Problem 1

HW3

Hanyuan Chi(chixx105), Zhi Shen(shenx704)

February 19, 2017

```
suppressPackageStartupMessages({
  library(TSA)
  library(forecast)
  library(ggplot2)
  library(dplyr)
})
```

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.

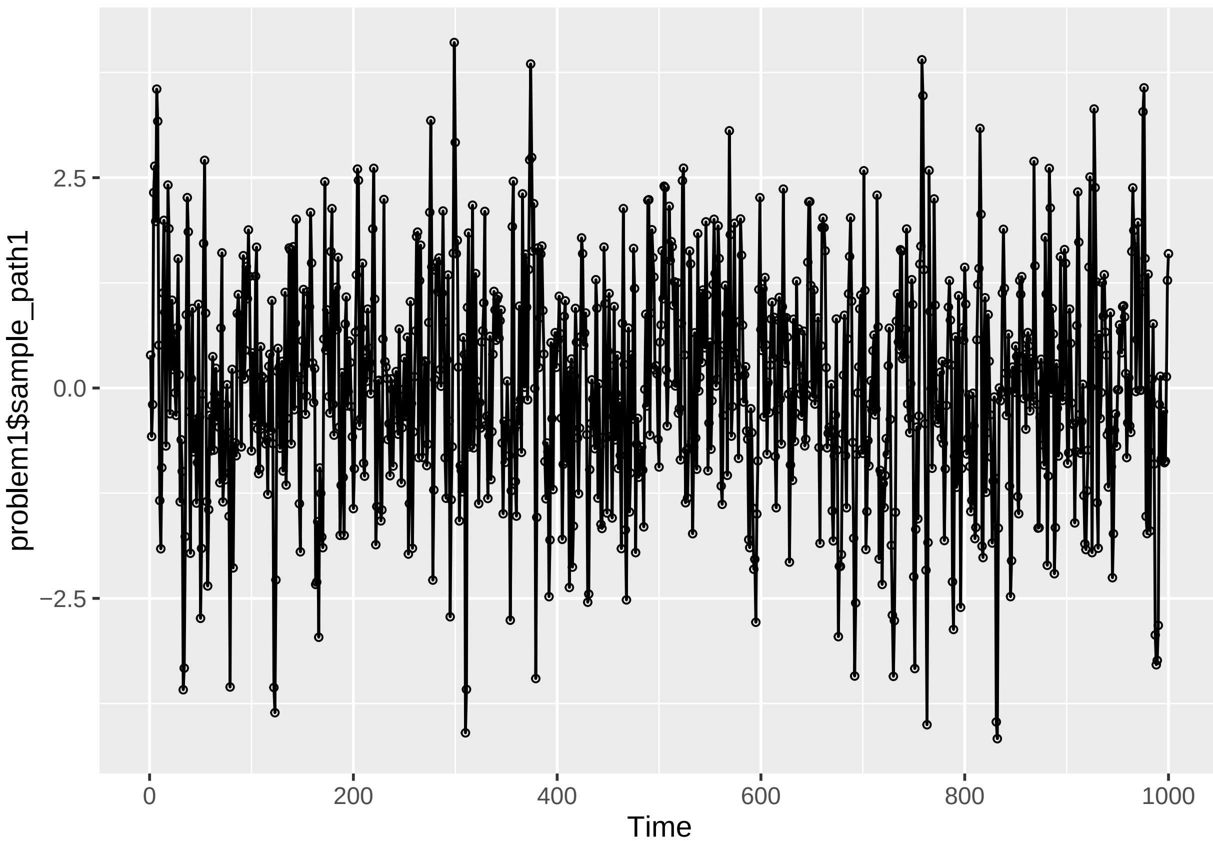
General Requirements

- 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`.

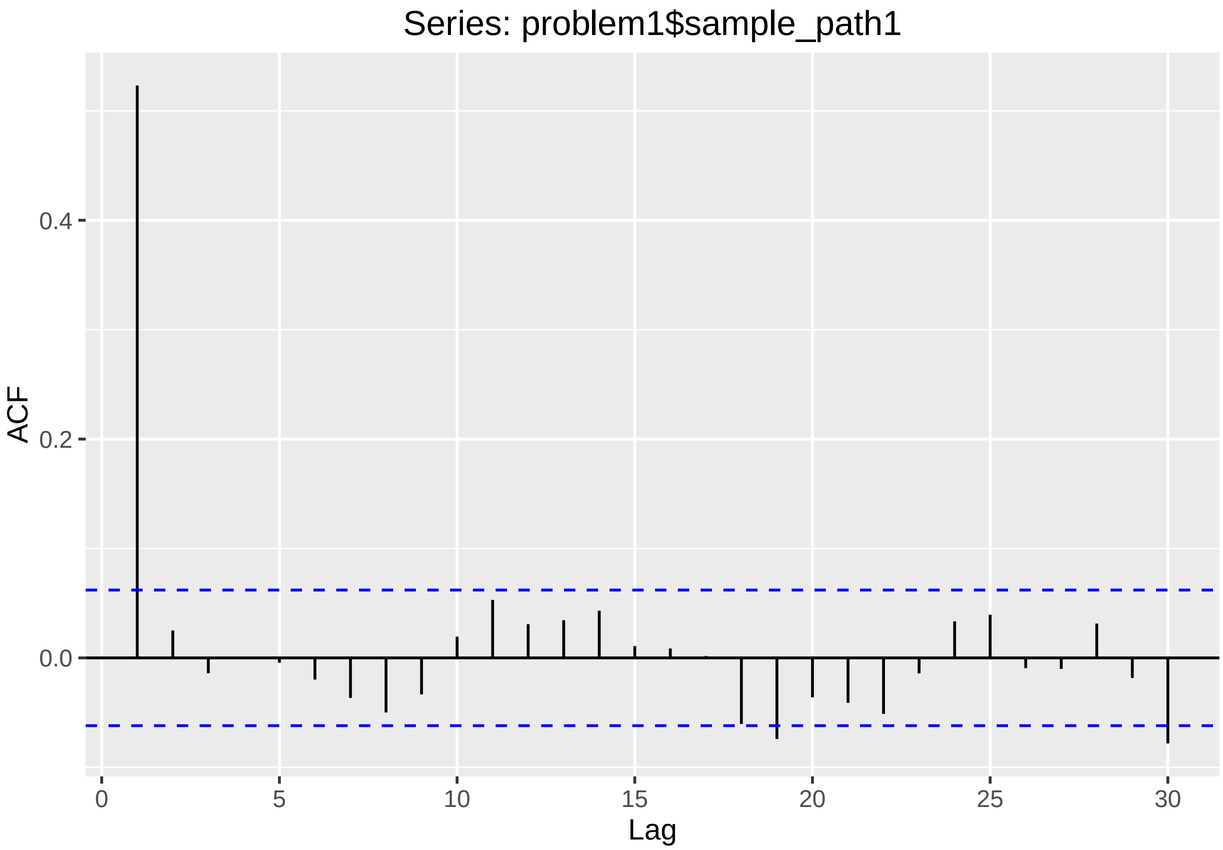
Question 1

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

```
# Please do your analysis below
autoplot(problem1$sample_path1) + geom_point(shape = 1, size = 1)
```




```
ggAcf(problem1$sample_path1)
```




