

Time Series Analysis - STAT 478 - HW #5

Alex Towell (atowell@siue.edu)

Contents

Problem 1	1
Preliminary analysis	1
Part (a)	2
Part (b)	2
Part (c)	2
Part (d)	3
Problem 2	5
Preliminary analysis	5
Part (a)	5
Part (b)	5
Part (c)	6
Part (d)	7
Problem 3	9
Preliminary steps	9
Part (a)	9
Part (b)	9
Part (c)	10
Additional experimentation	11
Problem 4	11
Preliminary steps	11
Part (a)	12
Part (b)	12
Part (c)	14
Additional experimentation	15

Problem 1

Consider the MA(2) process, where all the $\{e_t\}$ values are independent white noise with variance σ^2 .

$$Y_t = e_t - 0.5e_{t-1} - 0.3e_{t-2}$$

Preliminary analysis

In general, a MA(2) process has the form

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

where $\{e_t\}$ is i.i.d. WN with mean 0 and variance σ^2 .

We denote such a moving average process by MA(2; $\mu, \theta_1, \theta_2, \sigma$) which we model with the following R function:

```
MA2 <- function(mu, theta1, theta2, sigma=1)
{
  return(function(N)
  {
    et <- rnorm(n=N+2, mean=0, sd=sigma)
    yt <- mu +
      et[1:N] -
      theta1 * et[2:(N+1)] -
      theta2 * et[3:(N+2)]
    return(yt)
  })
}
```

This function takes three parameters, μ , θ_1 , and θ_2 , and optionally a fourth parameter σ , and returns an anonymous function that accepts a single parameter, n , which specifies the number of points to generate from the process.

Matching $e_t - 0.5e_{t-1} - 0.3e_{t-2}$ and $\mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$ term by term, we see that $\mu = 0$, $\theta_1 = 0.5$, and $\theta_2 = 0.3$, or in other words

$$\{Y_t\} \sim \text{MA2}(\mu = 0, \theta_1 = 0.5, \theta_2 = 0.3, \sigma).$$

Part (a)

Find $E(Y_t)$.

A moving average process of order 2 with $\mu = 0$ has an expectation of zero. However, just this once, we will manually derive the characteristic. By the linearity of expectation,

$$E(Y_t) = E(e_t) - 0.5 E(e_{t-1}) - 0.3 E(e_{t-2}) \quad (1)$$

$$= 0 - 0.5(0) - 0.3(0) = 0. \quad (2)$$

Part (b)

Find $\text{Cov}(Y_t, Y_t) = \text{Var}(Y_t)$.

Since $\{Y_t\} \sim \text{MA}(2; \mu = 0, \theta = 0.5, \theta_2 = 0.3, \sigma)$, the variance is given by

$$\text{Var}(Y_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2) = \sigma^2(1 + 0.5^2 + 0.3^2) = 1.34\sigma^2.$$

However, we may manually derive it using the computational variance formula,

$$\begin{aligned} \text{Var}(Y_t) &= E(Y_t^2) - E^2(Y_t) \\ &= E(e_t - 0.5e_{t-1} - 0.3e_{t-2})^2 \\ &= E\left(e_t^2 - e_t e_{t-1} - \frac{3}{5}e_t e_{t-2} + \frac{1}{4}e_{t-1}^2 + \frac{3}{10}e_{t-1}e_{t-2} + \frac{9}{100}e_{t-2}^2\right) \\ &= \sigma^2 - 0 - \frac{3}{5}0 + \frac{1}{4}\sigma^2 + \frac{3}{10}0 + \frac{9}{100}\sigma^2 \\ &= \sigma^2 \left(1 + \frac{1}{4} + \frac{9}{100}\right) \\ &= 1.34\sigma^2. \end{aligned}$$

Part (c)

Find $\gamma_k = \text{Cov}(Y_t, Y_{t+k})$. and, from this, find the ACF, ρ_k .

Since $\{Y_t\} \sim \text{MA}(2; \mu = 0, \theta = 0.5, \theta_2 = 0.3, \sigma)$, the covariance of Y_t and Y_{t+k} is given by

$$\text{Cov}(Y_t, Y_{t+k}) = \begin{cases} \text{Var}(Y_t) & k = 0, \\ \sigma^2(-\theta_1 + \theta_1\theta_2) & k = 1, \\ \sigma^2(-\theta_2) & k = 2, \\ 0 & k > 2. \end{cases} \quad (3)$$

We already derived the variance of Y_t and, since $\theta_1 = 0.5$ and $\theta_2 = 0.3$, we may rewrite the above as

$$\gamma_k = \begin{cases} 1.34\sigma^2 & k = 0, \\ -0.35\sigma^2 & k = 1, \\ -0.3\sigma^2 & k = 2, \\ 0 & k > 2 \end{cases} \quad (4)$$

and since $\rho_k = \gamma_k/\gamma_0$,

$$\rho_k = \begin{cases} 1 & k = 0, \\ -0.261 & k = 1, \\ -0.224 & k = 2, \\ 0 & k > 2. \end{cases} \quad (5)$$

Part (d)

Generate time series datasets of length $n = 200$ according to this MA(2) process. Plot the observed time series and the sample ACF and PACF. Do the plots agree with what you know to be true?

In code, we model $\{Y_t\} \sim \text{MA2}(\theta_1 = 0.5, \theta_2 = 0.3, \mu = 0)$ with:

