When we plot the residuals of the ARIMA(0,1,1) model, they fit well with the second sample path.
- Our final conclusion is that this is an ARIMA(0,1,1) model.

```
# Please specify the estimated ARIMA(p,d,q) orders below
Q2_p <- 0
Q2_d <- 1
Q2_q <- 1
Q2_Sp <- NA # [Optional] Seasonal AR() order
Q2_Sd <- NA # [Optional] Seasonal I() order
Q2_Sq <- NA # [Optional] Seasonal MA() order
Q2_S <- NA # [Optional] seasonal period
```

Question 3 (3 credits)

Please look at the `problem1$sample_path3` and identify the ARIMA order of the underlying stochastic process

```
# Please do your analysis below
y = problem1$sample_path3
ts.plot(y)
```




```
adf.test(y)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -1.4452, Lag order = 9, p-value = 0.8131
## alternative hypothesis: stationary
```

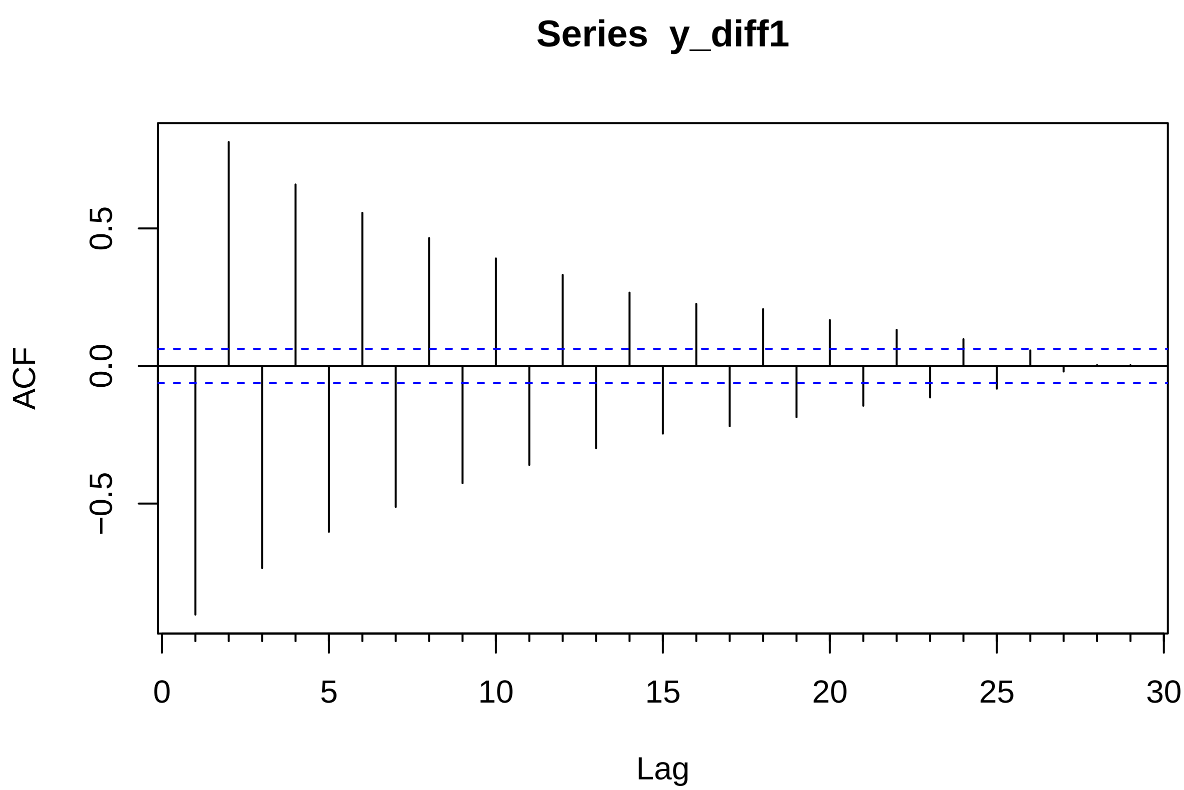
```
y_diff1=diff(y, differences = 1)
adf.test(y_diff1)
```

```
## Warning in adf.test(y_diff1): p-value smaller than printed p-value
```