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

Please describe your reasoning below

The process seems to be stationary with no seasonality based on the plot. After plotting the autocorrelation function, we can see that autocorrelation at lag 1 clearly sticks outside of the band/noise threshold, which means a strong auto-correlation at lag 1 and nothing beyond that. This is a clear indication of the moving average process of an order 1, with no autoregressive process.

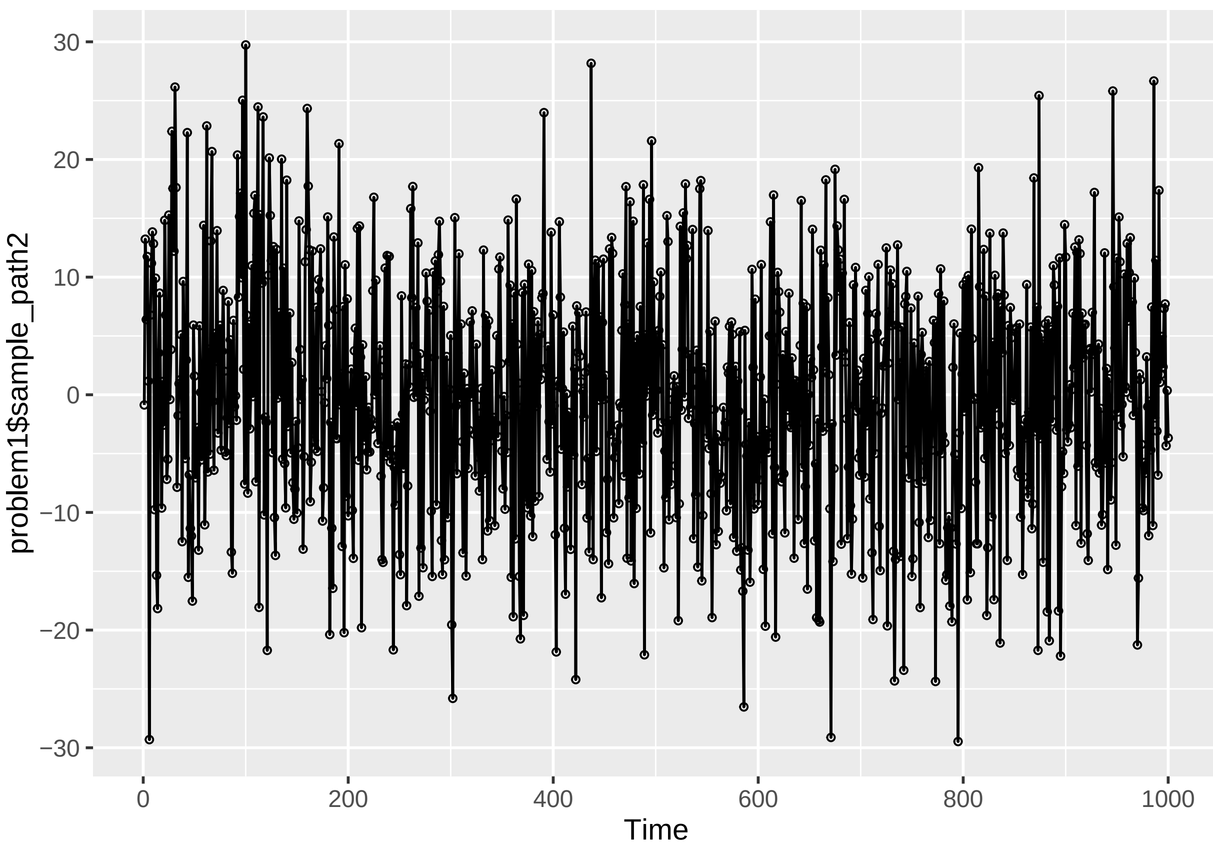
Please describe your reasoning above

```
# Please specify the estimated ARIMA(p,d,q) orders below
Q1_p <- 0
Q1_d <- 0
Q1_q <- 1
Q1_Sp <- 0 # [Optional] Seasonal AR() order
Q1_Sd <- 0 # [Optional] Seasonal I() order
Q1_Sq <- 0 # [Optional] Seasonal MA() order
Q1_S <- NA # [Optional] seasonal period
```

Question 2

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

```
# Please do your analysis below
autoplot(problem1$sample_path2) + geom_point(shape = 1, size = 1)
```

