[STAT 4540] HW-3

Michael Ghattas

3/2/2022

Problem 1

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```r
library(gridExtra)
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
df <- load("/Users/Home/Documents/Michael_Ghattas/School/CU_Boulder/2022/Spring 2022/STAT - 4540/HW/3/CO-GHCND-TN
-TX.RData");
cityIndex = 125;
dat <- data.frame(mo = mo[yr %in% c(1900:2010)], da = da[yr %in% c(1900:2010)], yr = yr[yr %in% c(1900:2010)], te
mp = TN[yr %in% c(1900:2010), cityIndex]);
agg = aggregate(temp ~ mo + yr, dat, mean);
```

(a)

```r
cos.c <- cos(2 * pi * agg$mo / 12);
sin.c <- sin(2 * pi *  agg$mo / 12);
fit <- lm(temp ~  poly(1:length(agg$temp), degree = 1) + cos.c + sin.c, data = agg);

summary(fit);
```

```
##
## Call:
## lm(formula = temp ~ poly(1:length(agg$temp), degree = 1) + cos.c +
##     sin.c, data = agg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.0399 -1.0427  0.0435  1.1312  5.2928
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            1.18015    0.05040   23.41   <2e-16 ***
## poly(1:length(agg$temp), degree = 1) 34.68043    1.83828   18.87   <2e-16 ***
## cos.c                                 -9.81995    0.07130 -137.73   <2e-16 ***
## sin.c                                 -6.07307    0.07127  -85.21   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.838 on 1326 degrees of freedom
## Multiple R-squared:  0.9525, Adjusted R-squared:  0.9524
## F-statistic:  8872 on 3 and 1326 DF,  p-value: < 2.2e-16
```

```r
coef(fit);
```

```
##                          (Intercept) poly(1:length(agg$temp), degree = 1)
##                             1.180152                            34.680432
##                                cos.c                                sin.c
##                            -9.819951                            -6.073068
```

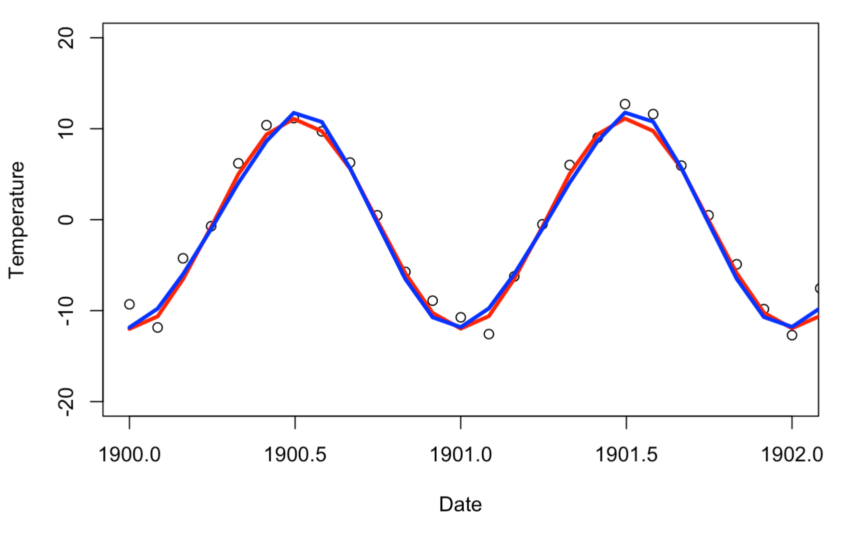
(b)

```r
m <- vector(mode = "list", length = 12);
for (i in 1:12 )
{
  m[[i]] = ifelse( agg$mo == i, 1, 0);
}
agg$date <- as.Date(make_datetime(year = agg$yr, month = agg$mo));
agg$Date <- decimal_date(agg$date);

lmod <- lm(temp ~ Date + m[[1]] + m[[2]] + m[[3]] + m [[4]] + m[[5]] + m[[6]] + m[[7]] + m[[8]] + m[[9]] + m[[10]
] + m[[11]] + m[[12]], data = agg);
summary(lmod)
```