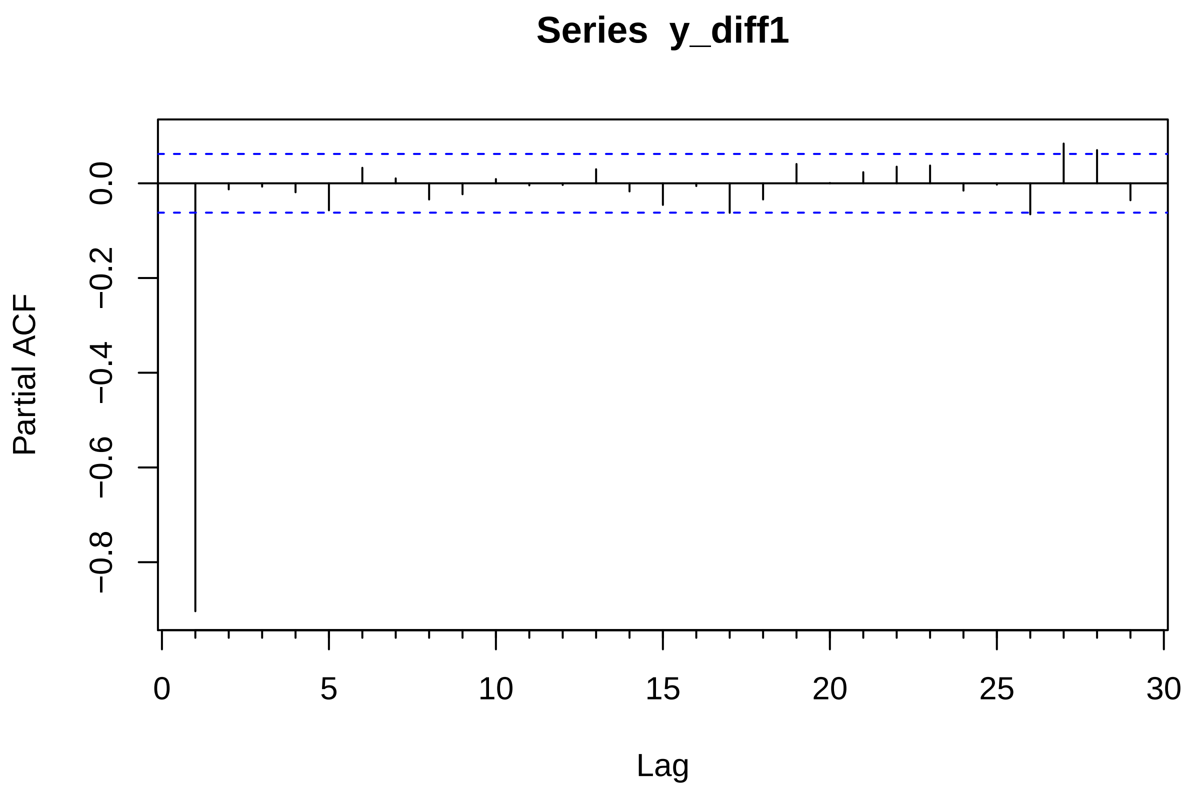
```
##
```

```
##  Augmented Dickey-Fuller Test
##
## data:  y_diff1
## Dickey-Fuller = -10.854, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```r
Acf(y_diff1)
```




```r
Pacf(y_diff1)
```



