Problem 1 (12 credits)

HW3

Maya Carnie, Xianging He, Olivia Liang, Lauren Moore, Ilana Novakoski, William Wu; carni015, he000273, linag625, moor1985, novak560, wuxx1066

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

```
## Warning: package 'TSA' was built under R version 3.5.3

## Warning: package 'forecast' was built under R version 3.5.3

## Warning: package 'ggplot2' was built under R version 3.5.3

## Warning: package 'dplyr' was built under R version 3.5.3

## Warning: package 'tseries' was built under R version 3.5.3
```

Identify given stochastic processes

Please load the data from file `problem1.Rds`

```r
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.

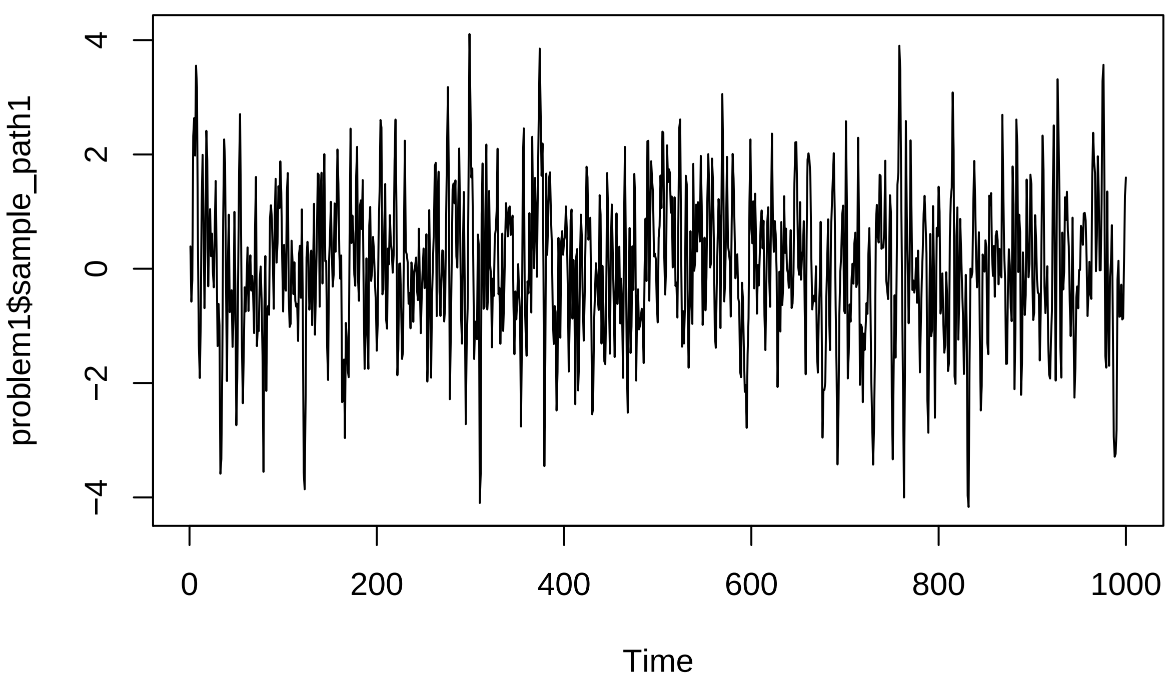
General Requirements

- 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

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




```r
#test whether stationary or not
adf.test(problem1$sample_path1)
```

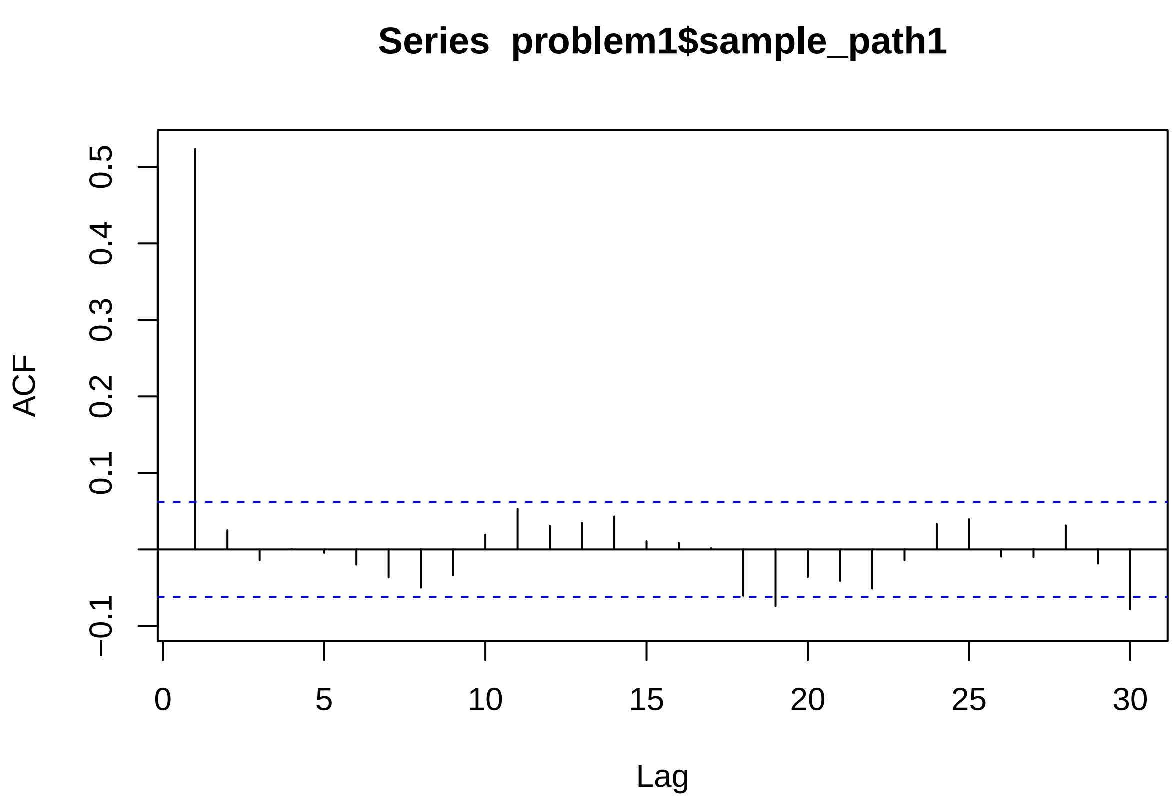
```
## Warning in adf.test(problem1$sample_path1): p-value smaller than printed p-
## value

##
##  Augmented Dickey-Fuller Test
##
## data:  problem1$sample_path1
## Dickey-Fuller = -9.7278, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