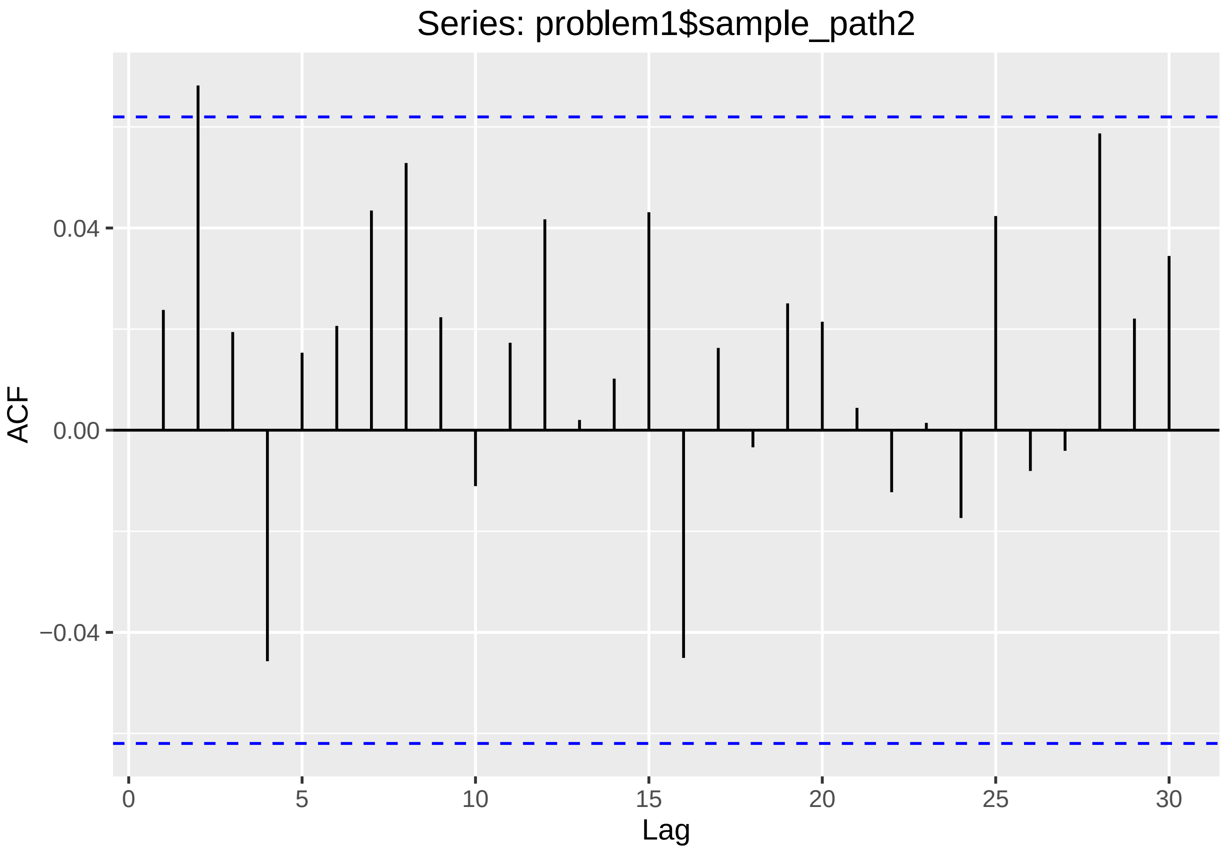



```r
adf.test(problem1$sample_path2,alternative = "stationary")
```

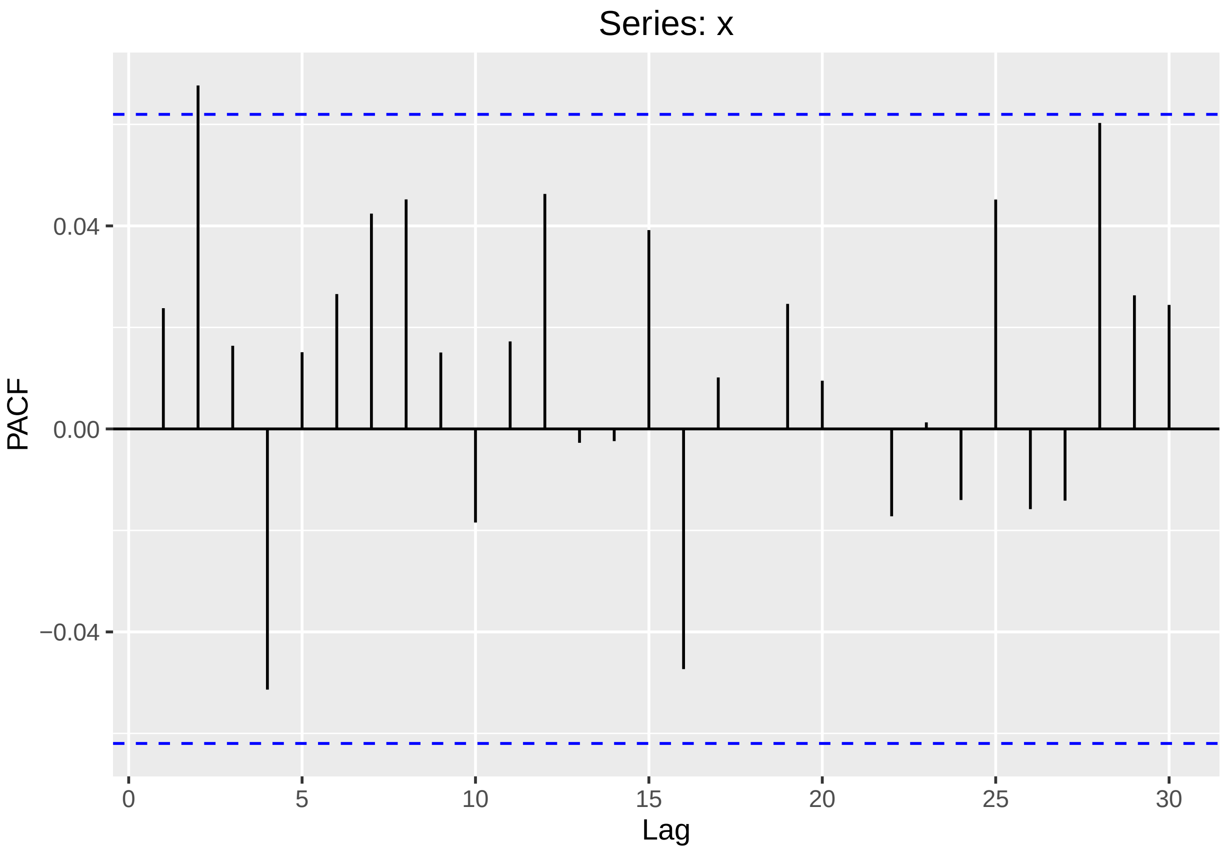
```
## Warning in adf.test(problem1$sample_path2, alternative = "stationary"): p-
## value smaller than printed p-value
```

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

```r
ggAcf(problem1$sample_path2)
```




```r
ggPacf(problem1$sample_path2)
```




```r
Box.test(problem1$sample_path2, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  problem1$sample_path2
## X-squared = 0.56751, df = 1, p-value = 0.4512
```

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

Please describe your reasoning below

The process seems to be stationary with no seasonality based on the plot. Using adf test, we can tell there is no random walk component based on the small p-value, so the process is stationary. After plotting the autocorrelation function and partial autocorrelation function, we can see that all autocorrelations and partial autocorrelations are within the band or slightly outside the band, so we can't confirm the existence of autocorrelation or the order of it. We used Ljung-Box test to see whether autocorrelations of this sample path are different from zero. The p-value is larger than 0.05, so we can't reject that values of this sample path are independent, which means this times series is white noise. So the original process should be ARIMA(0,0,0) .