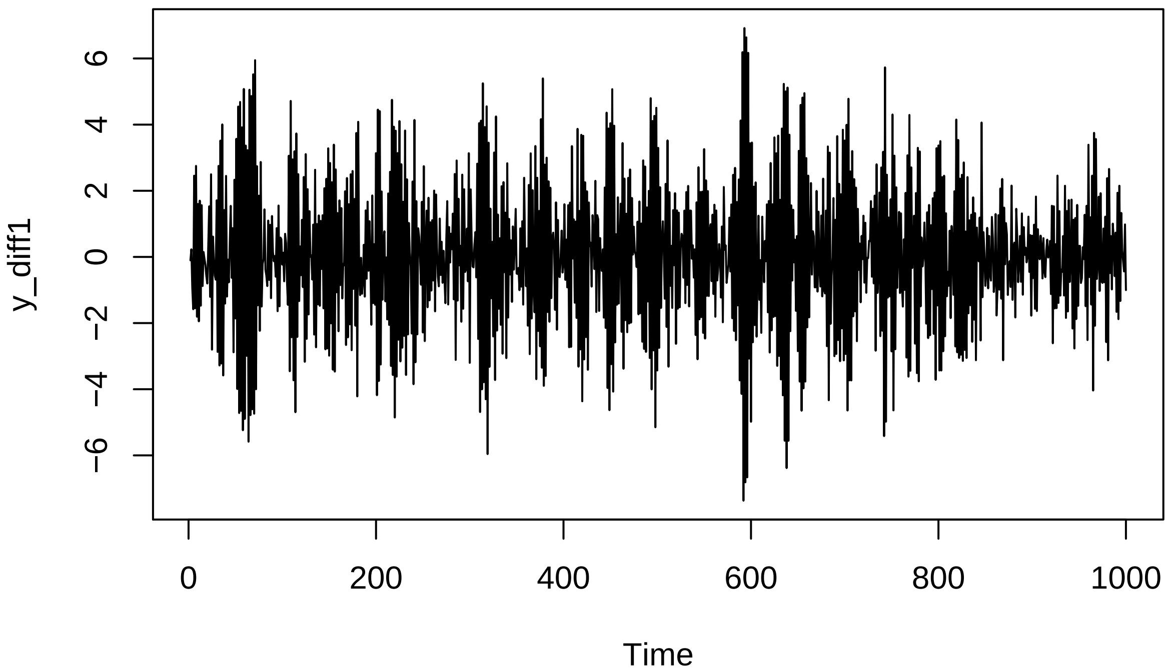

```r
eacf(y_diff1)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x x x x x x  x  x  x
## 1 o o 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 o o o o o o o o o o  o  o  o
## 4 x o x o o o o o o o o  o  o  o
## 5 x x x x o o o o o o o  o  o  o
## 6 x x x x o o o o o o o  o  o  o
## 7 x x x x o o o o o o o  o  o  o
```

```r
auto.arima(y)
```

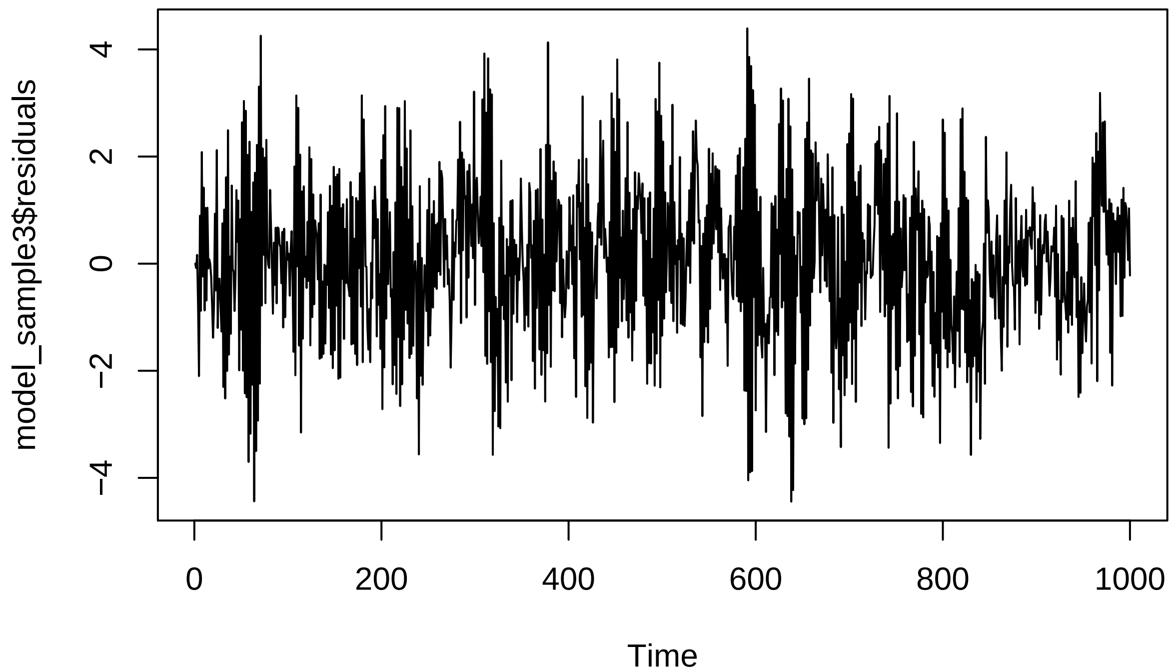
```
## Series: y
## ARIMA(1,1,0)
##
## Coefficients:
##           ar1
##       -0.9028
## s.e.   0.0135
##
## sigma^2 estimated as 1.007:  log likelihood=-1421.58
## AIC=2847.16   AICc=2847.17   BIC=2856.98
```

```r
ts.plot(y_diff1)
```