```
##
## Call:
## lm(formula = temp ~ Date + m[[1]] + m[[2]] + m[[3]] + m[[4]] +
##     m[[5]] + m[[6]] + m[[7]] + m[[8]] + m[[9]] + m[[10]] + m[[11]] +
##     m[[12]], data = agg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.1993 -0.8744  0.1097  1.0454  5.5631
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -67.222004   2.903378 -23.153  < 2e-16 ***
## Date          0.029715   0.001482  20.050  < 2e-16 ***
## m[[1]]       -1.082591   0.233139  -4.644 3.77e-06 ***
## m[[2]]        0.984766   0.233139   4.224 2.57e-05 ***
## m[[3]]        4.768590   0.233138  20.454  < 2e-16 ***
## m[[4]]        9.732902   0.233665  41.653  < 2e-16 ***
## m[[5]]       14.798052   0.233137  63.474  < 2e-16 ***
## m[[6]]       19.394080   0.233137  83.188  < 2e-16 ***
## m[[7]]       22.493444   0.233136  96.482  < 2e-16 ***
## m[[8]]       21.490379   0.233136  92.180  < 2e-16 ***
## m[[9]]       16.414000   0.233136  70.405  < 2e-16 ***
## m[[10]]      10.361500   0.233136  44.444  < 2e-16 ***
## m[[11]]       4.205637   0.233136  18.039  < 2e-16 ***
## m[[12]]             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.733 on 1317 degrees of freedom
## Multiple R-squared:  0.9581, Adjusted R-squared:  0.9577
## F-statistic:  2510 on 12 and 1317 DF,  p-value: < 2.2e-16
```

(c)

```r
plot(agg$temp ~ agg$Date, xlab = "Date", ylab = "Temperature", xlim = c(1900, 1902), ylim = c(-20,20));
lines(fit$fitted ~ agg$Date, col = "red", type = "l", pch = 20, lw = 3);
lines(lmod$fitted ~ agg$Date, col = "blue", type = "l", pch = 20, lw = 3)
```



(d)

```r
acf(residuals(fit), lar.max = 2);
```

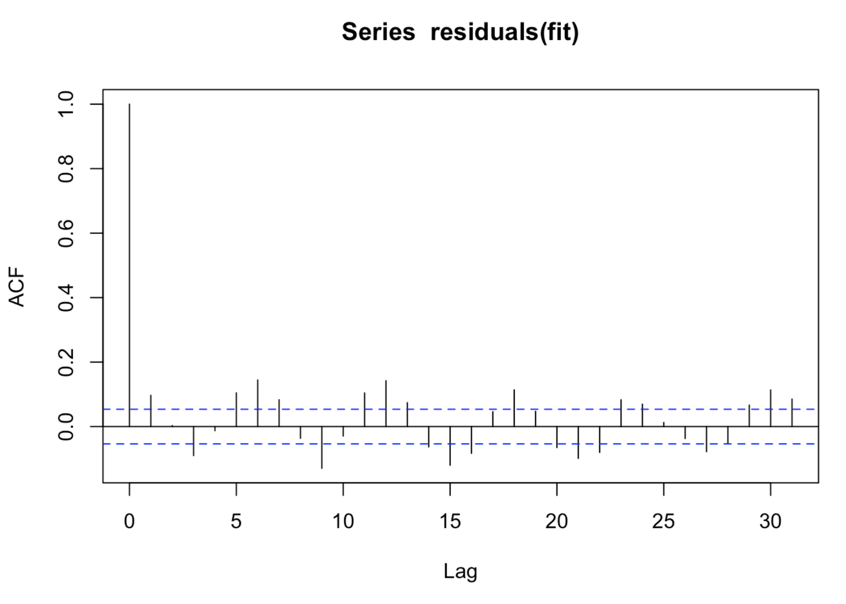
```
## Warning in plot.window(...): "lar.max" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "lar.max" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "lar.max" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "lar.max" is not a
## graphical parameter
```