Please describe your reasoning above

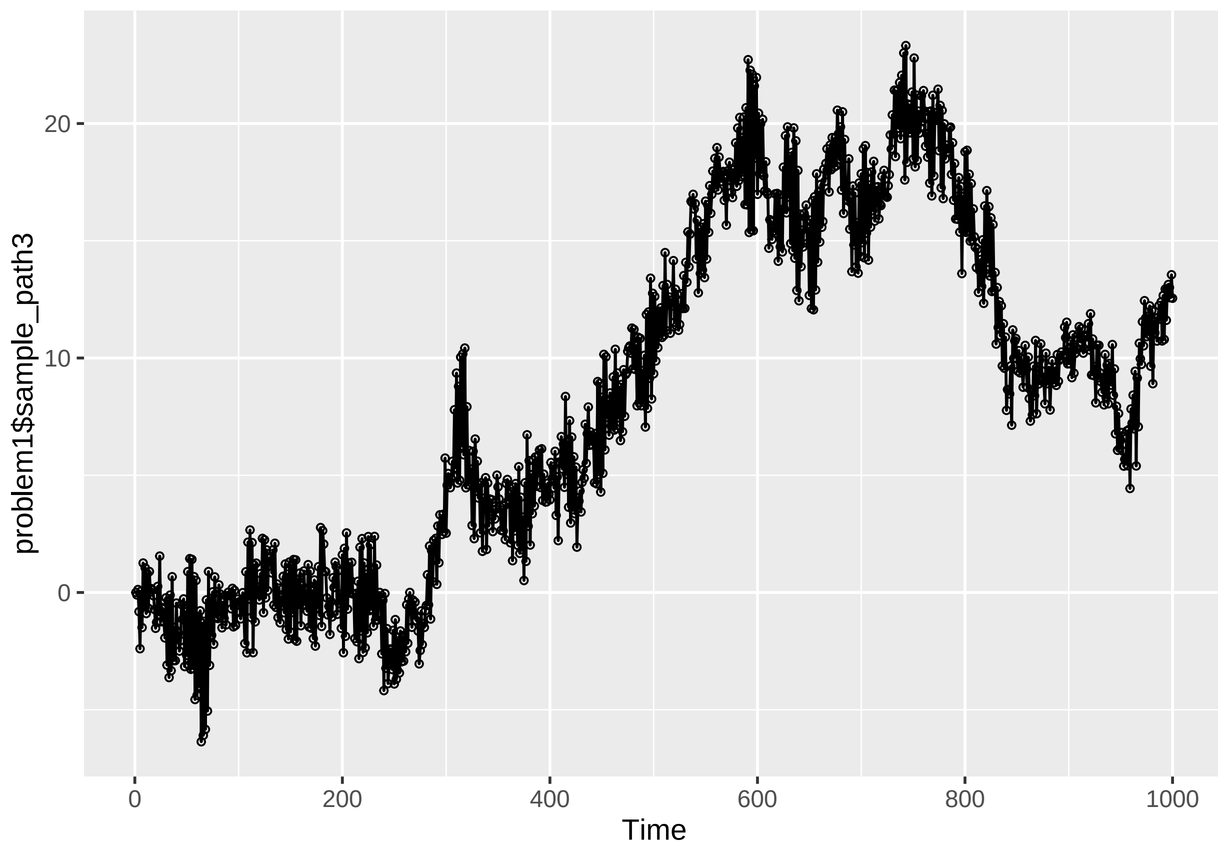
```r
# Please specify the estimated ARIMA(p,d,q) orders below
Q2_p <- 0
Q2_d <- 0
Q2_q <- 0
Q2_Sp <- 0 # [Optional] Seasonal AR() order
Q2_Sd <- 0 # [Optional] Seasonal I() order
Q2_Sq <- 0 # [Optional] Seasonal MA() order
```

```r
Q2_S <- NA # [Optional] seasonal period
```

Question 3

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

```r
# Please do your analysis below

autoplot(problem1$sample_path3) + geom_point(shape = 1, size = 1)
```




```r
adf.test(problem1$sample_path3,alternative = "stationary")
```

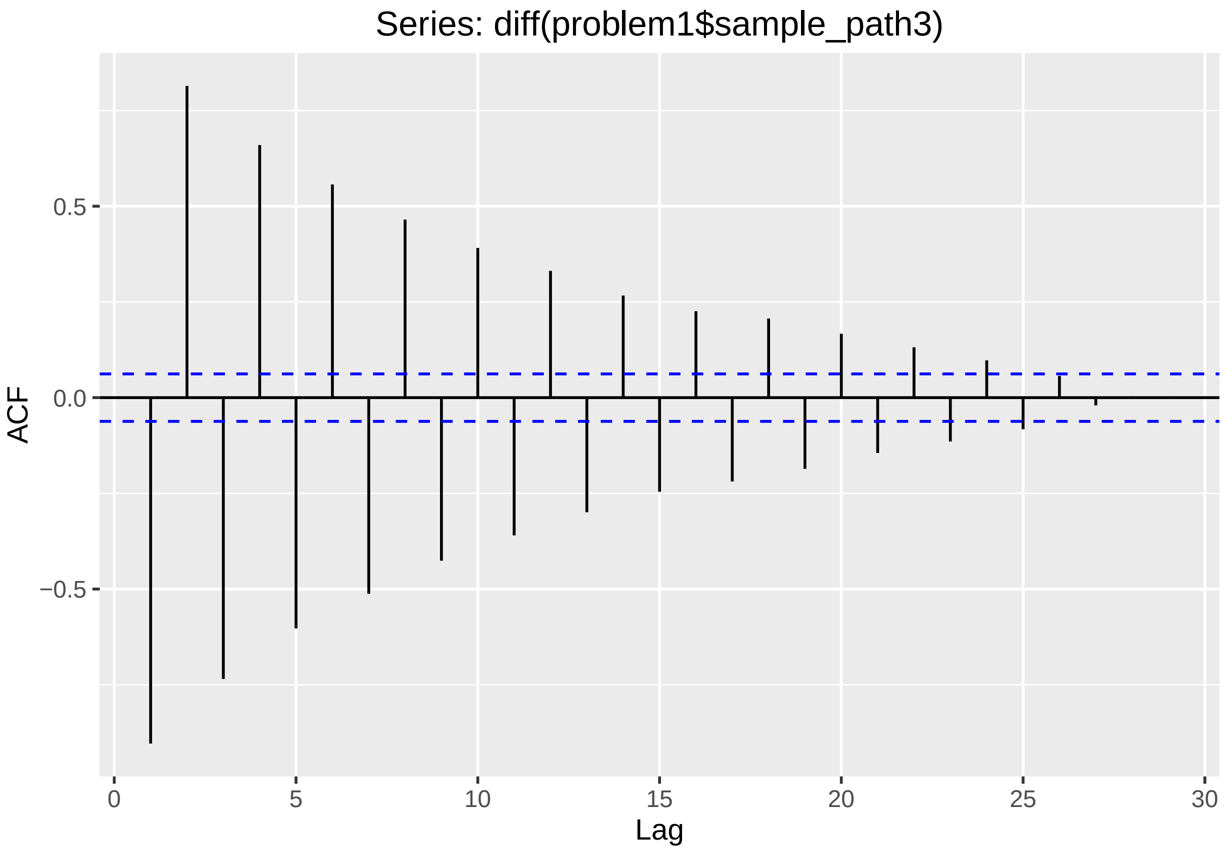
```
##
##  Augmented Dickey-Fuller Test
##
## data:  problem1$sample_path3
## Dickey-Fuller = -1.4452, Lag order = 9, p-value = 0.8131
## alternative hypothesis: stationary
```

```r
adf.test(diff(problem1$sample_path3),alternative = "stationary")
```