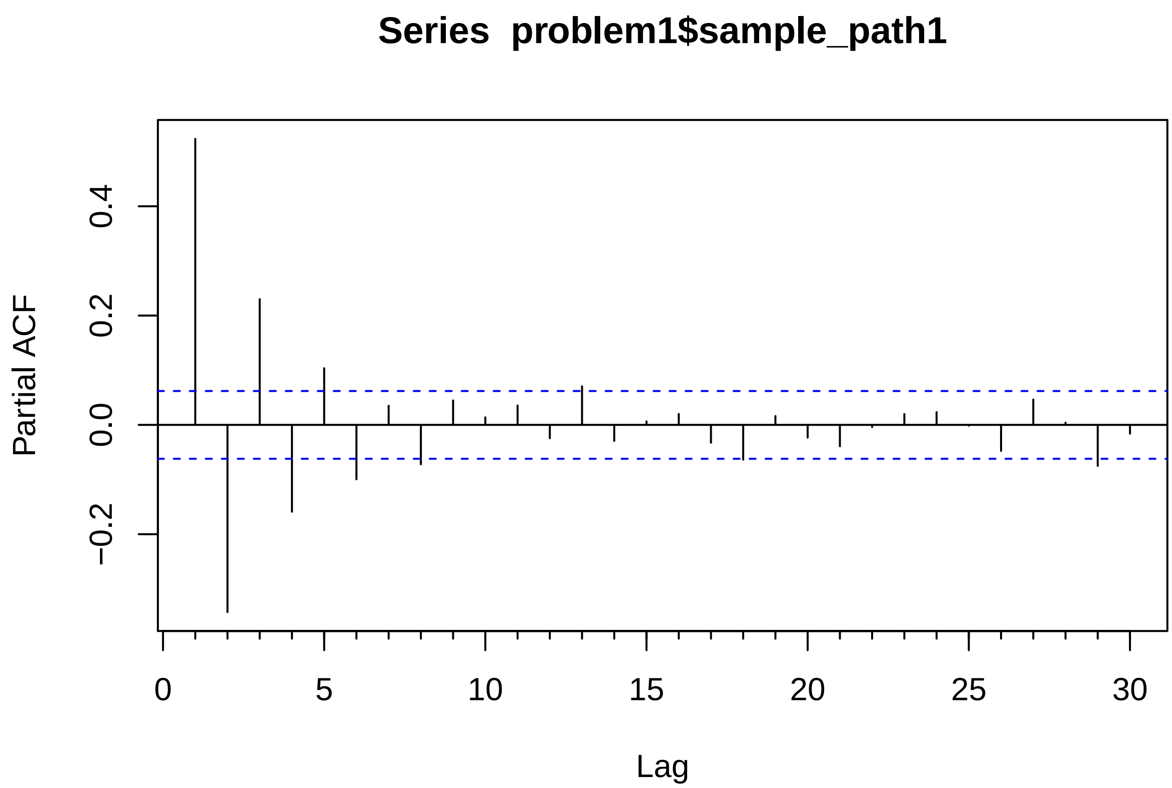
```
# the p-value is 0.01, which means that it may be stationary
model1= auto.arima(problem1$sample_path1)
model1
```

```
## Series: problem1$sample_path1
## 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
```

```
Acf(problem1$sample_path1)
```



```r
#ACF is used for moving average(q)
Pacf(problem1$sample_path1)
```




```r
#determine AR(p)
eacf(problem1$sample_path1)
```

```
## 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
#determine(p,q) in arima
Arima(problem1$sample_path1, order = c(0,0,1))
```

```
## Series: problem1$sample_path1
## 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(problem1$sample_path1, order = c(0,0,2))
```

```
## Series: problem1$sample_path1
## 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
# Please do your analysis above
```

Please describe your reasoning below

> ADF test shows the process is likely to be stationary. The ACF plot matches an MA(q) process. It has 1 tall bar followed by many shorter ones that don't follow a tailing off pattern. The PACF plot appears to tail off, supporting an MA(q) process The EACF plot suggests an ARMA(0,1) model

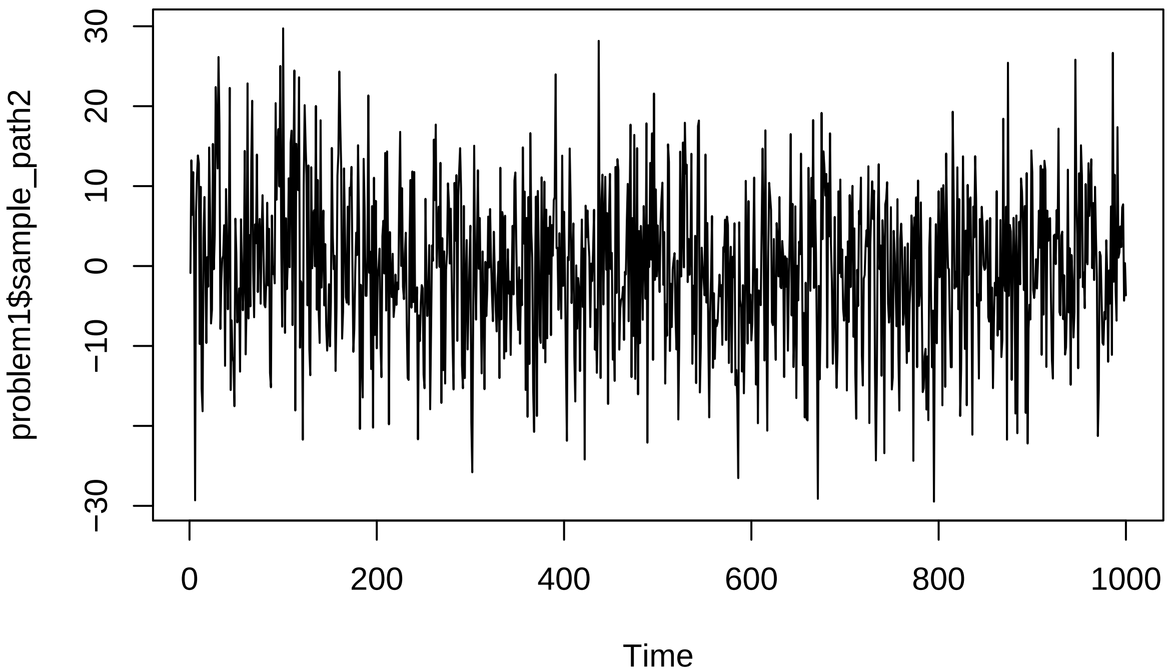
Please describe your reasoning above

```r
# Please specify the estimated ARIMA(p,d,q) orders below
Q1_p <- 0
Q1_d <- 0
Q1_q <- 1
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
ts.plot(problem1$sample_path2)
```

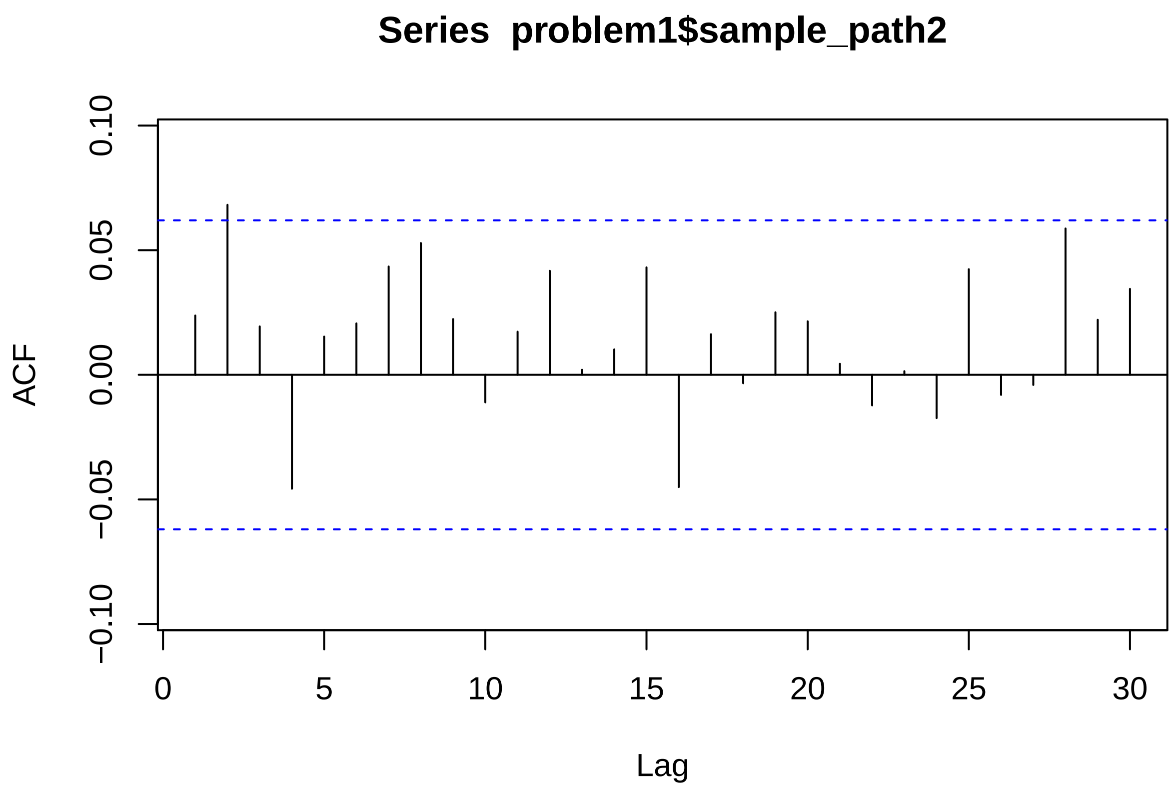



```r
adf.test(problem1$sample_path2)
```