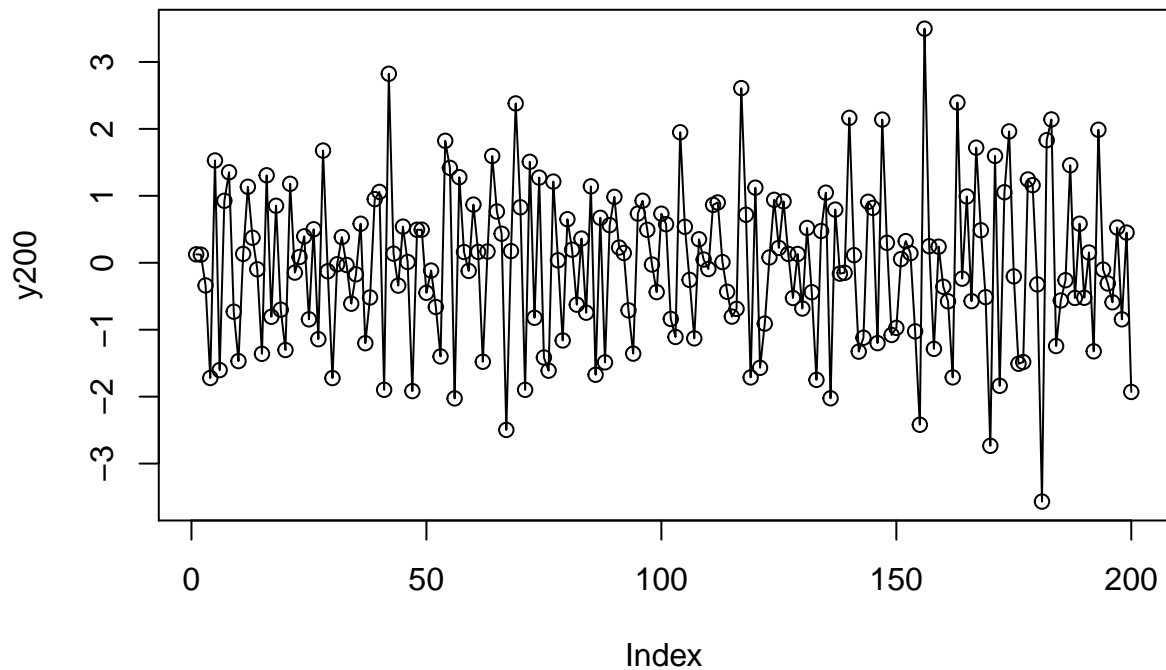
```
Yt <- MA2(theta1=.5, theta2=.3, mu=0)
```

We sample $n = 200$ values from $\{Y_t\}$ with

```
y200 <- Yt(N=200)
```

We plot the realization of the time series with the following R code:

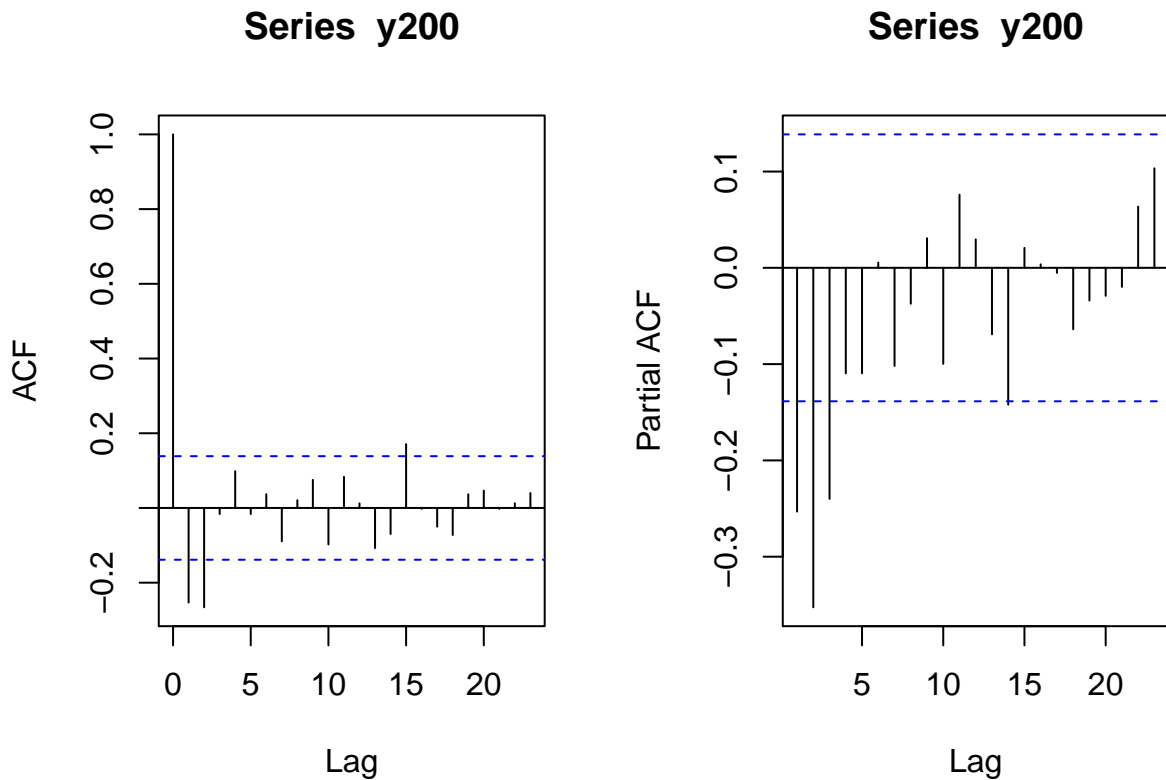
```
plot(y200, type="l")
points(y200)
```



The mean appears to be centered around 0 with a constant variance, as expected.

The sample ACF and PACF are given by:

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(y200)
pacf(y200)
```



As expected, since $\{Y_t\}$ models MA(2), the sample ACF cuts off after lag $k = 2$. The PACF is not as

informative, other than suggesting that $\{Y_t\}$ is not a good fit for an autoregressive model.

Problem 2

Consider the AR(1) process: $Y_t = \phi Y_{t-1} + e_t$, where all the $\{e_t\}$ values are independent white noise with variance σ^2 .

Preliminary analysis

In general, a AR(1) process is given by

$$Y_t = \Delta + \phi Y_{t-1} + e_t$$

where $\{e_t\}$ is i.i.d. WN with mean 0 and variance σ^2 .

We denote such an autoregressive process by $\text{AR}(1; \phi, \Delta, \sigma)$, which we model with the following R function:

```
AR1 <- function(phi, delta, sigma=1)
{
  return(function(N)
  {
    yt <- vector(length=N)
    yt[1] <- 0 #rnorm(n=1,mean=0,sd=sigma)
    for (i in 2:N)
    {
      et <- rnorm(n=1,mean=0,sd=sigma)
      yt[i] <- delta+phi*yt[i-1]+et
    }
    return(yt)
  })
}
```

This function takes three parameters, ϕ , Δ , and σ , and returns an anonymous function that accepts a single parameter, n , which specifies the number of points to generate from the process.

Observe that an autoregressive process $\text{AR}(1; \phi, \Delta, \sigma^2)$ has a mean $\mu = \Delta/(1 - \phi)$ and variance $\sigma_{Y_t}^2 = \sigma^2/(1 - \phi^2)$.

Part (a)

Show that if $|\phi| = 1$, the process cannot be stationary.

To be stationary, the variance must be constant. The variance of Y_t when $|\phi| = 1$ is given by

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(Y_{t-1}) + \sigma^2 \\ &= \text{Var}(Y_{t-2}) + 2\sigma^2 \\ &\vdots \\ &= \text{Var}(Y_{t-(t-2)}) + (t-2)\sigma^2 \\ &= \text{Var}(Y_{t-(t-1)}) + (t-1)\sigma^2 \\ &= t\sigma^2. \end{aligned}$$

Thus, if $|\phi| = 1$, Y_t has a variance of $t\sigma^2$, which is a function of time and is thus non-stationary.

Part (b)

Take $\phi = -0.6$, calculate find the ACF, ρ_k .

For a time series that models AR(1), the autocorrelation function ρ_k is given by ϕ^k . Thus, since $\{Y_t\}$ models AR(1; $\phi = -0.6, \Delta = 0, \sigma^2$), $\{Y_t\}$ has the autocorrelation function

$$\rho_k = (-0.6)^k$$

for $k = 0, 1, 2, \dots$

Part (c)

Take $\phi = -0.6$. Generate time series datasets of length $n = 200$ according to the AR(1) process. Plot the observed time series and the sample ACF and PACF. Do the plots agree with what you know to be true?

First, we know that $Y_t = \phi Y_{t-1} + e_t$. Matching this term by term to Matching $\Delta \phi Y_{t-1} + e_t$ shows that $\Delta = 0$ (and therefore $E(Y_t) = 0$).

In code, we model $\{Y_t\} \sim \text{AR1}(\phi = -0.6, \Delta = 0, \sigma = 1)$ with:

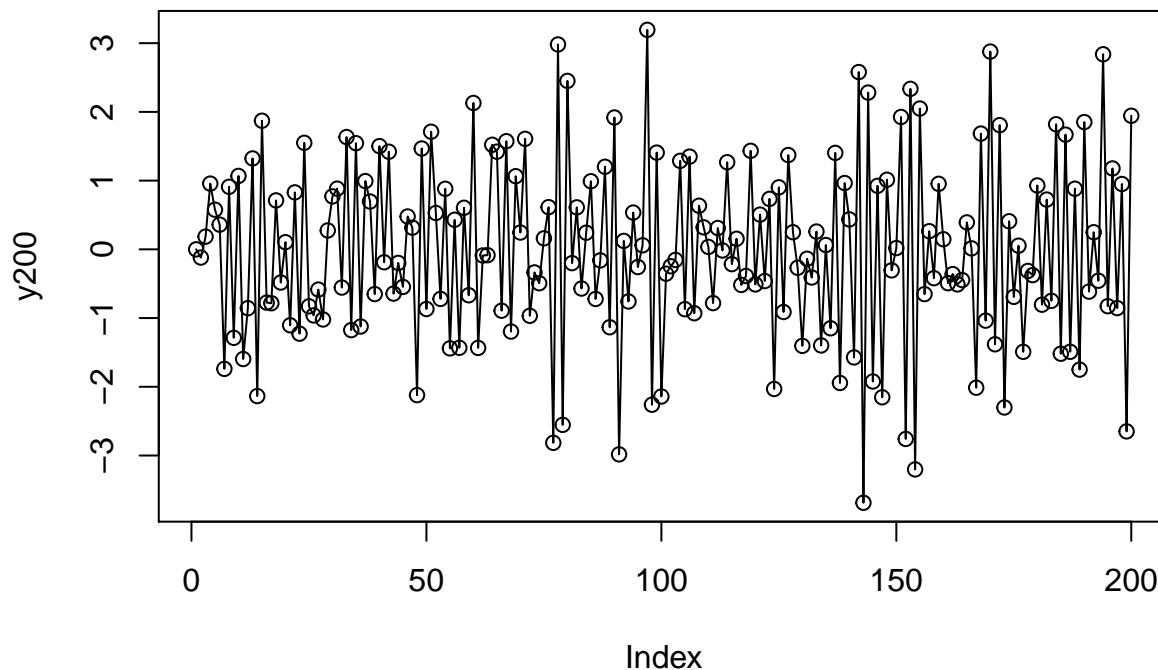
```
Yt <- AR1(phi=-0.6, delta=0, sigma=1)
```

We sample $n = 200$ values from $\{Y_t\}$ with:

```
y200 <- Yt(N=200)
```

We plot the realization of the time series with the following R code:

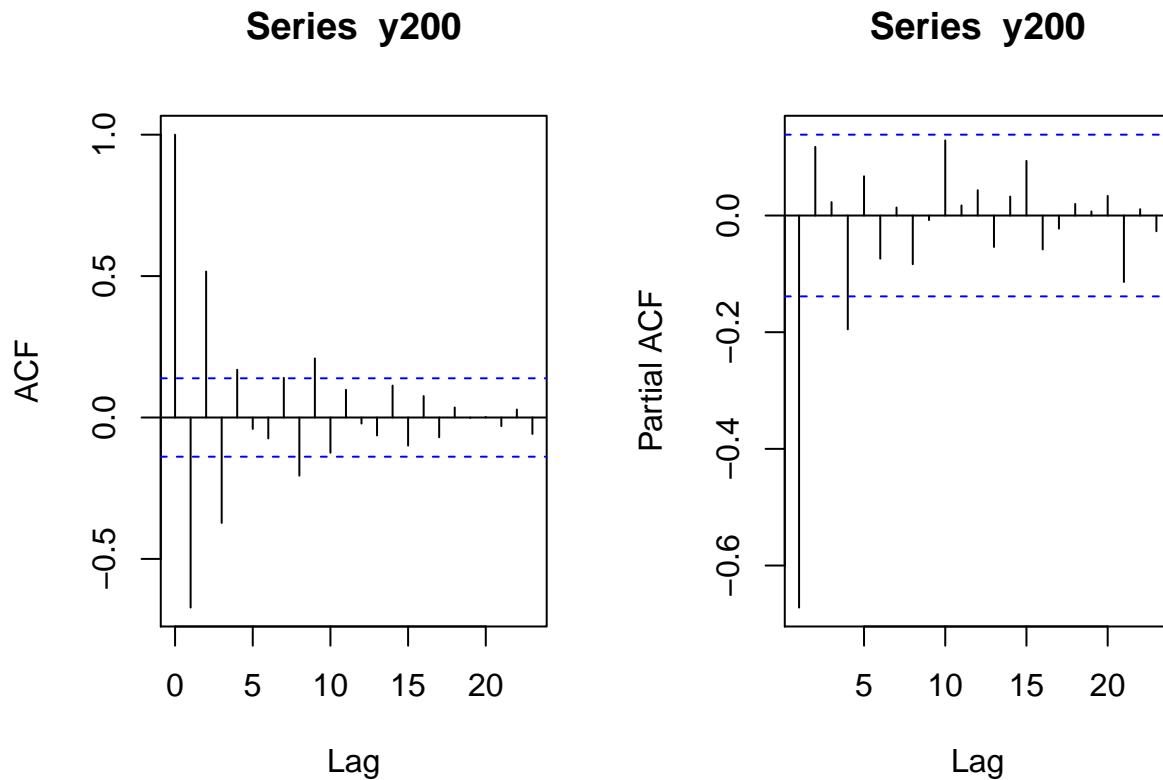
```
plot(y200, type="l")
points(y200)
```



The mean appears to be centered around 0 with a constant variance as expected. Also, since ϕ is negative, Y_t and Y_{t+1} are negatively correlated and thus exhibits jittery up and down movements as expected.

The sample ACF and PACF are given by:

```
par(mfrow=c(1,2), oma=c(0,0,0,0))
acf(y200)
pacf(y200)
```



The sample PACF cuts off (with a single outlier) after lag 1, as expected of a AR(1) process.

Part (d)

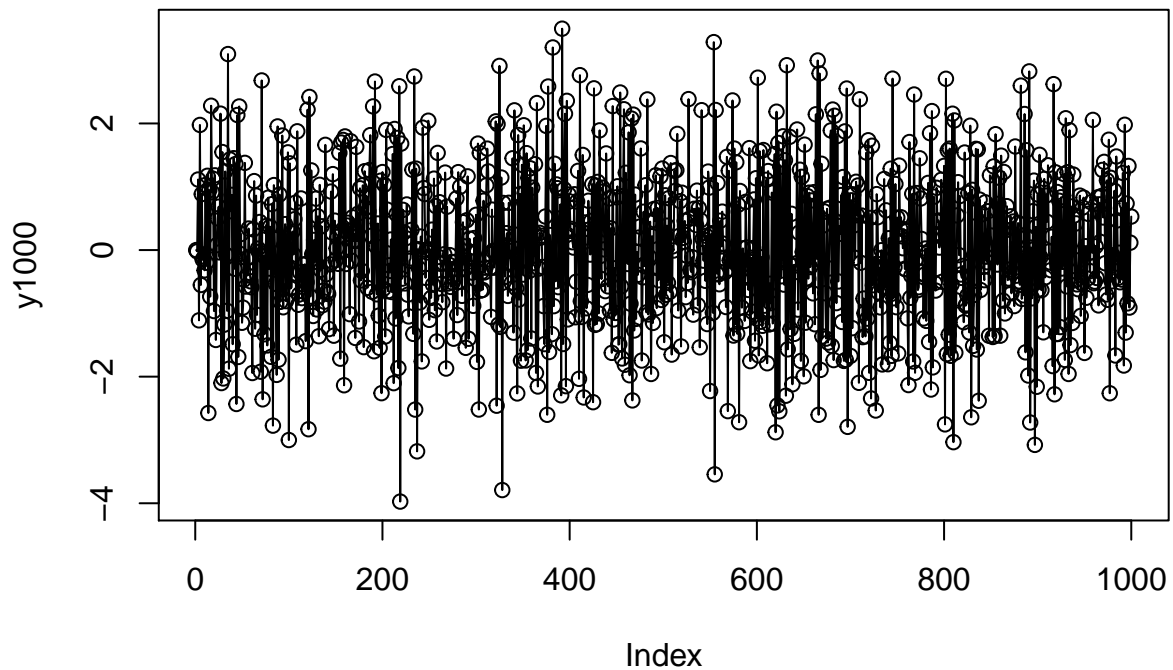
What happens when we increase the sample size? Repeat part (c) when $n = 1000$. Comment on your findings.

We sample $n = 1000$ values from $\{Y_t\}$ with

```
y1000 <- Yt(N=1000)
```

We plot the realization of the time series with the following R code:

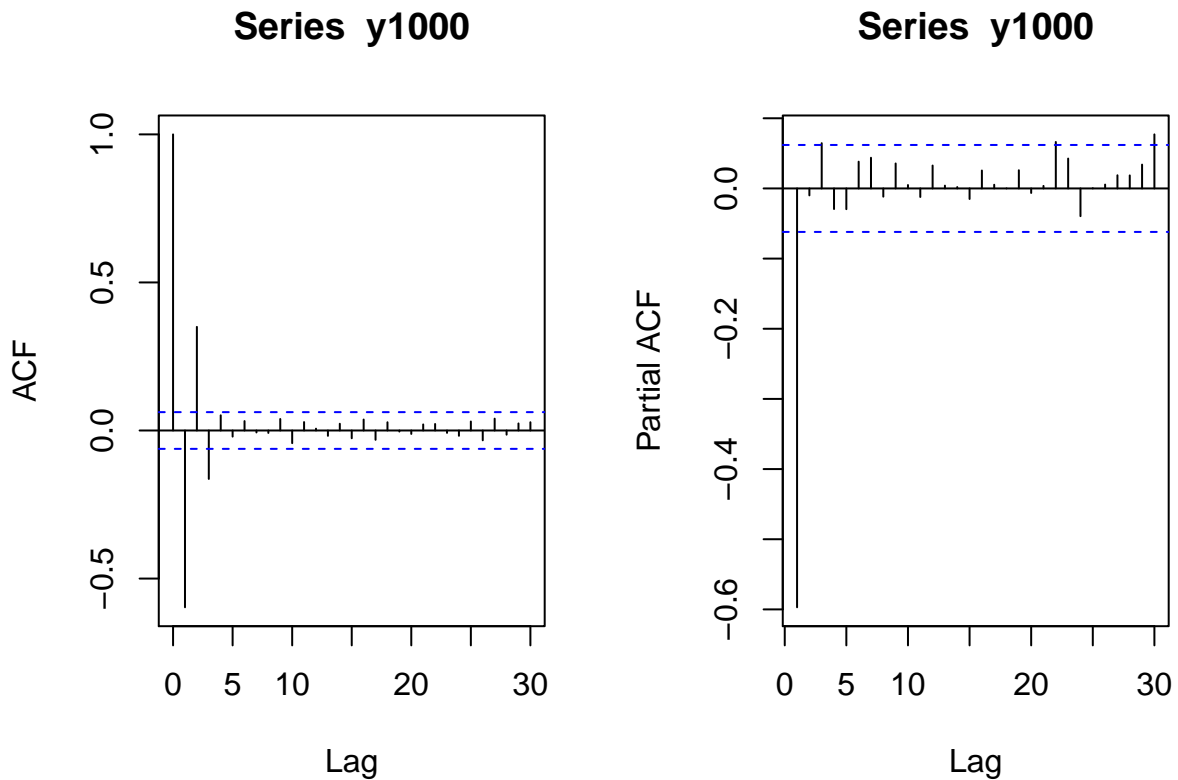
```
plot(y1000,type="l")
points(y1000)
```



The mean appears to be centered around 0 with a constant variance, as expected.

The sample ACF and PACF are given by:

```
par(mfrow=c(1,2), oma=c(0,0,0,0))
acf(y1000)
pacf(y1000)
```



We see the same general pattern as before in each case, except due to the increased sample size the confidence

intervals have narrowed.

Problem 3

A data set of 57 consecutive measurements from a machine tool are in the *deere3* object in the TSA package.

Preliminary steps

We load the requisite time series data with the following R code:

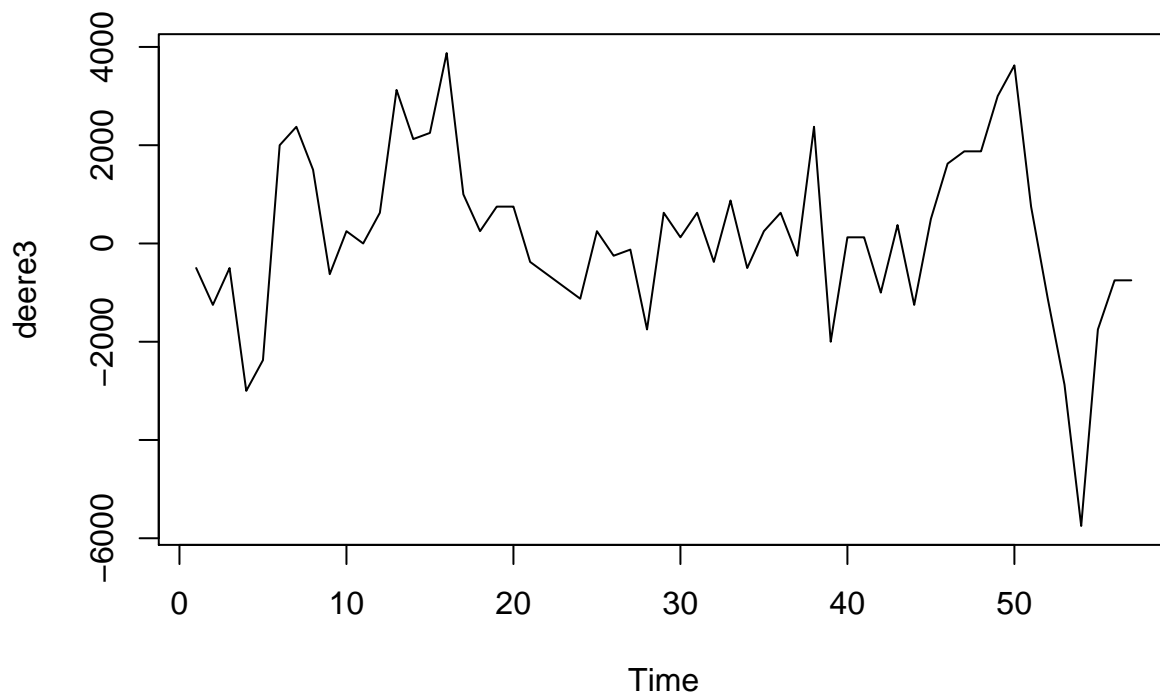
```
library(TSA)
data(deere3)
```

Part (a)

Plot the time series. What basic pattern do you see from the plot? Might a stationary model be appropriate for this plot?

We plot the time series with the following R code:

```
plot(deere3)
```

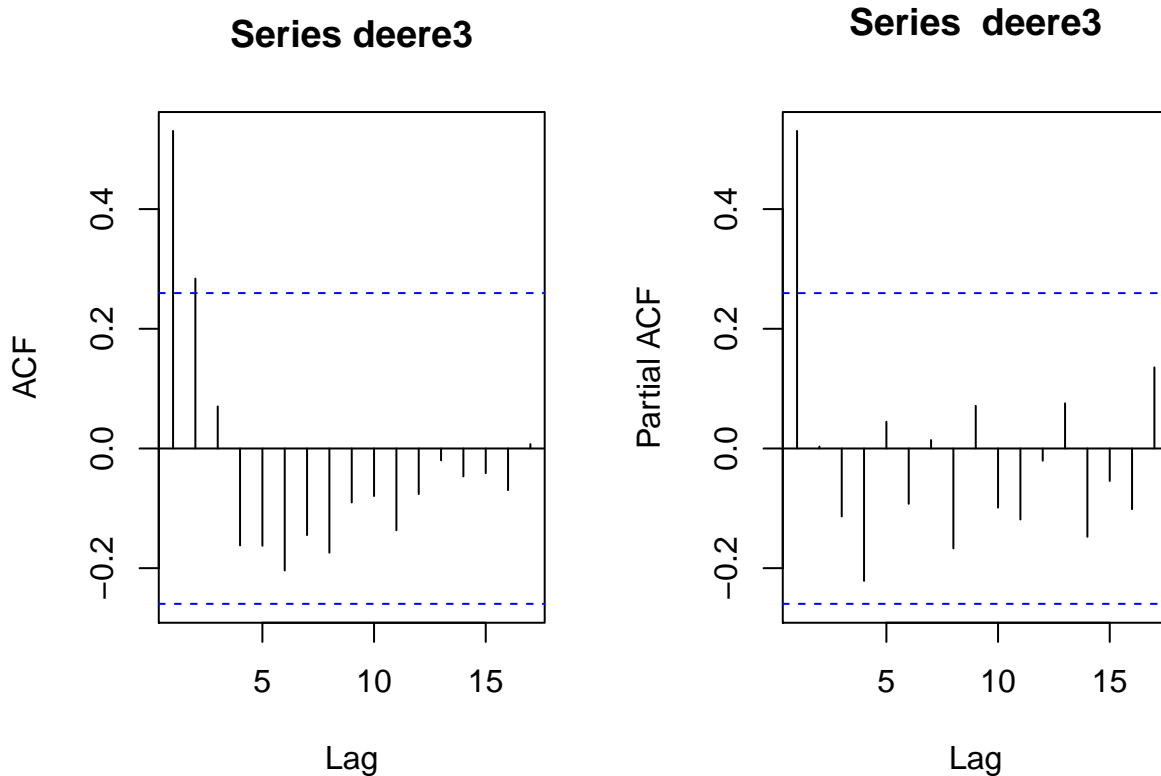


A stationary model may be appropriate since it fluctuates randomly around some central tendency and the fluctuations seem relatively constant (albeit quite large).

Part (b)

Plot the sample ACF and PACF. Tentatively specify the type of model (AR, MA, or ARMA) as well as the order(s) of the model. Write up detailed notes that describe how you decided on the model.

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(deere3)
pacf(deere3)
```



We seek a parsimonious model that sufficiently explains the data. Tentatively, we choose a AR(1) model.

The sample ACF decays and the sample PACF drops off to values compatible with 0 after lag $k = 1$. Thus, the sample ACF and PACF are compatible with a AR(1) process. Due to its parsimony and compatibility with the data, we choose AR(1).

Part (c)

Fit an AR(1) model using arima function in R and use it to forecast the next ten values of the series, and provides the forecasted values.

We seek to fit the time series to a first-order autoregressive model of the form

$$Y_t = \delta + \phi Y_{t-1} + e_t$$

where e_t is white noise with mean 0 and variance σ^2 and $\delta = \mu(1 - \phi)$.

We perform the fit using the following R code:

```
ar1 <- arima(deere3,order=c(1,0,0)) # fit an AR(1) model
ar1

##
## Call:
## arima(x = deere3, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
```

```
##      0.5255   124.3832
## s.e. 0.1108   394.2066
##
## sigma^2 estimated as 2069355: log likelihood = -495.51, aic = 995.02
```

We see that $\hat{\mu} = 124.3832$, $\hat{\phi} = 0.5255$, $\hat{\sigma}^2 = 2069355$, and $\hat{\delta} = \hat{\mu}(1 - \hat{\phi}) = 59.2$. Thus, we estimate that

$$\hat{Y}_t = 59.02 + 0.5255\hat{Y}_{t-1} + e_t.$$

We perform a 10-step ahead forecast with the following R code:

```
predict(ar1,n.ahead=10)$pred
```

```
## Time Series:
## Start = 58
## End = 67
## Frequency = 1
## [1] -335.145917 -117.120758 -2.538374  57.680010  89.327578 105.959850
## [7]  114.700885  119.294706 121.708973 122.977783
```

For small look-aheads, since nearby values in the time series are correlated, nearby previous values should have some measurable effect on the forecast. However, we expect that, as the look-ahead time for the forecast goes to infinity, the forecast converges to $\hat{\mu} = 124.3832$.

Additional experimentation

For fun, we construct a generative autoregressive model of order 1 with:

```
Xt <- AR1(phi=0.5255, delta=59.0198284, sigma=sqrt(2069355))
```

If we sample a large number of points from this generative model and then fit a AR(1) model to it, we get estimates similar to before:

```
arma(Xt(N=1000),order=c(1,0,0))
```

```
##
## Call:
## arima(x = Xt(N = 1000), order = c(1, 0, 0))
##
## Coefficients:
##      ar1 intercept
##      0.5197  202.0152
## s.e. 0.0270   93.6336
##
## sigma^2 estimated as 2026916: log likelihood = -8680.11, aic = 17364.22
```

Problem 4

A data set of durations until payment for 130 consecutive orders from a Winegrad distributor are in the days object in the TSA package.

Preliminary steps

We load the data with the following R code:

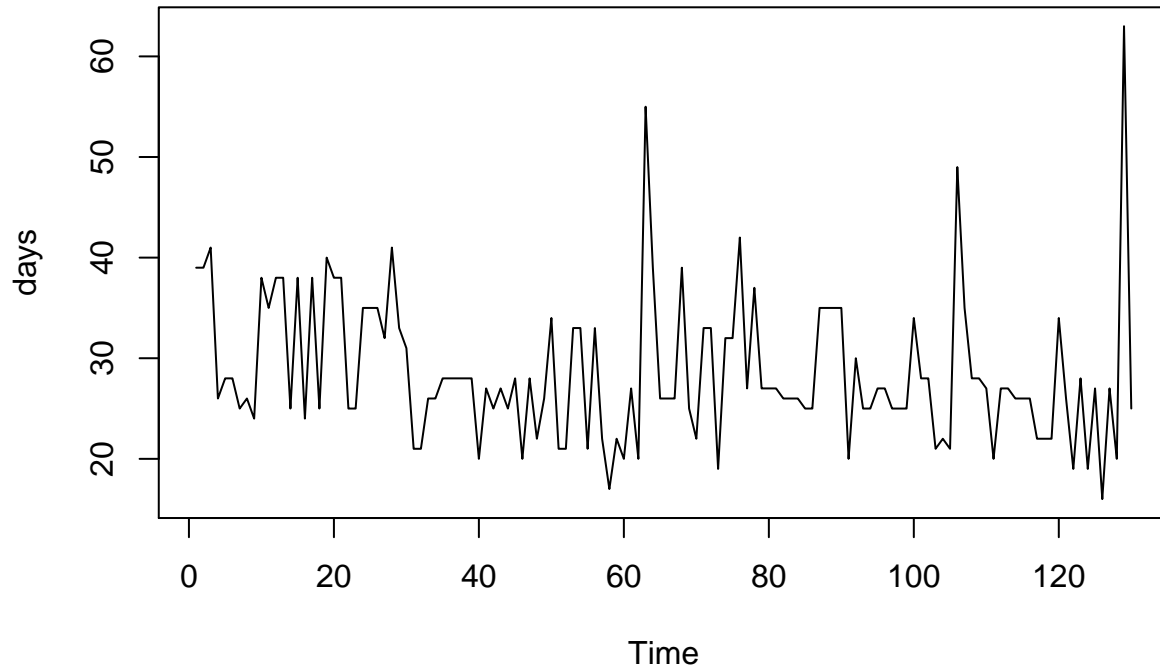
```
library(TSA)
data(days)
```

Part (a)

Plot the time series. What basic pattern do you see from the plot? Might a stationary model be appropriate for this plot?

We plot the time series with the following R code:

```
plot(days)
```



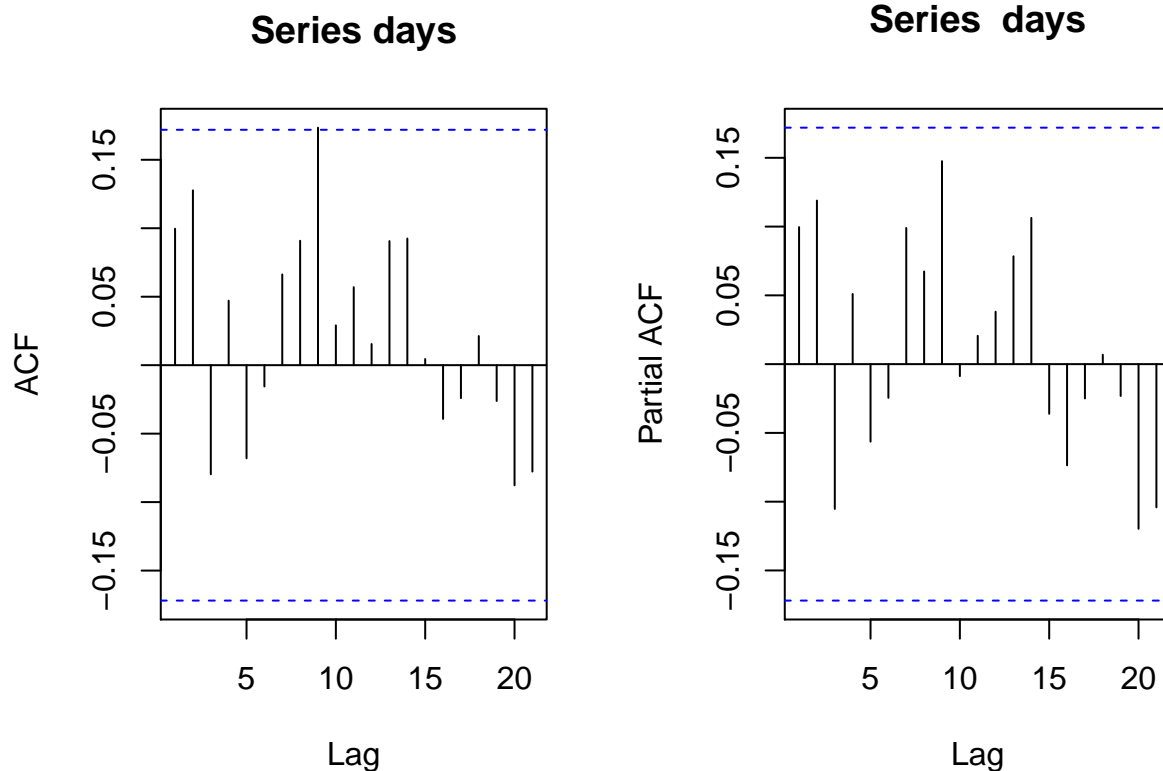
A stationary model may be appropriate since it fluctuates randomly around some central tendency and the fluctuations seem relatively constant (albeit quite large).

Part (b)

Plot the sample ACF and PACF. Tentatively specify the type of model (AR, MA, or ARMA) as well as the order(s) of the model. Write up detailed notes that describe how you decided on the model.

The sample ACF and PACF are given by:

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(days)
pacf(days)
```



There does not seem to be much of a pattern to the data. The ACF and PACF suggest that the data is compatible with the hypothesis that the values in the time series are linearly uncorrelated.

We choose a model that is a white noise process with a non-zero mean. We might think of this as a degenerate case of the ARMA model denoted by $\text{ARMA}(0, 0)$.

Model selection

In what follows, we analyze model selection for this time series in more depth. Given a set of candidate models that may hopefully sufficiently explain the data, one strategy of model selection is to choose the model with the lowest Akaike information criterion (AIC) on the given data set.

In particular, suppose the set of candidate models M is given by

$$M = \{\text{ARMA}(p, q) \mid p \in P, q \in Q\}$$

where P and Q are subsets of the natural numbers and we let the selected model m^* be defined as

$$m^* = \arg \min_{m \in M} \text{AIC}(m).$$

If M is a relatively small set, we can simply perform a brute-force exhaustive search through. We let $P = \{0, 1, 2, 3\}$ and $Q = \{0, 1, 2, 3\}$ and thus $|M| = 16$. We perform the exhaustive search with the following R code:

```
P <- c(0,1,2,3)
Q <- c(0,1,2,3)
aics <- matrix(nrow=length(P)*length(Q), ncol=3)
colnames(aics) <- c("AIC", "p", "q")
i <- 1
for (p in P)
{
```

```

for (q in Q)
{
  aics[i,1] <- arima(days,order=c(p,0,q))$aic
  aics[i,2] <- p
  aics[i,3] <- q
  i <- i + 1
}
}
aics <- aics[order(aics[,1],decreasing=FALSE),]
aics

```