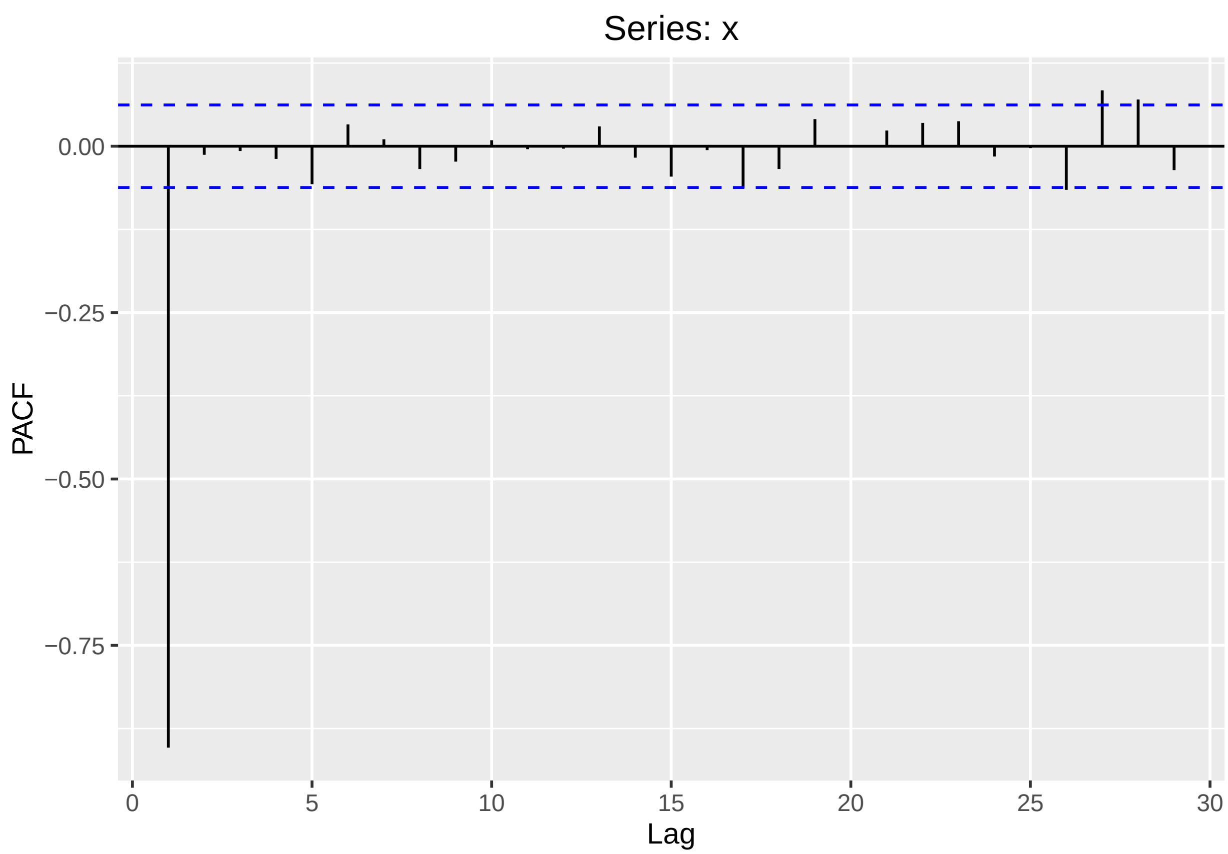
```
## Warning in adf.test(diff(problem1$sample_path3), alternative =
## "stationary"): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  diff(problem1$sample_path3)
## Dickey-Fuller = -10.854, Lag order = 9, p-value = 0.01
## alternative hypothesis: stationary
```

```r
ggAcf(diff(problem1$sample_path3))
```




```r
ggPacf(diff(problem1$sample_path3))
```




```r
eacf(diff(problem1$sample_path3))
```