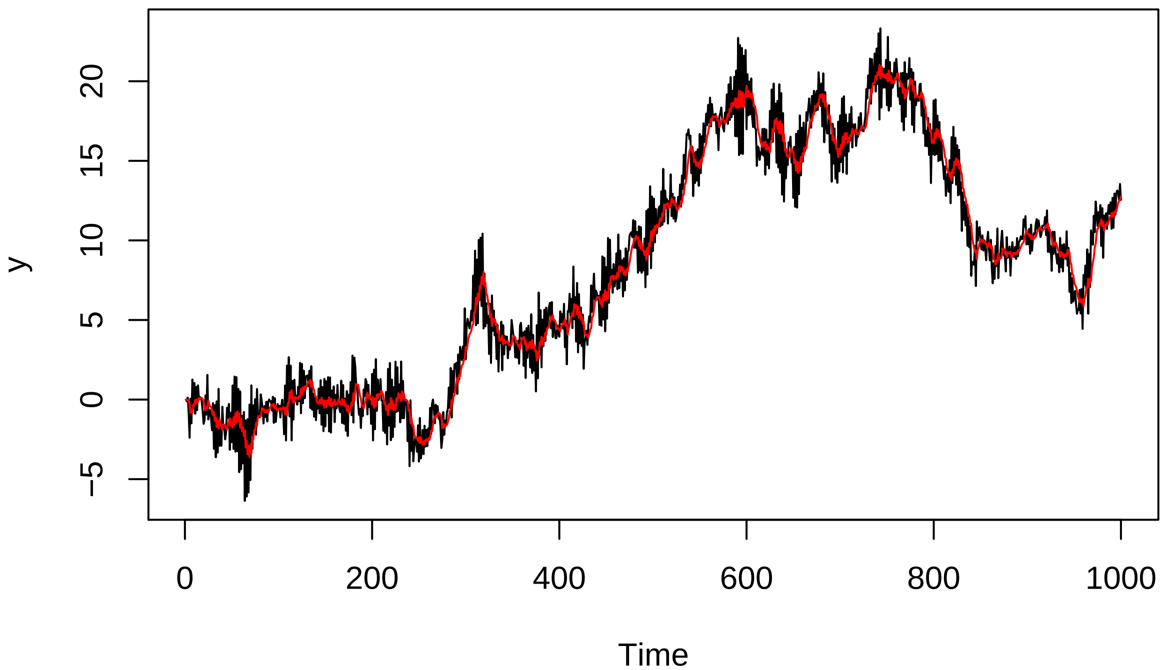



```r
model_sample3 = Arima(y,order=c(0,1,1))

plot(model_sample3$residuals)
```




```r
plot(y)
lines(y - model_sample3$residuals, col="red")
```



Please describe your reasoning below:

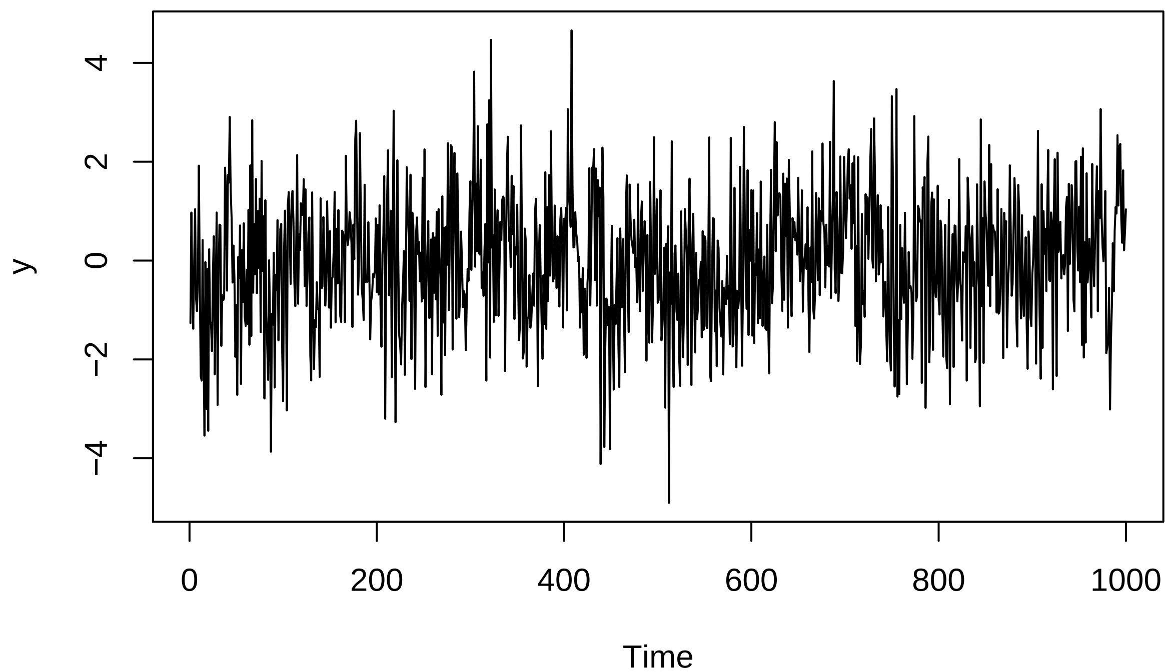
- The initial time series plot does not have consistent mean of 0, and the ADF test does not confirm stationarity (p-value = 0.8131).
- After differencing the sample path by 1, the ADF test now confirms stationarity (p-value = 0.01).
- The PACF and EACF plots show that the sample path may be following an AR(1) model.
- The auto.arima output on this sample path gives us a model of ARIMA(1,1,0); this is consistent with our PACF and EACF plots.
- When we plot the residuals of the ARIMA(0,1,1) model, they fit well with the third sample path.
- Our final conclusion is that this is an ARIMA(1,1,0) model.

```r
# Please specify the estimated ARIMA(p,d,q) orders below
Q3_p <- 1
Q3_d <- 1
Q3_q <- 0
Q3_Sp <- NA # [Optional] Seasonal AR() order
Q3_Sd <- NA # [Optional] Seasonal I() order
Q3_Sq <- NA # [Optional] Seasonal MA() order
Q3_S <- NA # [Optional] seasonal period
```

Question 4 (3 credits)

Please look at the `problem1$sample_path4` and identify the ARIMA order of the underlying stochastic process

```r
# Please do your analysis below
y = problem1$sample_path4
ts.plot(y)
```