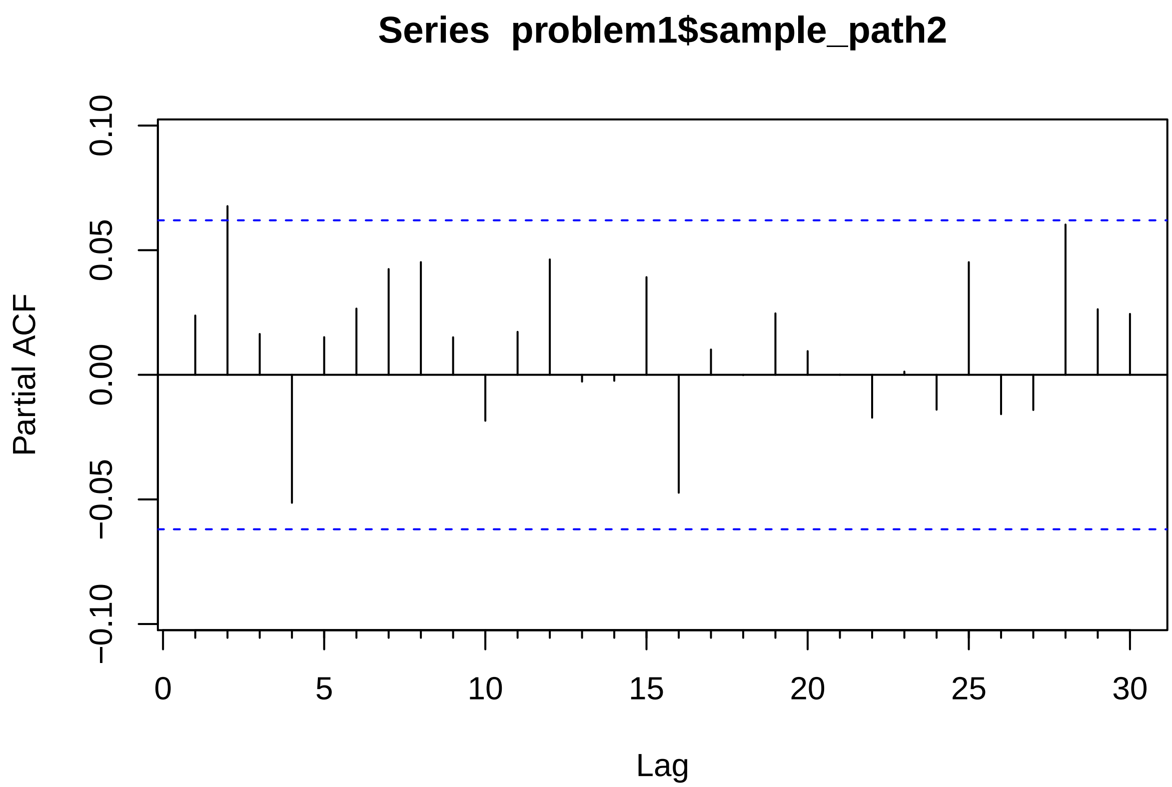
```
## Warning in adf.test(problem1$sample_path2): p-value smaller than printed p-
## value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  problem1$sample_path2
## Dickey-Fuller = -9.3444, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```r
Acf(problem1$sample_path2)
```