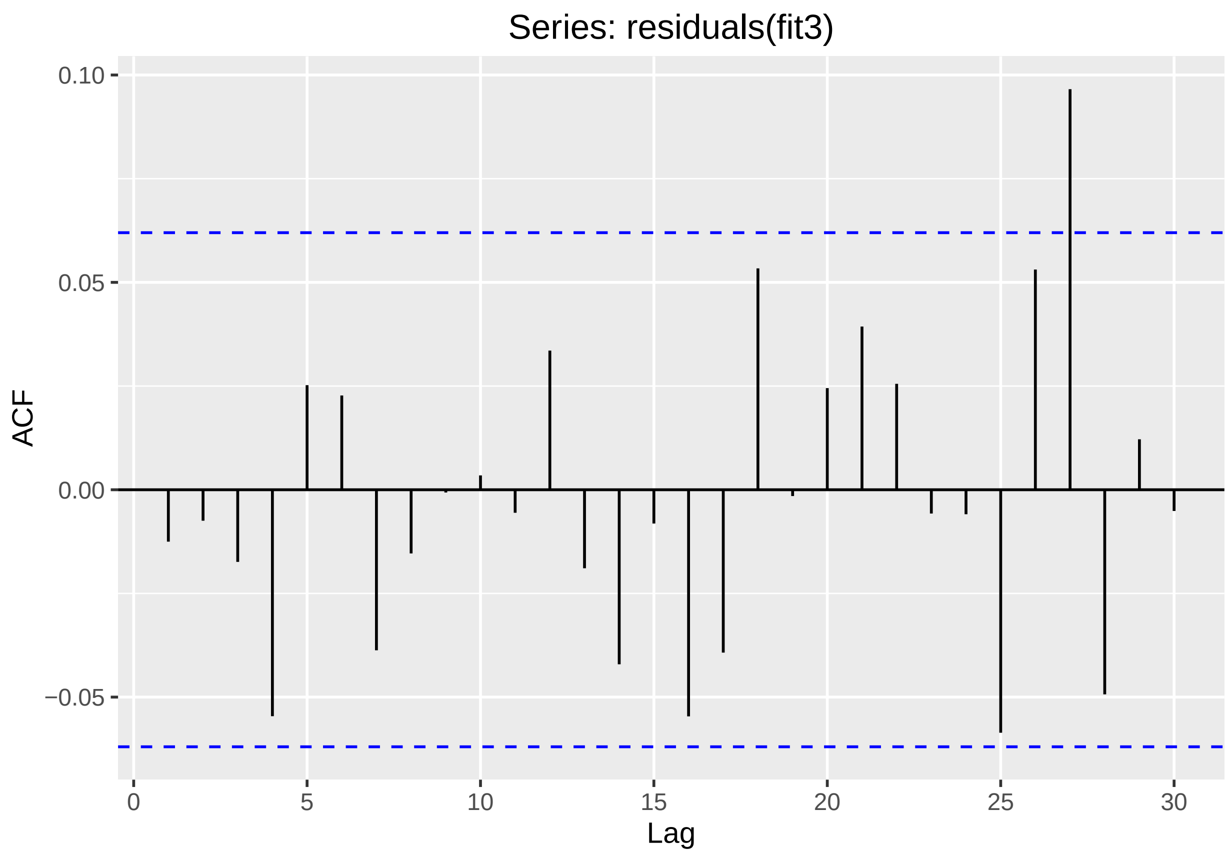
```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x x x x x x  x  x  x
## 1 o o o o o o o o o o o  o  o  o
## 2 x o o o o o o o o o o  o  o  o
## 3 x o o o o o o o o o o  o  o  o
## 4 x o x o o o o o o o o  o  o  o
## 5 x x x x o o o o o o o  o  o  o
## 6 x x x x o o o o o o o  o  o  o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

```r
fit3 <- Arima(problem1$sample_path3, order=c(1,1,0))
summary(fit3)
```

```
## Series: problem1$sample_path3
## 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
##
```

```
## Training set error measures:
##                       ME     RMSE       MAE     MPE     MAPE      MASE
## Training set 0.02484092 1.002725 0.8014718 338.383 408.3771 0.4203521
##                     ACF1
## Training set -0.01251398
```