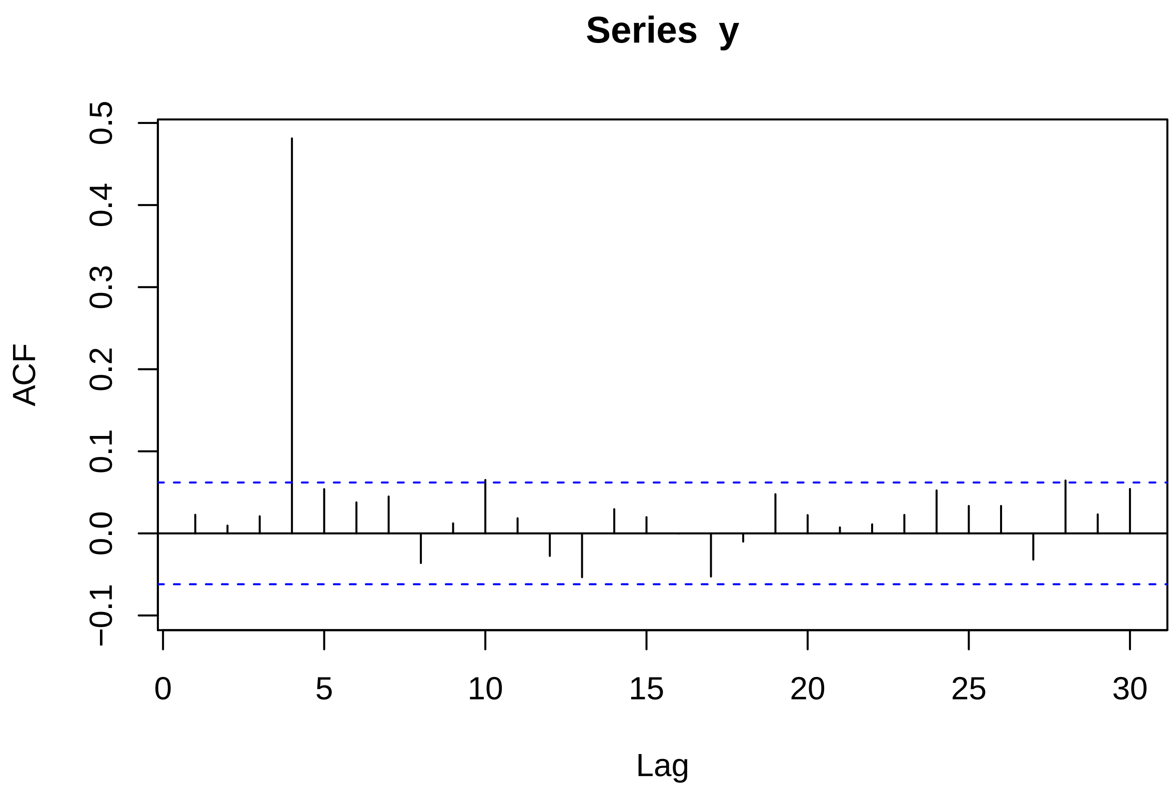



```r
adf.test(y)
```

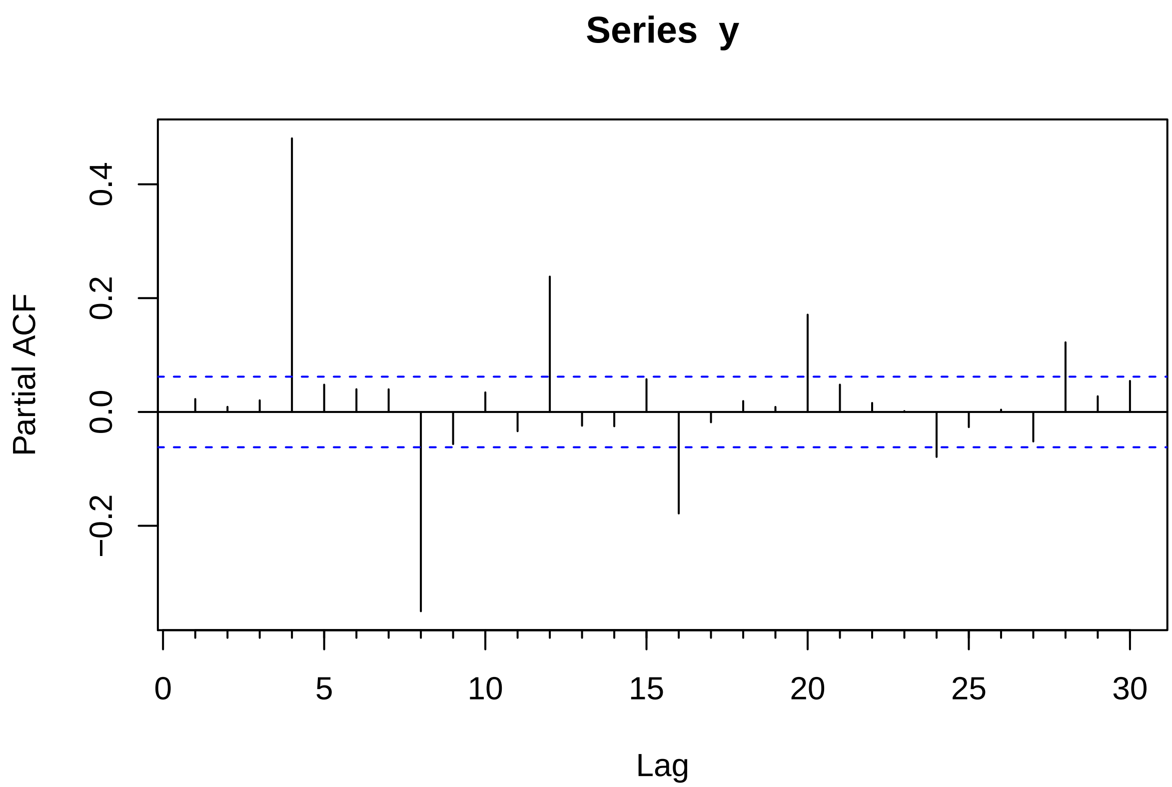
```
## Warning in adf.test(y): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  y
## Dickey-Fuller = -9.3712, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```r
Acf(y)
```




```
Pacf(y)
```




```r
eacf(y)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o o x o o o o o x o  o  o  o
## 1 x o o x x o o o o o o  o  o  o
## 2 x o o x x o o o o o o  o  o  o
## 3 o o o x x x x o o o o  o  o  o
## 4 x x x x x x x x o o o  o  o  o
## 5 x o o x x x x o x o o  o  o  o
## 6 x o o x x x x o o x o  o  o  o
## 7 x x x x x x x x o x x  o  o  o
```

```r
auto.arima(y)
```

```
## Series: y
## ARIMA(4,0,1) with non-zero mean
##
## Coefficients:
##          ar1     ar2     ar3     ar4      ma1     mean
##       0.1012  0.0026  0.0090  0.4792  -0.1147  -0.0327
## s.e.  0.0586  0.0279  0.0279  0.0280   0.0675   0.0813
##
## sigma^2 estimated as 1.426:  log likelihood=-1593.99
## AIC=3201.99   AICc=3202.1   BIC=3236.34
```

```r
Arima(y,order=c(4,0,1))
```