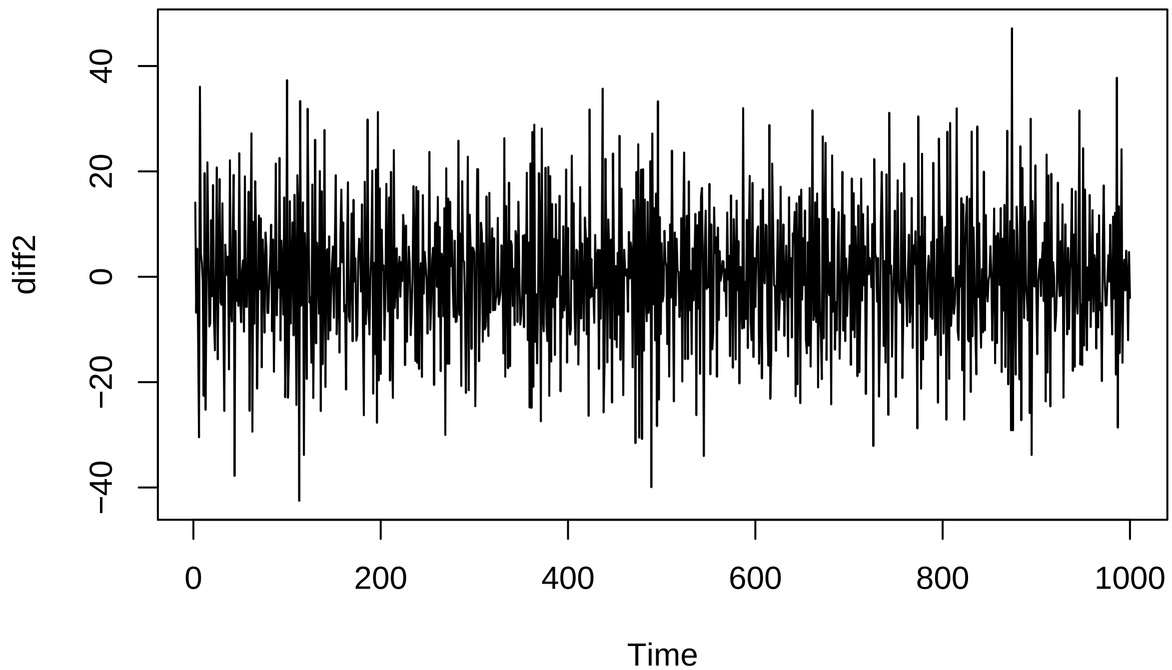

```r
Pacf(problem1$sample_path2)
```



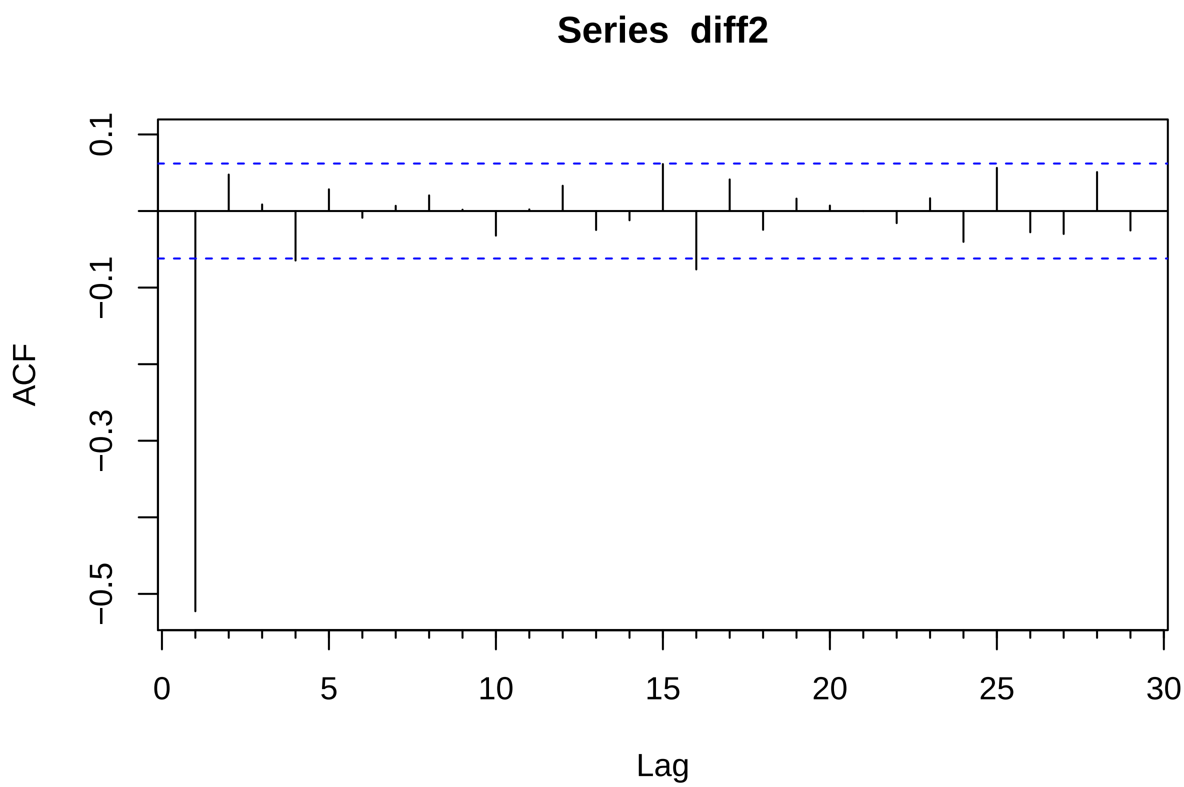

```r
eacf(problem1$sample_path2)
```

```
## 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
```

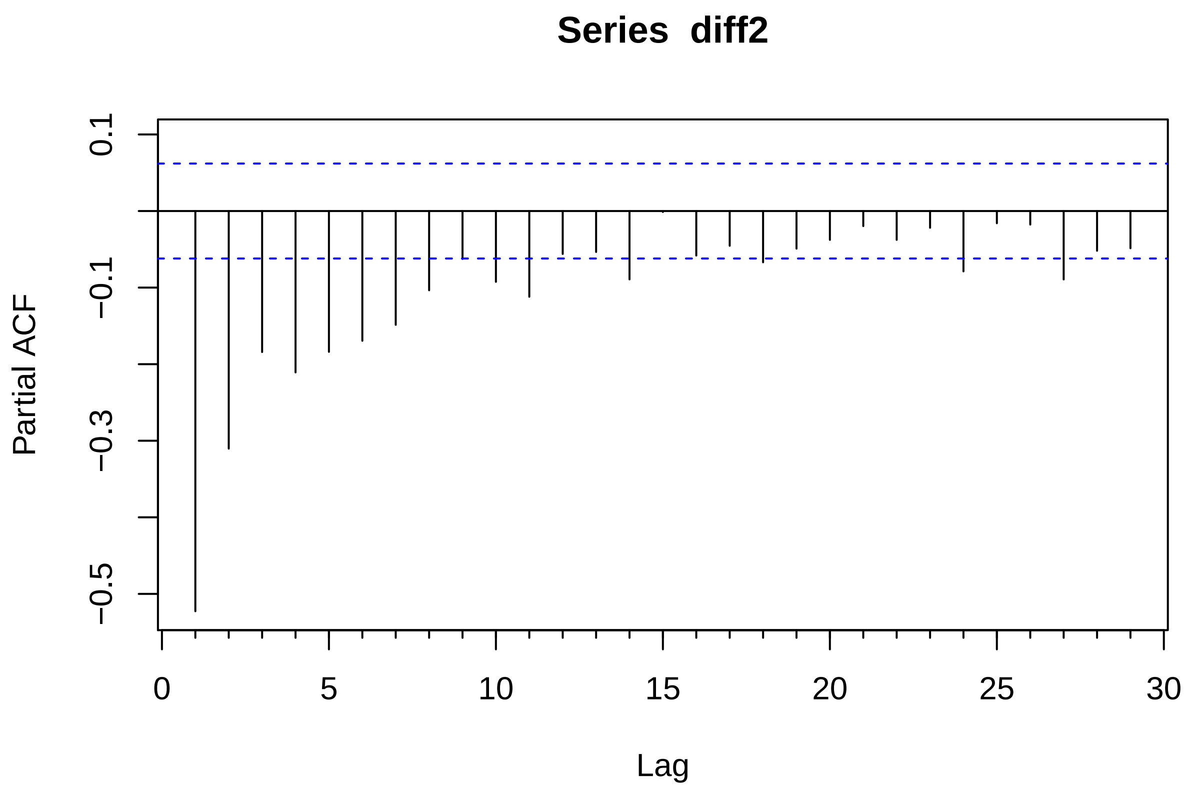
```r
diff2 = diff(problem1$sample_path2)
ts.plot(diff2)
```




```
Acf(diff2)
```




```r
Pacf(diff2)
```




```
eacf(diff2)
```

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

```
#ACF is used for moving average(q)
#determine AR(p)

modelsp2 = auto.arima(problem1$sample_path2, seasonal = TRUE)#arima(0,1,1)
Arima(problem1$sample_path2, order = c(0,1,1))
```

```
## Series: problem1$sample_path2
## 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
#determine(p,q) in arima
# it is an arma model

# Please do your analysis above
```

Please describe your reasoning below

> ADF test suggests the process is stationary ACF and PACF plots don't show any significant pattern Adding a change in the model shows something different ACF and PACF follows an MA(q) model, showing one peak and tailing off respectively EACF suggest order for ARMA is (0,1)