```
## Warning in box(...): "lar.max" is not a graphical parameter
```

```
## Warning in title(...): "lar.max" is not a graphical parameter
```




```r
acf(residuals(lmod), lar.max = 2)
```

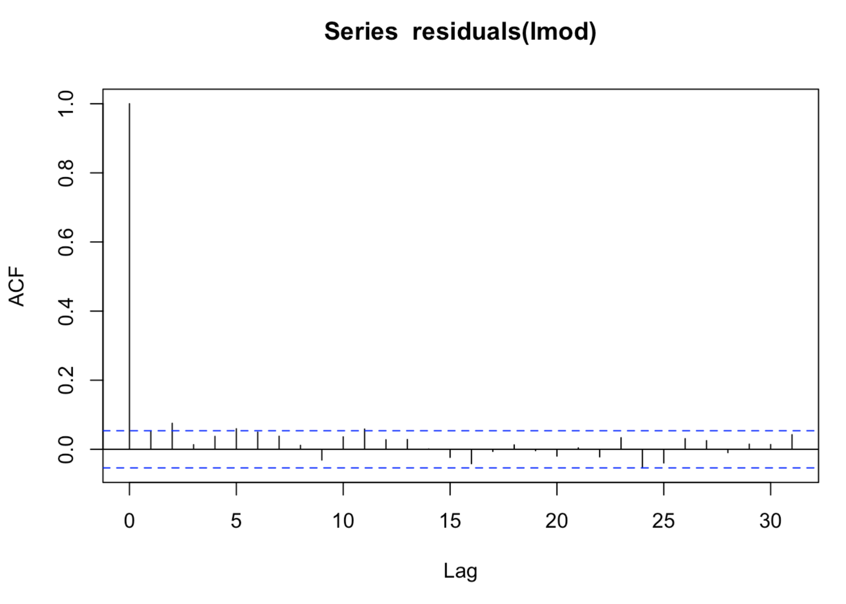
```
## Warning in plot.window(...): "lar.max" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "lar.max" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "lar.max" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "lar.max" is not a
## graphical parameter
```

```
## Warning in box(...): "lar.max" is not a graphical parameter
```

```
## Warning in title(...): "lar.max" is not a graphical parameter
```



Both moddels are very similar, though we can see the Sinusodial model has more correlation than the Dummy model. Thus we can say that the Dunny model is slightly better.

Problem 2

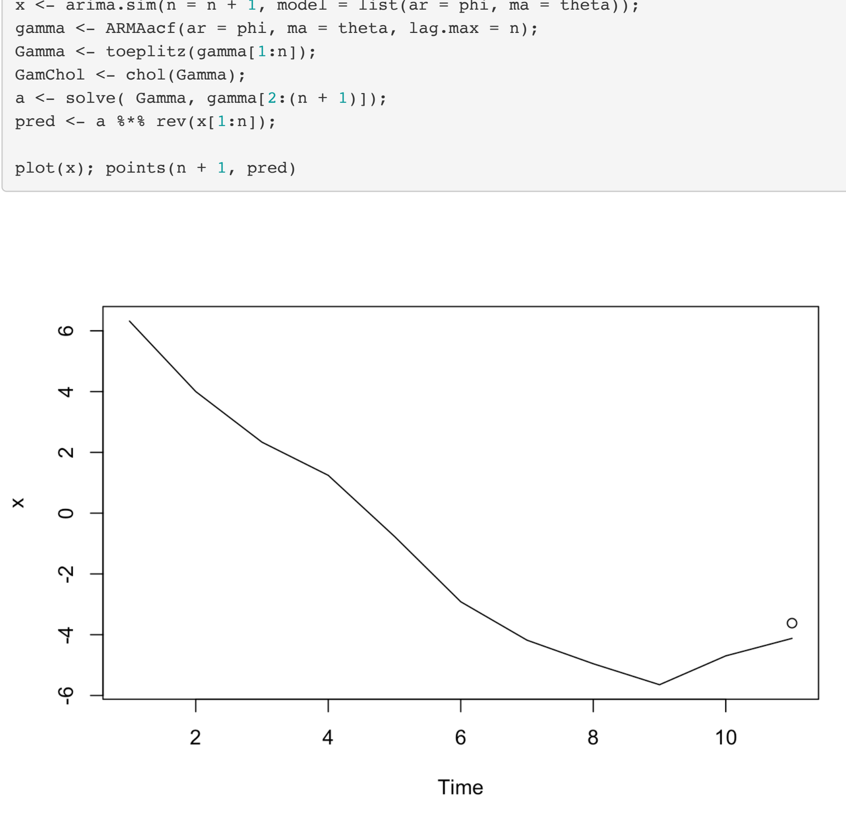
$X_t - \phi \cdot X_{t-1} = Z_t + \theta \cdot Z_{t-1}$, and assume that $|\phi| < 1$, then our ARMA(1, 1) $\{X_t\}$ process is causal.
Thus: $X_t = Z_t + (\phi + \theta) \cdot \sum_{j=1}^{\infty} \phi^{j-1} \cdot Z_{t-j}$ is our MA(∞) process.
We are given that $Z_t \sim WN(0, \sigma^2)$, so we proceed accordingly to derive:

$$
\begin{aligned}
Var(X_t) &= Var(Z_t + (\phi + \theta) \cdot \sum_{j=1}^{\infty} \phi^{j-1} \cdot Z_{t-j}) \\
&= Var(Z_t) + (\phi + \theta)^2 \cdot \sum_{j=1}^{\infty} (\phi^{j-1})^2 \cdot Var(Z_{t-j}) \\
&= \sigma^2 + (\phi + \theta)^2 \cdot \sum_{j=1}^{\infty} (\phi^2)^j \cdot \sigma^2 \\
&= \sigma^2 \cdot (1 + \frac{(\phi + \theta)^2}{1 - \phi^2})
\end{aligned}
$$

Problem 3

(a)