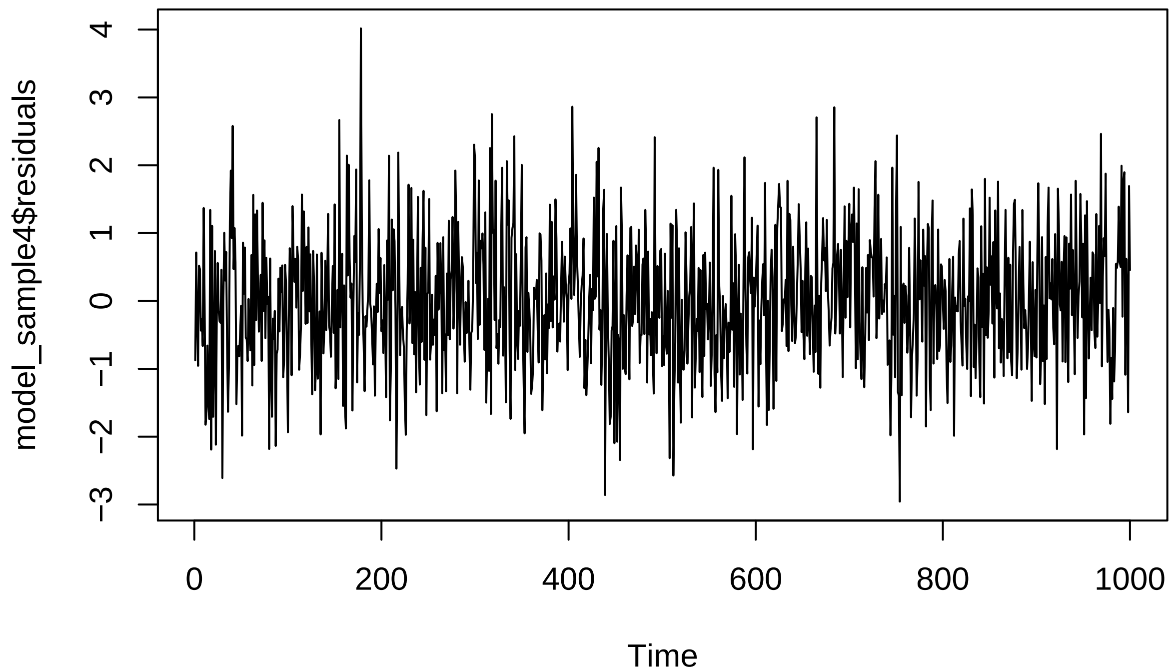
```
## Series: y
## ARIMA(4,0,1) with non-zero mean
##
## Coefficients:
##          ar1     ar2     ar3     ar4      ma1     mean
##       0.1012  0.0026  0.0090  0.4792  -0.1147  -0.0327
## s.e.  0.0586  0.0279  0.0279  0.0280   0.0675   0.0813
##
## sigma^2 estimated as 1.426:  log likelihood=-1593.99
## AIC=3201.99   AICc=3202.1   BIC=3236.34
```

```r
Arima(y, seasonal=list(order=c(0,0,1), period = 4))
```