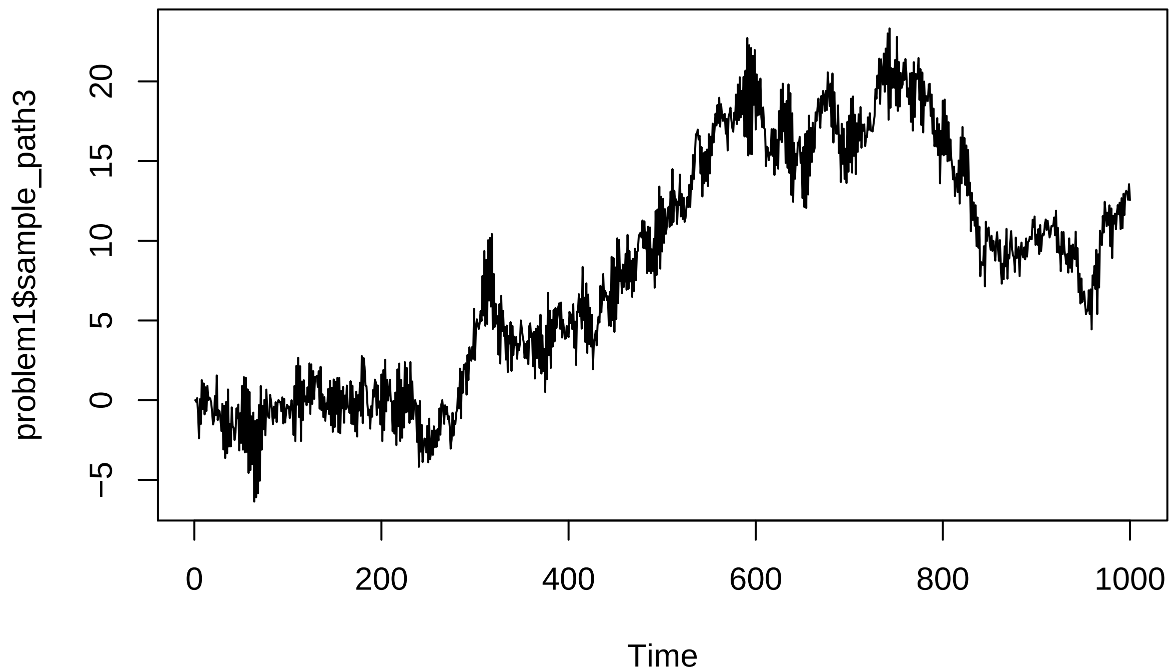
Please describe your reasoning above

```r
# 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

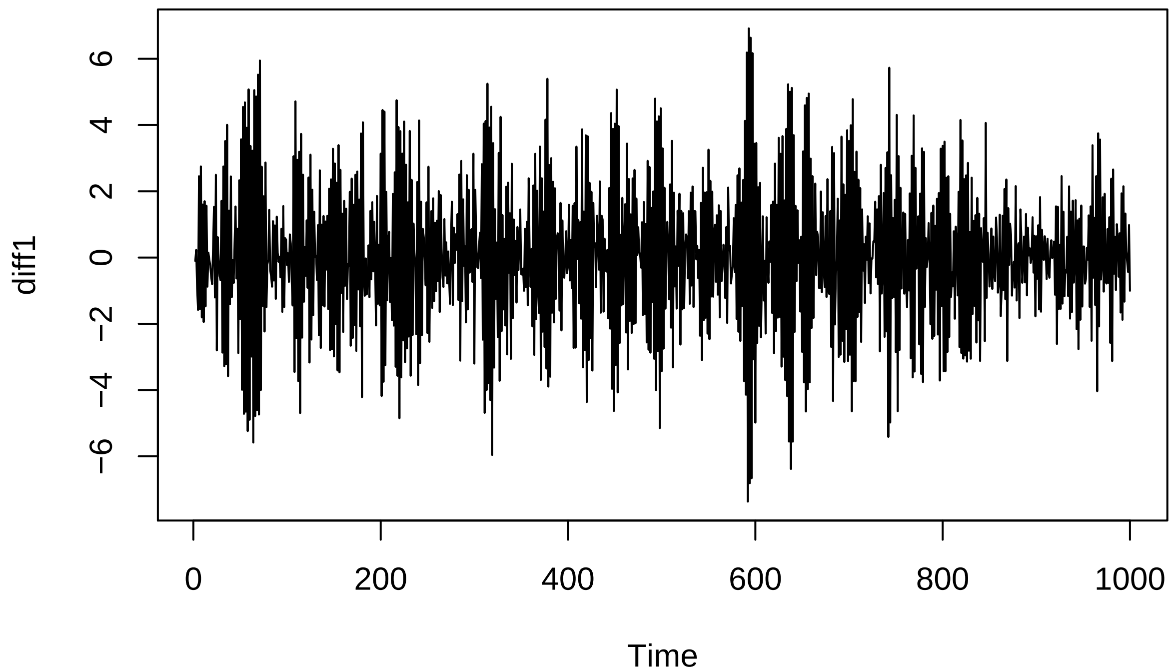
```r
# Please do your analysis below
ts.plot(problem1$sample_path3)
```



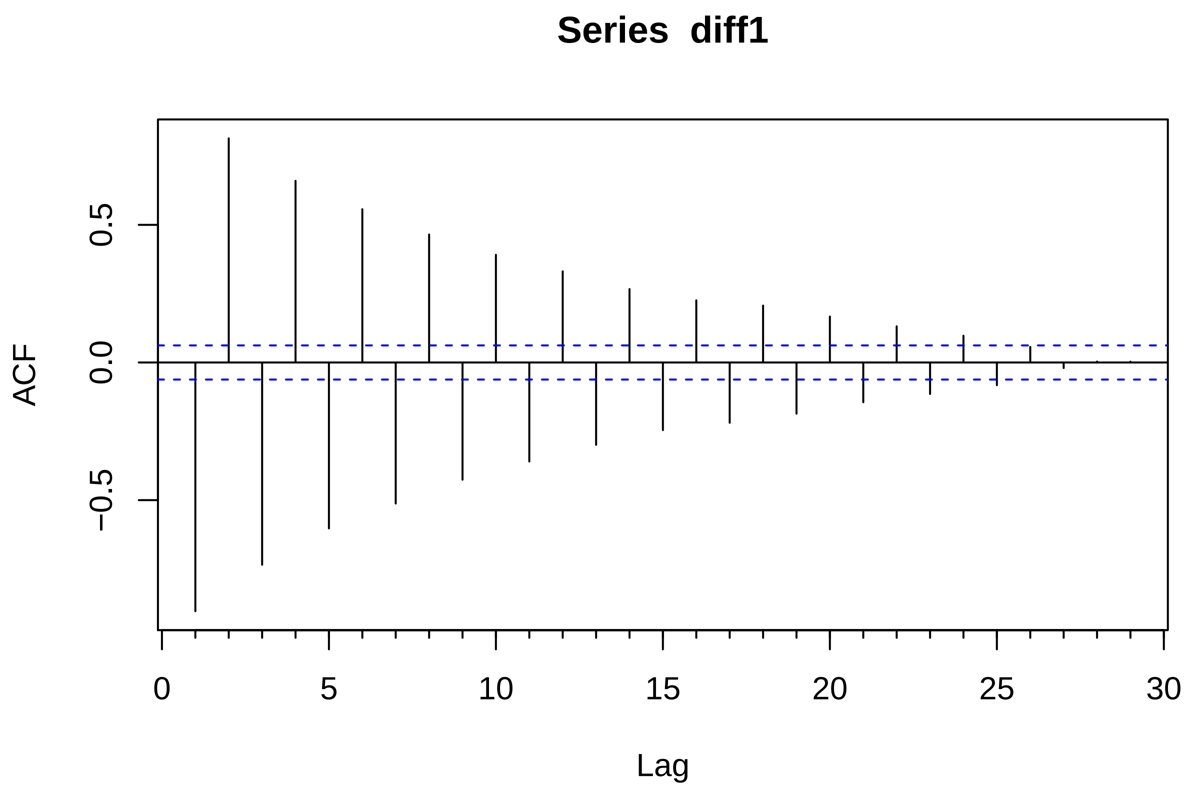