```r
set.seed(16)

phi = 0.9;
theta = 0.7;
n = 10;

x <- arima.sim(n = n + 1, model = list(ar = phi, ma = theta));
gamma <- ARMAacf(ar = phi, ma = theta, lag.max = n);
Gamma <- toeplitz(gamma[1:n]);
GamChol <- chol(Gamma);
a <- solve( Gamma, gamma[2:(n + 1)]);
pred <- a %*% rev(x[1:n]);

plot(x); points(n + 1, pred)
```

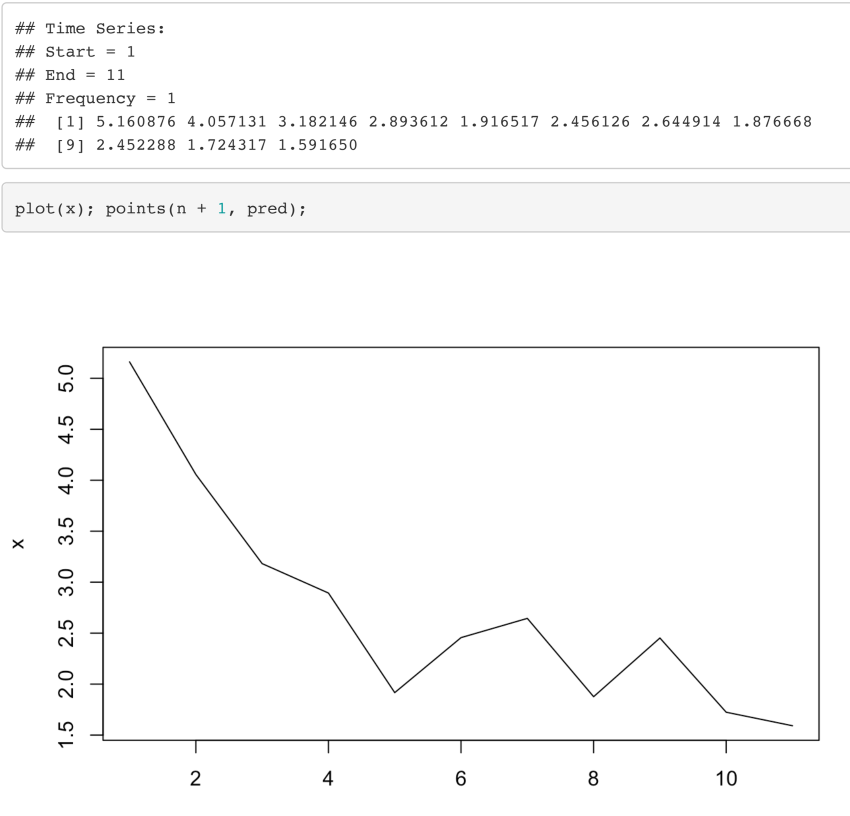


###(b)

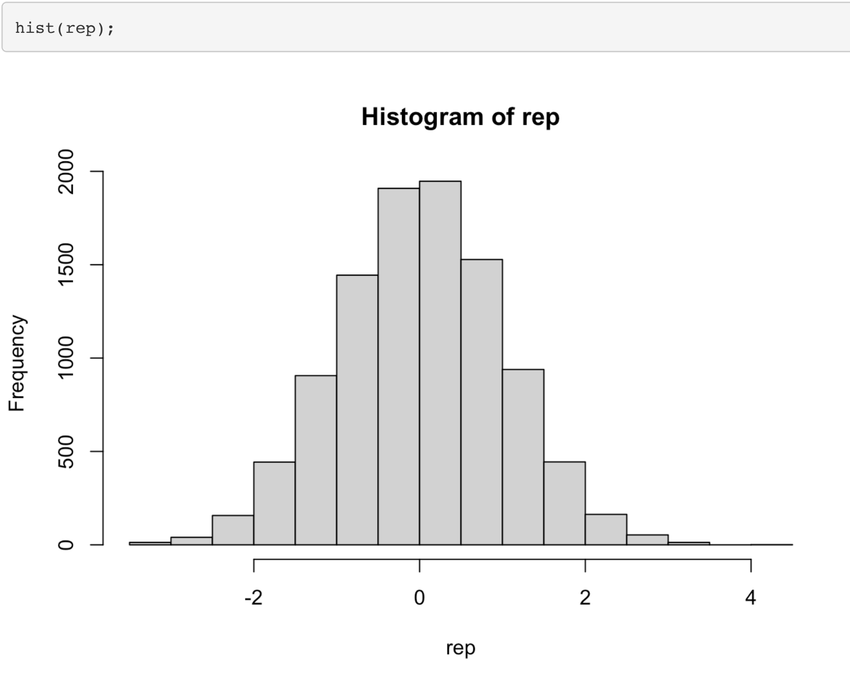
```r
set.seed(16)

num = 10000;
phi = 0.9;
theta = 0.5;
n = 10;

gamma <- ARMAacf(ar = phi, ma = theta, lag.max = n);
Gamma = toeplitz(gamma[1:n]);
GamChol = chol(Gamma);
a <- backsolve(GamChol, forwardsolve(GamChol, gamma[2:(n + 1)], transpose = TRUE, upper.tri = TRUE));

rep <- rep(0, num);
for( i in 1:num )
{
  x <- arima.sim(n = n + 1, model = list(ar = phi, ma = theta), sd = 1);
  pred <- a %*% rev(x[1:n]);
  rep[i] <- x[n + 1] - pred;
}
x;
```

```
## Time Series:
## Start = 1
## End = 11
## Frequency = 1
##  [1] 5.160876 4.057131 3.182146 2.893612 1.916517 2.456126 2.644914 1.876668
##  [9] 2.452288 1.724317 1.591650
```

```r
plot(x); points(n + 1, pred);
```




```r
hist(rep);
```




```r
m = mean(rep); m
```

```
## [1] 0.01468379
```

```r
s = var(rep); s
```

```
## [1] 0.993473
```

Not sure what to expect, though the estimate seem to be unbiased.

(c)

```r
gam <- 1 + (phi + theta)^2/(1 - phi^2);
seg2 = gam * (gamma[1] - a %*% gamma[2:(n + 1)]); seg2
```

```
##          [,1]
## [1,] 1.000001
```