```r
ggAcf(residuals(fit3))
```




```r
Box.test(residuals(fit3), type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit3)
## X-squared = 0.15707, df = 1, p-value = 0.6919
```

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

Please describe your reasoning below

The process seems to be nonstationary with no seasonality based on the plot. Using adf test, we can't reject that there is a random walk component since p-value is large, so the process is nonstationary. We first take the first order difference of y_t and check the stationality of Y_t-Y_t-1 using adf test. The result shows that after taking the first order difference, the process has become stationary. After plotting the autocorrelation function on first order difference, we can see that autocorrelations have been decreasing with time, which is an indicator for AR component. Partial autocorrelations indicates the order of AR process is 1 because only lag 1 sticks out. After checking extended auto-correlation, we notice when AR(1) process is taken away, no significant autocorrelation exists for MA process. So we think the process may be ARIMA(1,1,0). After fitting this sample path to ARIMA(1,1,0), we plot autocorrelation of residuals and also use Ljung-Box test to see whether autocorrelations of these residuals are different from zero. The p-value is larger than 0.05, so we can't reject that these residuals are independent, which means the residual times series is white noise. So the original process should be ARIMA(1,1,0).

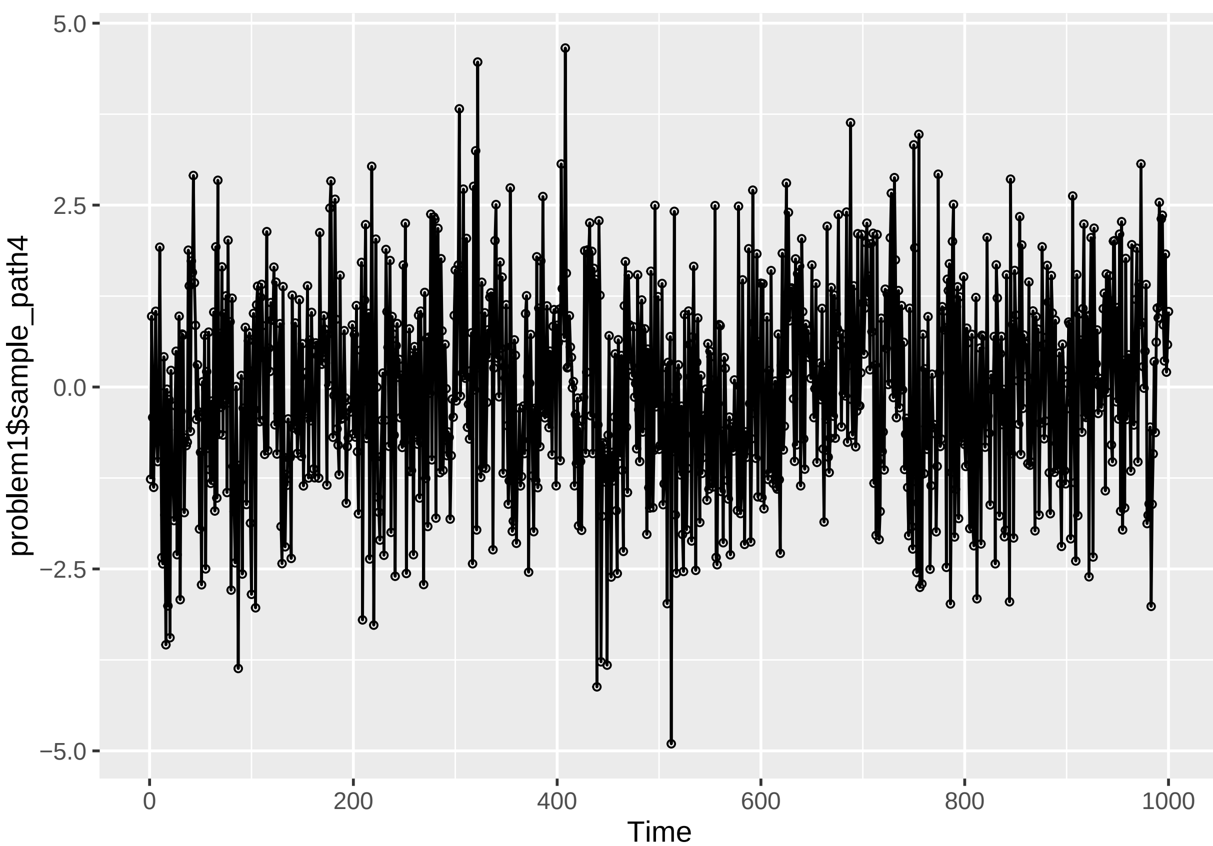
Please describe your reasoning above

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

Question 4

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

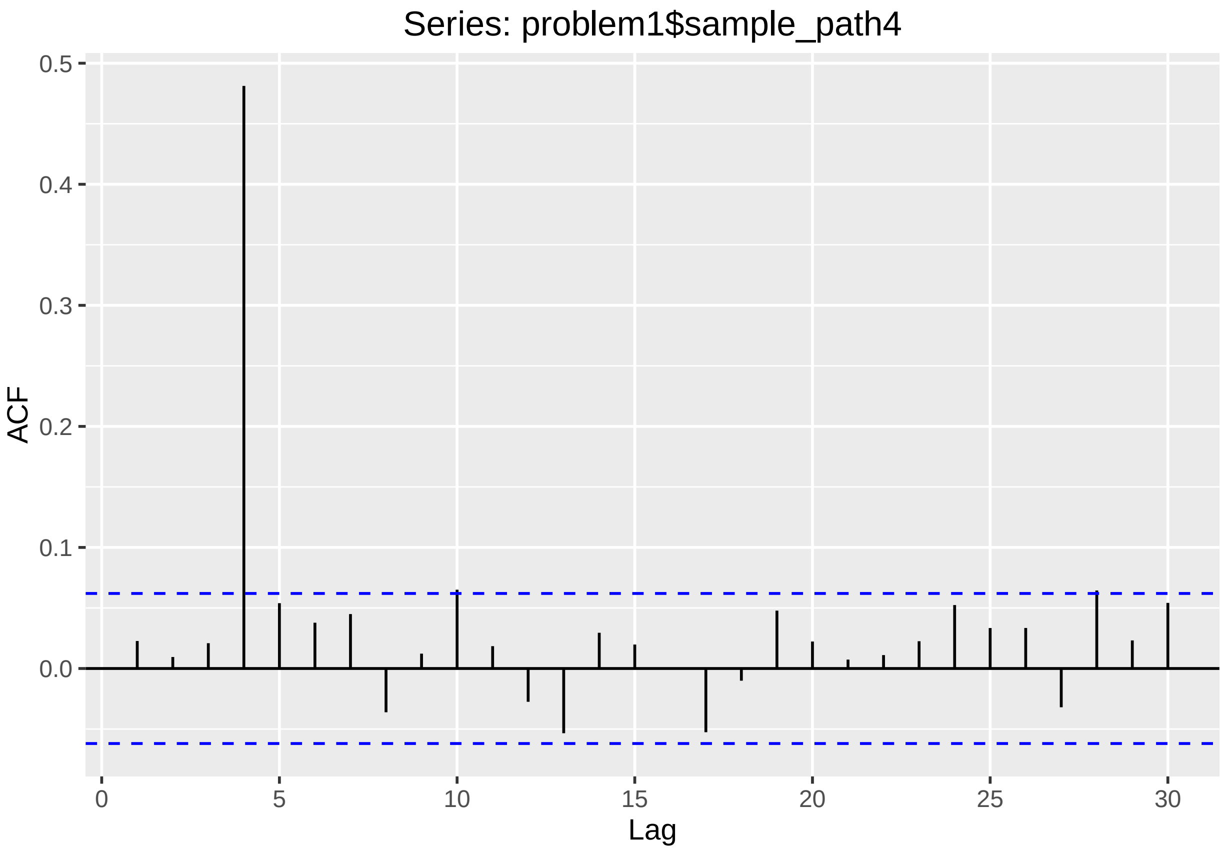
```
# Please do your analysis below

autoplot(problem1$sample_path4) + geom_point(shape = 1, size = 1)
```



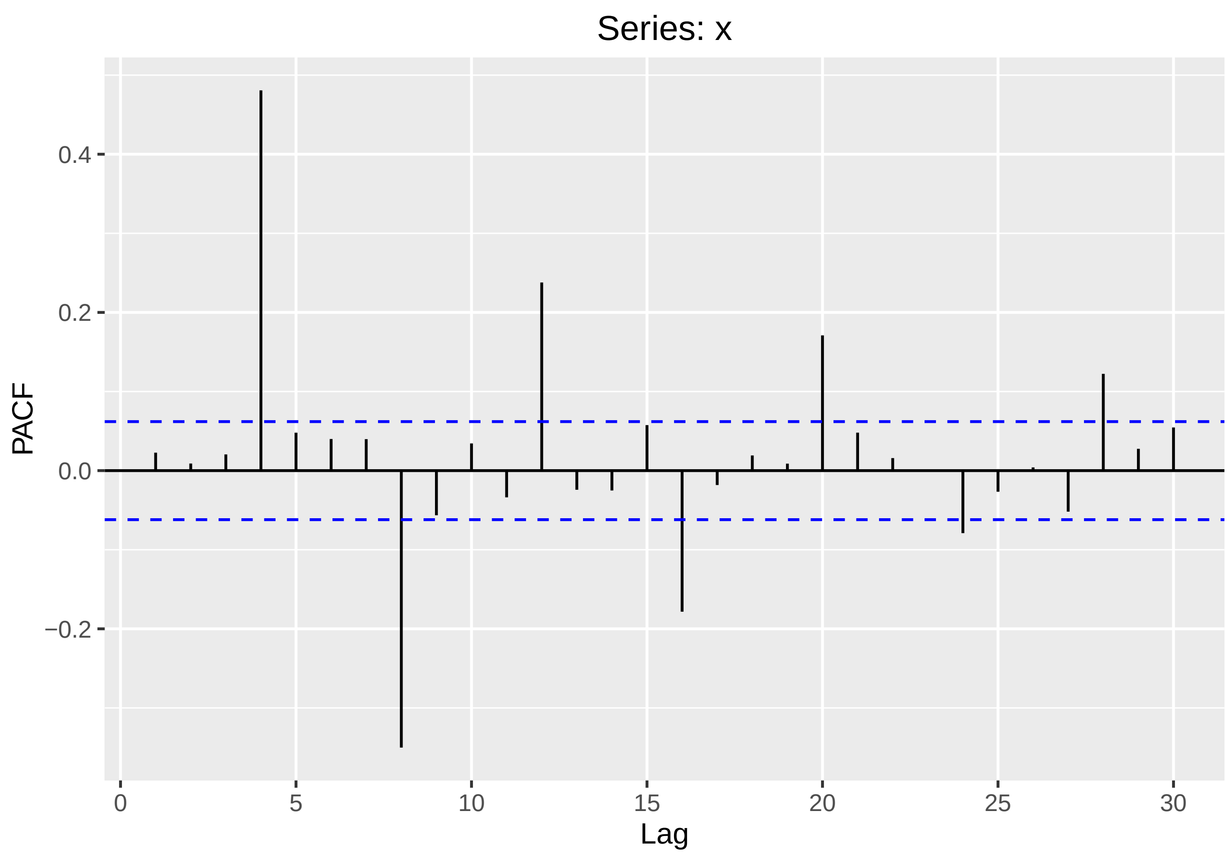