```

##           AIC p q
## [1,] 886.2239 1 2
## [2,] 886.9111 0 0
## [3,] 886.9546 0 2
## [4,] 887.0030 2 1
## [5,] 887.2617 3 0
## [6,] 887.3115 2 0
## [7,] 887.6010 1 0
## [8,] 887.8065 0 3
## [9,] 887.8736 0 1
## [10,] 887.9656 2 2
## [11,] 888.5414 3 1
## [12,] 888.8066 3 2
## [13,] 888.9095 1 1
## [14,] 890.0033 2 3
## [15,] 890.8047 3 3
## [16,] 890.8777 1 3

```

```

cat("m* = ARIMA(",aics[1,2],",",aics[1,3],")\n")

```

```

## m* = ARIMA( 1 , 2 )

```

We see that $m^* = \text{ARMA}(1, 2)$. However, the next best model according to AIC on M is $\text{ARMA}(0, 0)$. The *third* best model is $\text{ARMA}(0, 2)$, which is identical to $\text{MA}(2)$.

Given the simplicity of white noise (with a non-zero mean), as an ad hoc decision I remain inclined to accept $\text{ARMA}(0, 0)$ as the most likely model for the data. In other words, my *prior* on M more heavily weighs the simpler models with smaller p and q than simply choosing the minimum AIC on M .

Part (c)

Fit an $\text{MA}(2)$ model using `arima` function in R and use it to forecast the next ten values of the series, and list the forecasted values.

We seek to fit the time series to a second-order moving average model of the form

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

where e_t is white noise with mean 0 and variance σ^2 .

We perform the fit using the following R code:

```

ma2 <- arima(days,order=c(0,0,2))
ma2

```

```

##
## Call:

```

```
## arima(x = days, order = c(0, 0, 2))
##
## Coefficients:
##          ma1      ma2  intercept
##          0.1113  0.1557   28.6931
## s.e.    0.0894  0.0884    0.7946
##
## sigma^2 estimated as 51.33:  log likelihood = -440.48,  aic = 886.95
```

We see that $\hat{\mu} = 28.6931$, $\hat{\theta}_1 = 0.1113$, and $\hat{\theta}_2 = 0.1557$. Thus, we estimate that

$$\hat{Y}_t = 28.6931 + e_t - 0.1113e_{t-1} - 0.1557e_{t-2}.$$

We perform a 10-step ahead forecast with the following R code:

```
predict(ma2,n.ahead=10)$pred
```

```
## Time Series:
## Start = 131
## End = 140
## Frequency = 1
## [1] 33.43453 27.67666 28.69310 28.69310 28.69310 28.69310 28.69310 28.69310
## [9] 28.69310 28.69310
```

We see that the first two forecasts are a function of the last two observations. After that, the forecast is $\hat{\mu} = 28.6931$, which is consistent with the assumption of a constant mean stationary process.

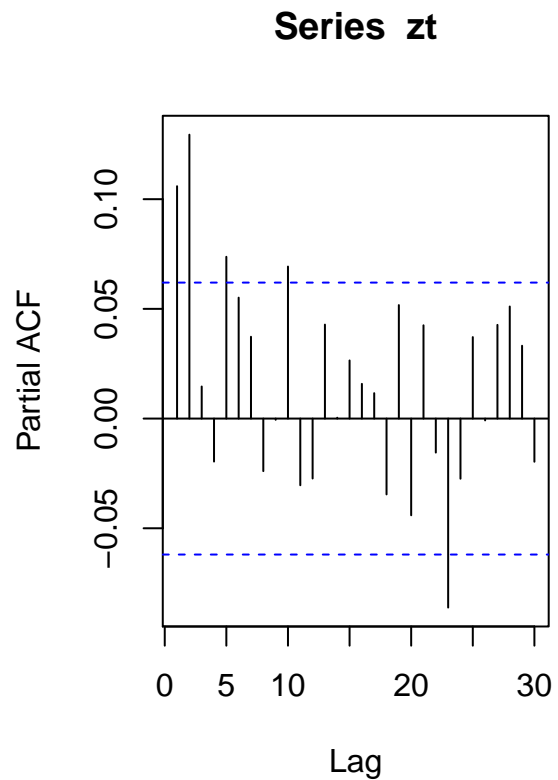
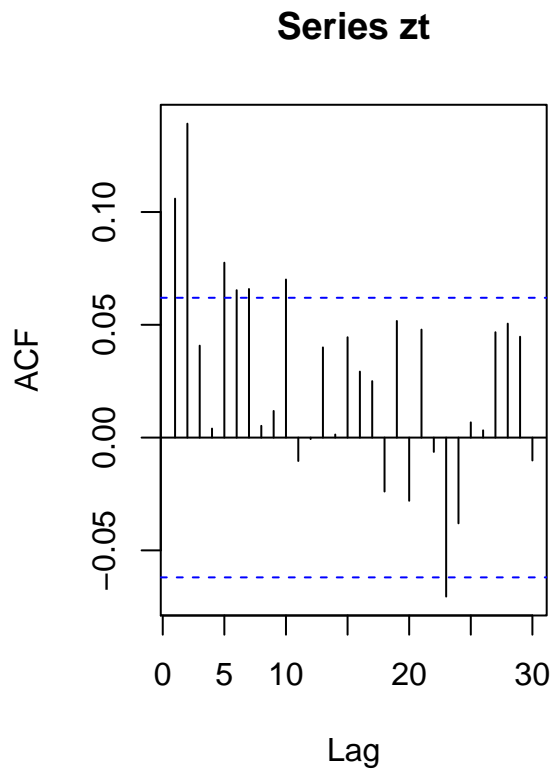
Additional experimentation

For fun, we construct a generative moving average model of order 2 with:

```
Zt <- MA2(mu=28.6931,theta1=-0.1113,theta2=-0.1557,sigma=51.33)
```

If we sample a large number of points from this generative model and then fit a MA(2) model to it, we get estimates similar to before:

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
zt <- Zt(N=1000)
acf(zt)
pacf(zt)
```



```
arima(zt,order=c(0,0,2))
```

```
##
## Call:
## arima(x = zt, order = c(0, 0, 2))
##
## Coefficients:
##      ma1      ma2  intercept
##    0.0874  0.1431    32.1212
## s.e.  0.0311  0.0324     1.8686
##
## sigma^2 estimated as 2307:  log likelihood = -5290.87,  aic = 10587.75
```