```r
adf.test(problem1$sample_path3)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  problem1$sample_path3
## Dickey-Fuller = -1.4452, Lag order = 9, p-value = 0.8131
## alternative hypothesis: stationary
```

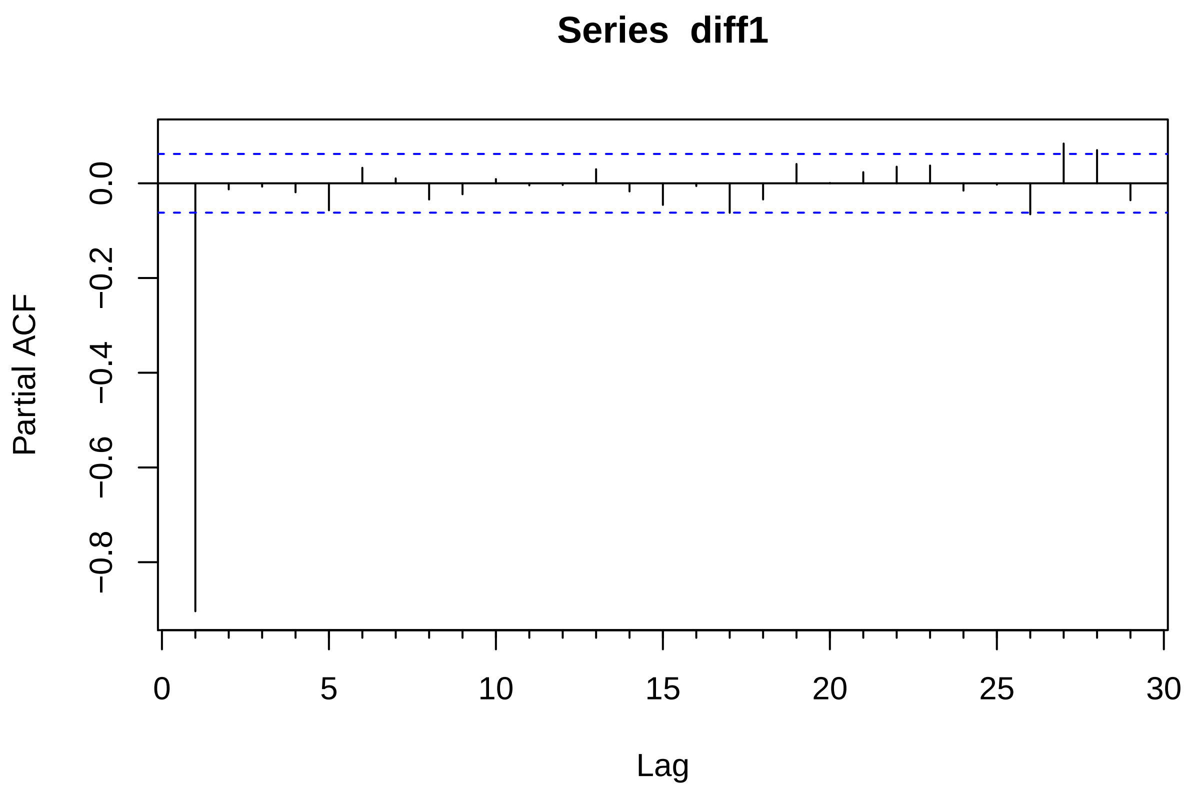
```r
# it is obviously not stationary
diff1 = diff(problem1$sample_path3)
# diff1 is stationary
ts.plot(diff1)
```




```
Acf(diff1)
```




```r
#ACF is used for moving average(q)
Pacf(diff1)
```




```r
#determine AR(p)
eacf(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
# Please do your analysis above
```

Please describe your reasoning below

> ADF test shows that this process is likely not stationary After adding a difference to the process, the ACF and PACF align with an AR(p) model The EACF chart supports an of ARMA(1,0)

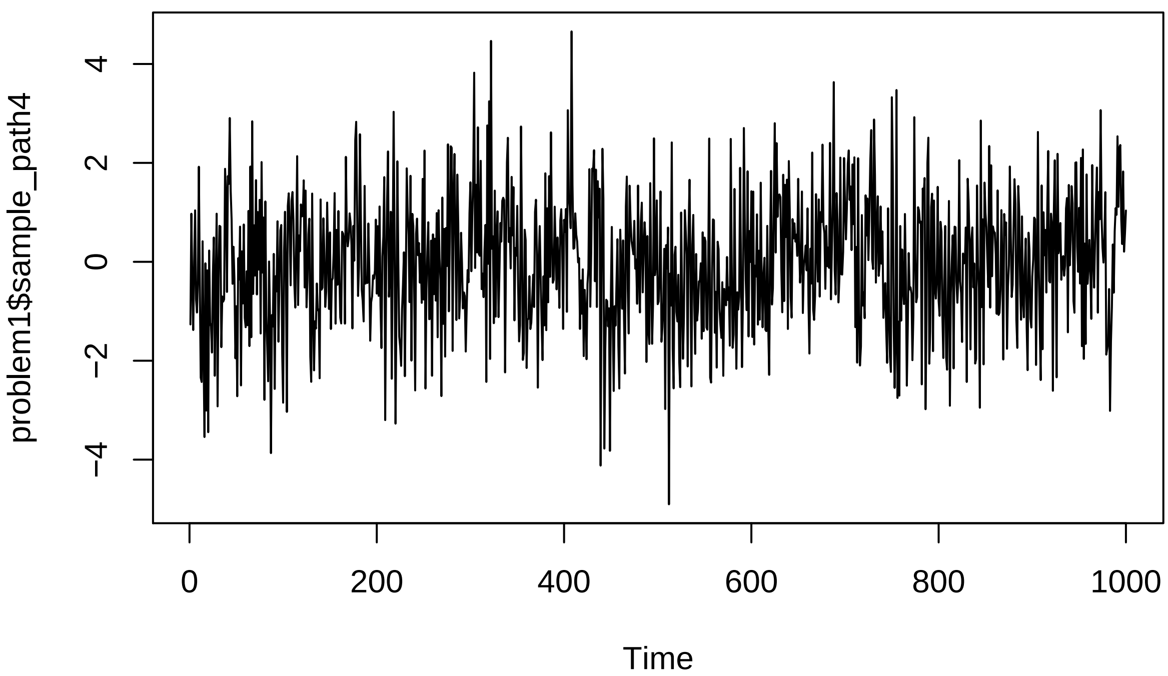
Please describe your reasoning above