```
adf.test(problem1$sample_path4,alternative = "stationary")
```

```
## Warning in adf.test(problem1$sample_path4, alternative = "stationary"): 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
```

```
ggAcf(problem1$sample_path4)
```



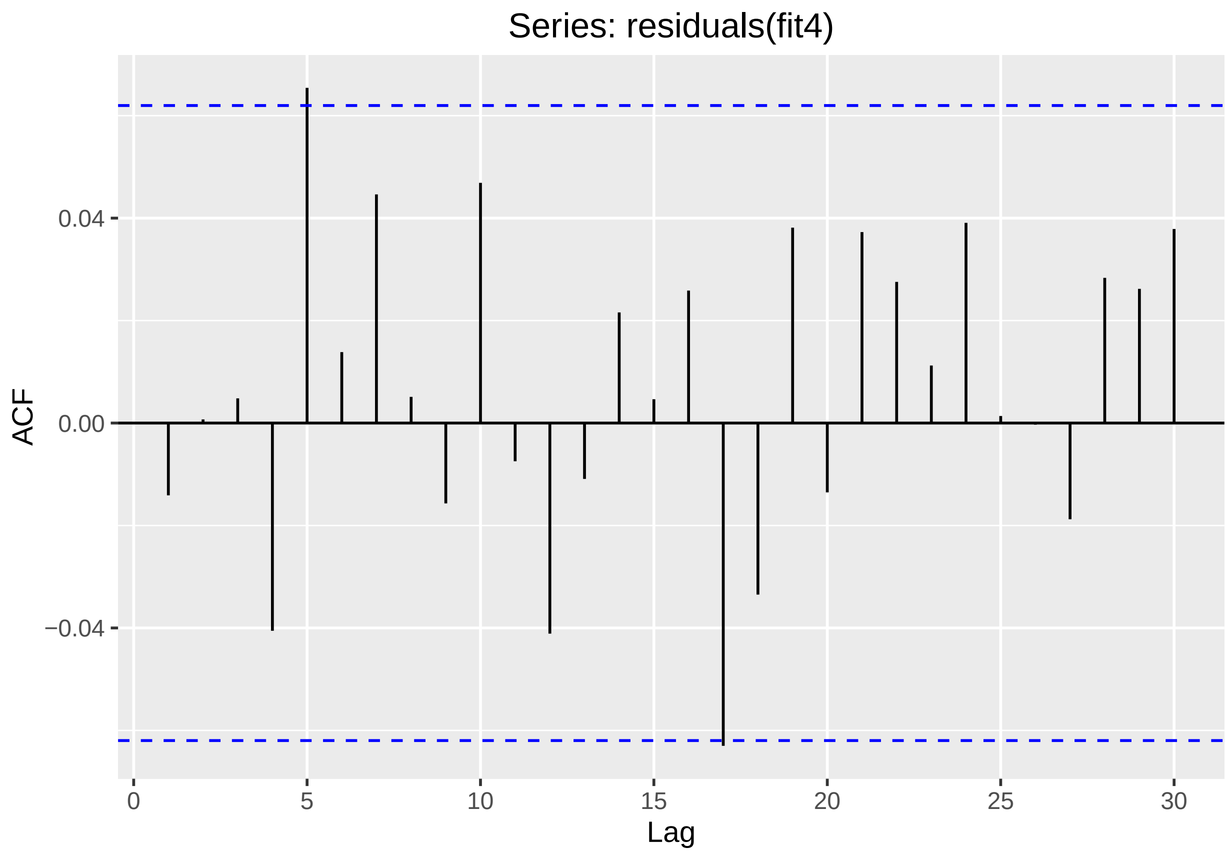

```
ggPacf(problem1$sample_path4)
```




```r
fit4 <- arima(problem1$sample_path4, order = c(0,0,0),
              seasonal = list(order = c(0,0,1), period = 4))

summary(fit4)
```

```
##
## Call:
## arima(x = problem1$sample_path4, order = c(0, 0, 0), seasonal = list(order = c(0,
##     0, 1), period = 4))
##
## Coefficients:
##         sma1  intercept
##       0.9908    -0.0340
## s.e.  0.0117     0.0621
##
## sigma^2 estimated as 0.9763:  log likelihood = -1414.94,  aic = 2833.88
##
## Training set error measures:

## Warning in trainingaccuracy(f, test, d, D): test elements must be within
## sample

##                 ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

```r
ggAcf(residuals(fit4))
```




```r
Box.test(residuals(fit4), type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  residuals(fit4)
## X-squared = 0.19982, df = 1, p-value = 0.6549
```

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

Please describe your reasoning below

Using adf test, we can tell there is no random walk component based on the small p-value. After plotting the autocorrelation function and partial autocorrelation function, we can see a spike at lag 4 in the ACF but no other significant spikes. The PACF shows exponential decay in the seasonal lags; that is, at lags 4, 8, 12, ..,so there is a seasonal MA(1) component with a period of 4. So we think the seasonal part of ARIMA may be (0,0,1), but we are not sure about the non-seasonal part. After fitting this sample path to ARIMA(0,0,0)(1,1,0)4, we plot autocorrelation of residuals and also use Ljung-Box test to see whether autocorrelations of these residuals are different from zero. The p-value is larger than 0.05, so we can't reject that these residuals are independent, which means the residual times series is white noise. So the seasonal part of ARIMA should be (0,0,1).

Please describe your reasoning above

```r
# 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
```