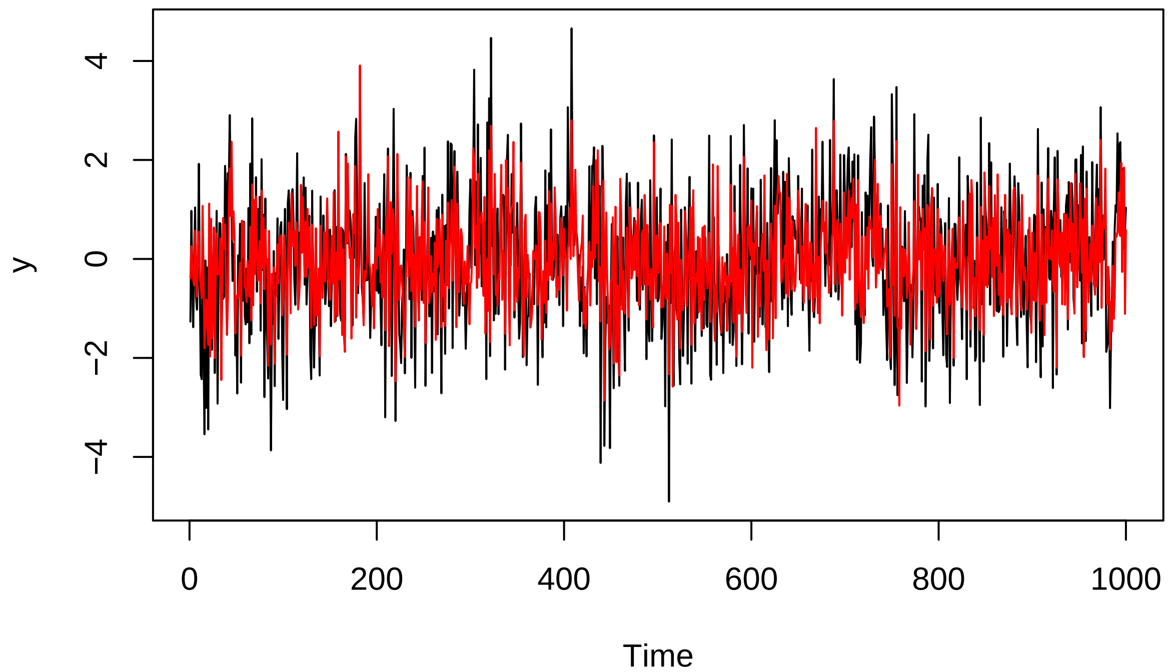
```
## Series: y
## ARIMA(0,0,0)(0,0,1)[4] with non-zero mean
##
## Coefficients:
##         sma1     mean
##       0.9908  -0.0340
## s.e.  0.0117   0.0621
##
## sigma^2 estimated as 0.9783:  log likelihood=-1414.94
## AIC=2835.88   AICc=2835.9   BIC=2850.6
```

```r
model_sample4 = Arima(y, seasonal=list(order=c(0,0,1), period = 4))

plot(model_sample4$residuals)
```




```
plot(y)
lines(y - model_sample4$residuals, col="red")
```



Please describe your reasoning below:

- The initial time series plot appears to have consistent mean of 0, and the ADF test confirms stationarity (p-value = 0.01).
- The ACF plot shows that the sample path may be a MA(1) model, with a lag of 4.
- The PACF plot shows a seasonal trend of 4.
- The auto.arima output on this sample path shows ARIMA(4,0,1) which gives the output from the ACF plot more validity that the sample path is an MA(1); however, this does not capture the seasonal nature of the sample path.
- When we compare the AIC & BIC scores between our ARIMA(0,0,0)(0,0,1)[4] and ARIMA(4,0,1) models, we have a lower scores for the ARIMA(0,0,0)(0,0,1)[4] model (AIC=2835.88).
- When we plot the residuals of the ARIMA(0,0,0)(0,0,1)[4] model, they fit well with the fourth sample path.
- Our final conclusion is that this is an ARIMA(0,0,0)(0,0,1)[4] model.

```
# Please specify the estimated ARIMA(p,d,q) orders below
Q4_p <- 0
Q4_d <- 0
Q4_q <- 0
Q4_Sp <- 0 # [Optional] Seasonal AR() order
Q4_Sd <- 0 # [Optional] Seasonal I() order
Q4_Sq <- 1 # [Optional] Seasonal MA() order
Q4_S <- 4 # [Optional] seasonal period
```