```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
ts.plot(problem1$sample_path4)
```



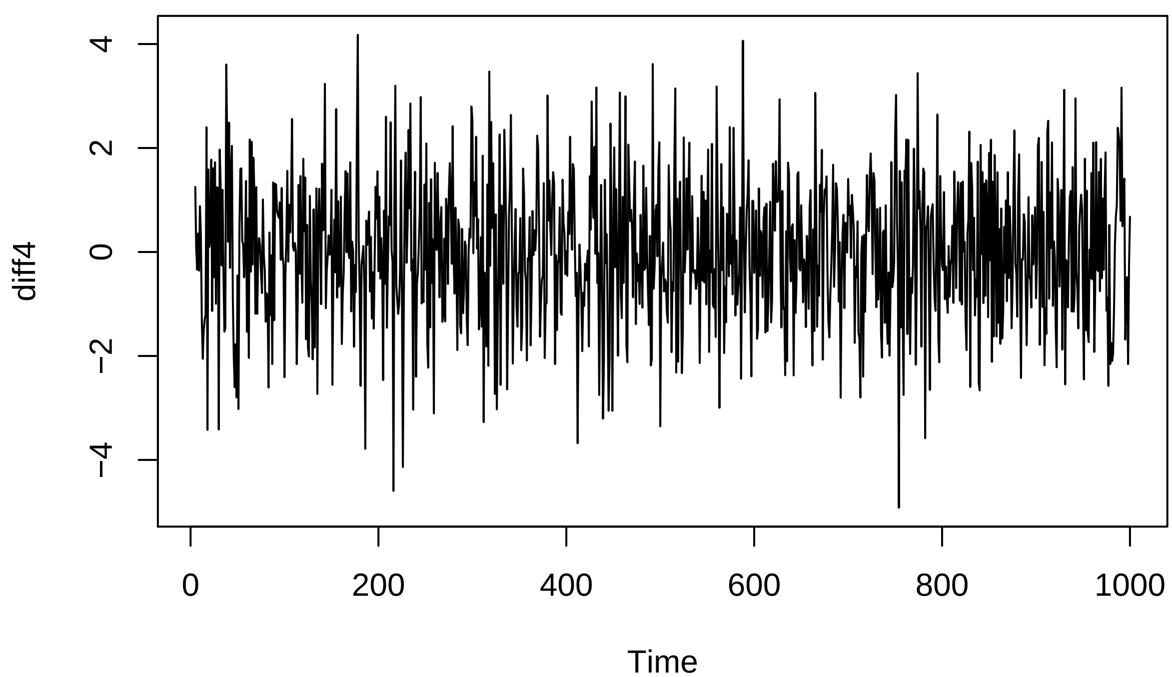

```r
adf.test(problem1$sample_path4)
```

```
## Warning in adf.test(problem1$sample_path4): p-value smaller than printed p-
## value
```

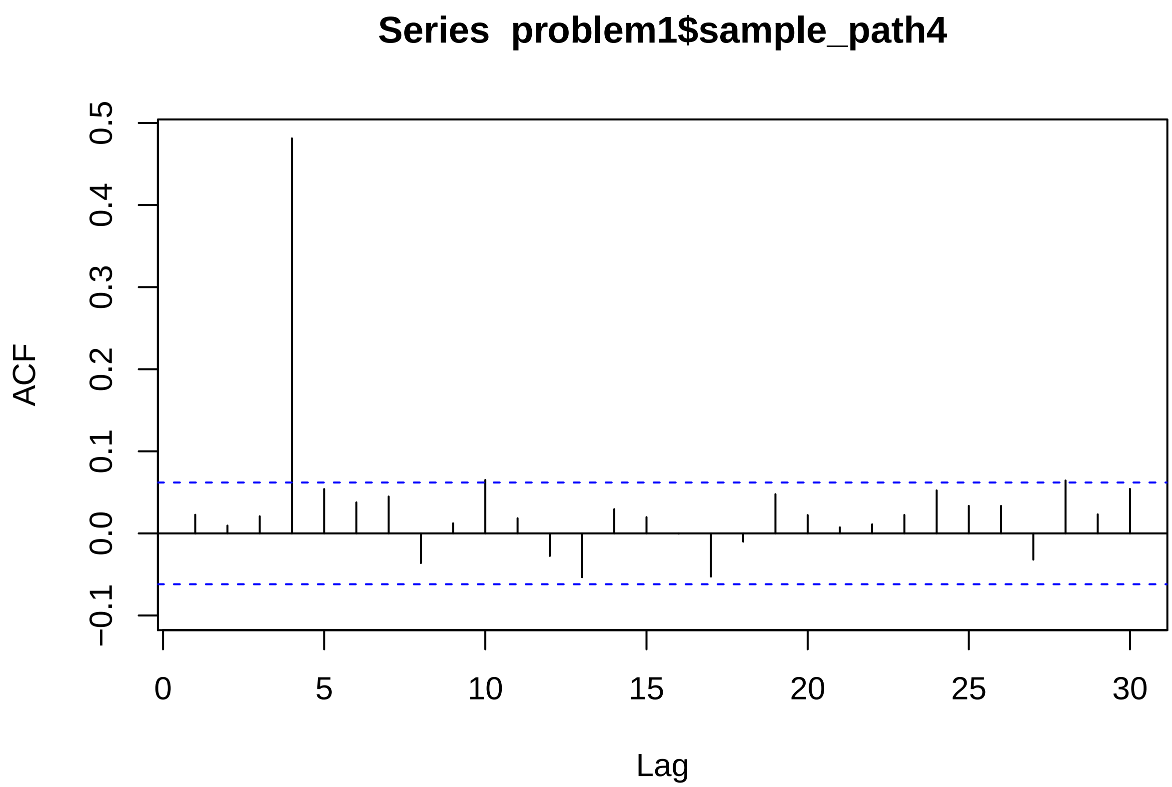
```
##
```

```
##  Augmented Dickey-Fuller Test
##
## data:  problem1$sample_path4
## Dickey-Fuller = -9.3712, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

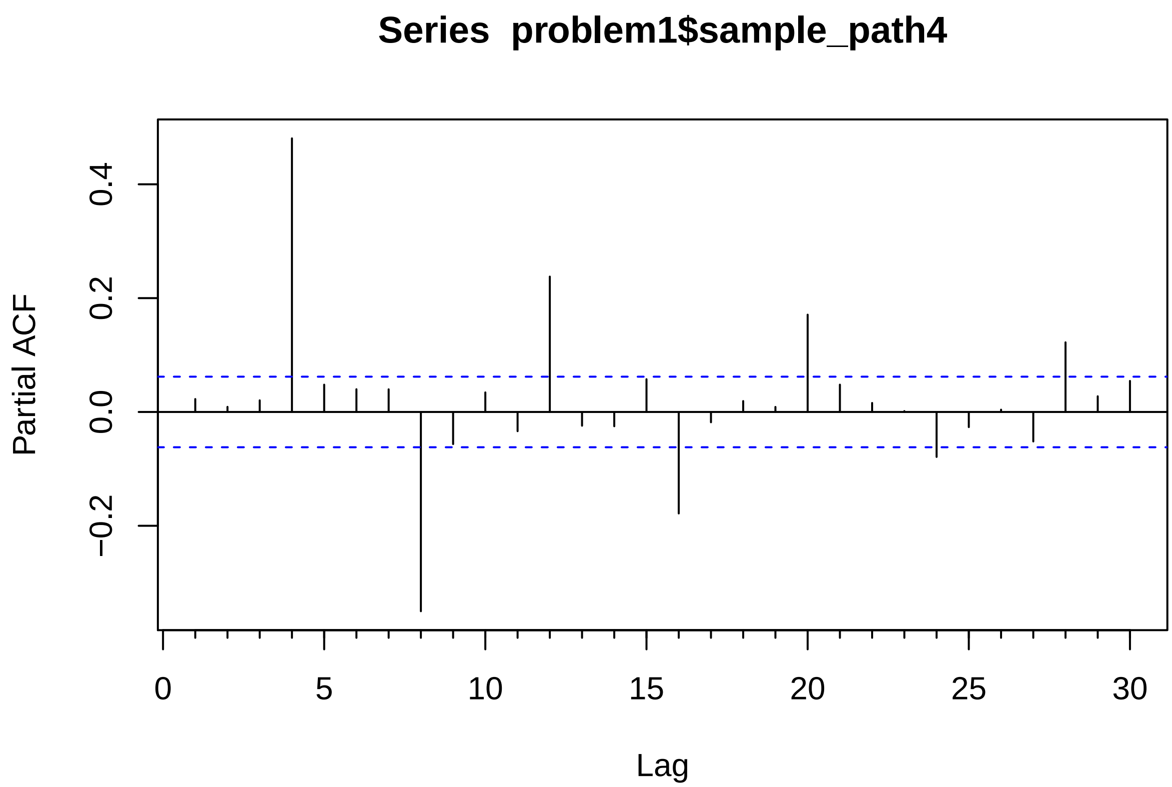
```r
diff4 = diff(problem1$sample_path4,difference = 1, lag = 4)
ts.plot(diff4)
```




```r
Acf(problem1$sample_path4)
```



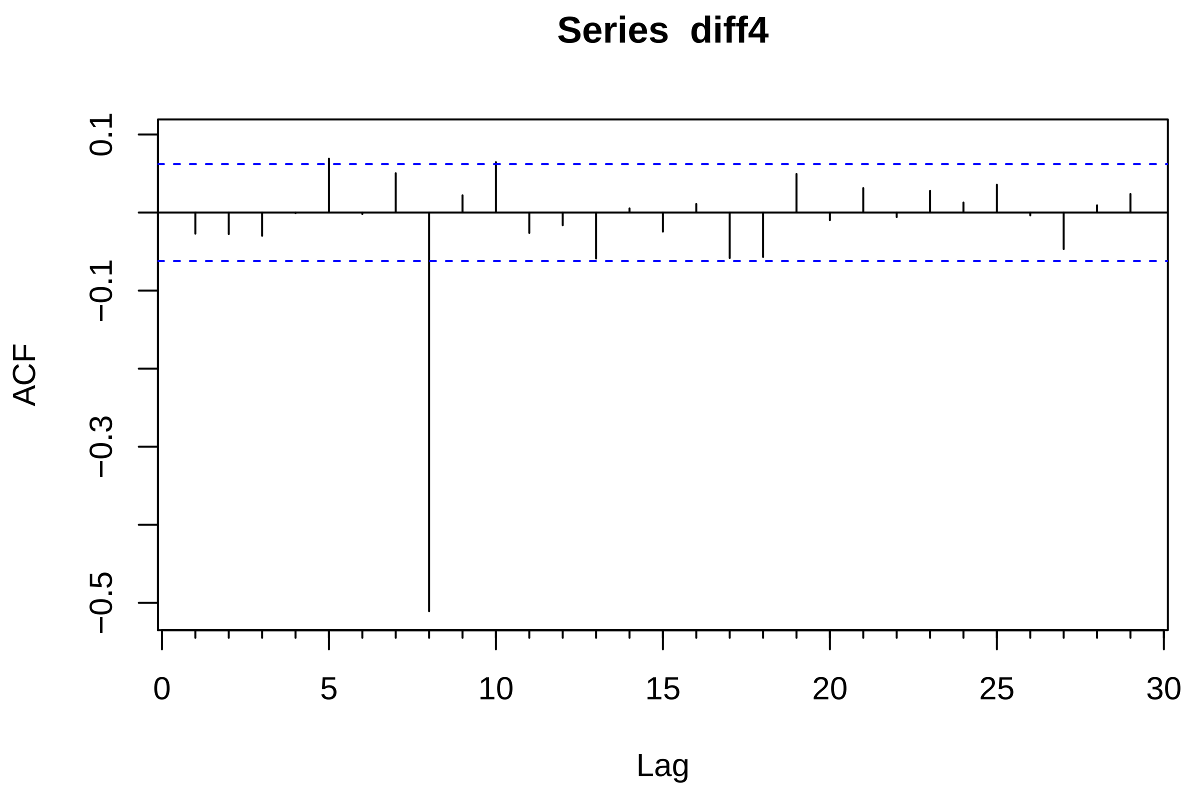

```r
#ACF is used for moving average(q)
Pacf(problem1$sample_path4)
```



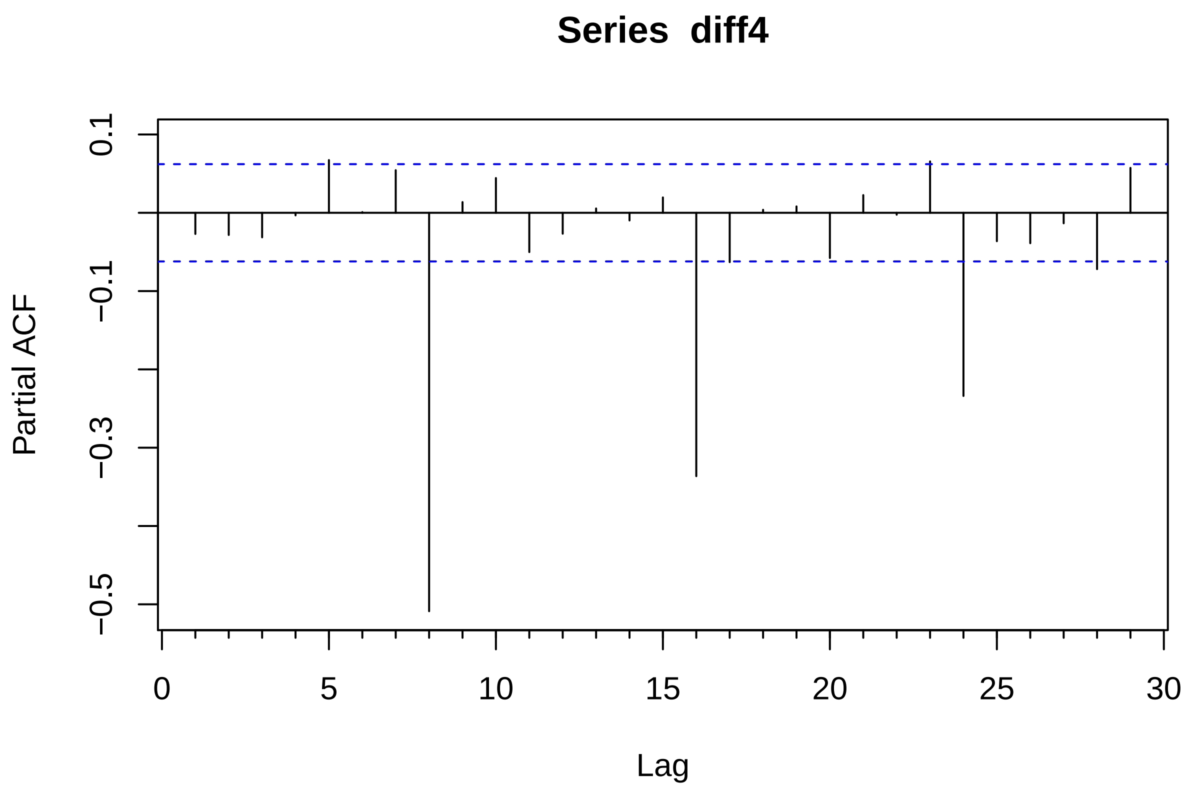

```r
#from the ACF and Pacf, we can find there are seasonality
#determine AR(p)
eacf(problem1$sample_path4)
```

```
## 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
Acf(diff4)
```




```
Pacf(diff4)
```




```r
eacf(diff4)
```

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

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

Please describe your reasoning below

> The PACF function seems to show that this is the seasonal model, with peaks at every 4th lag. The ACF also peaks at 4 and cuts off entirely. This implies an MA(q) model The ACF is a single peak, supporting an order of 1

Please describe your reasoning above

```r
# Please specify the estimated ARIMA(p,d,q) orders below
Q4_p <- NA
Q4_d <- NA
Q4_q <- NA
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
```