

## 36-618 HW2

Daniel Nason

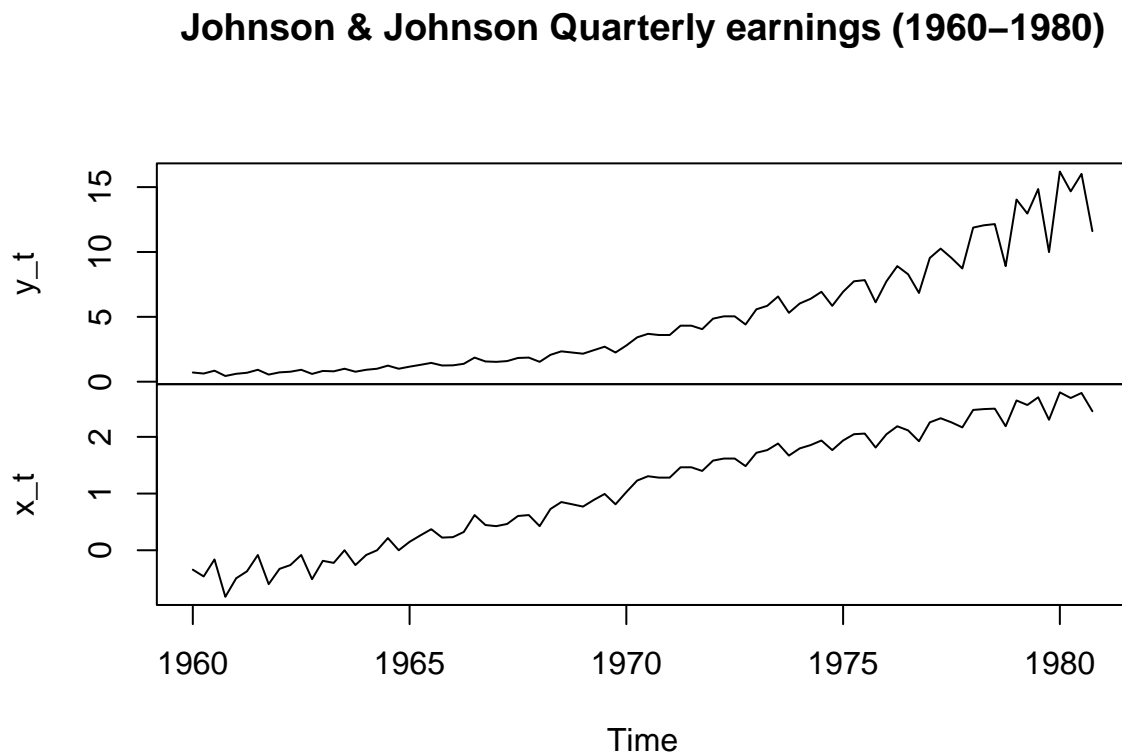
2/11/2022

```
library(astsa)
library(tidyverse)
```

### Question 1

a)

```
data <- cbind(astsa::jj, log(astsa::jj))
colnames(data) <- c("y_t", "x_t")
plot(data, main = "Johnson & Johnson Quarterly earnings (1960-1980)", ylab = "Earnings per share")
```



Based on the plot of the raw data ( $y_t$ ) over time, we see that the data resembles a multiplicative time-series model since there is both an increasing trend and increasing amplitude of the periodic trend, which illustrates non-constant autocovariance that does not depend only on time. As seen by  $x_t = \log(y_t)$ , taking the log eliminates the increasing amplitude of the period over time even though the trend remains, which lessens this non-constant autocovariance over time.

b)

```
t <- as.numeric(time(data))
tc <- t - 1970
q <- factor(cycle(data))
fit <- lm(data[,2] ~ I(tc) + I(tc^2) + I(tc^3) + q)
summary(fit)
```

```
##
## Call:
## lm(formula = data[, 2] ~ I(tc) + I(tc^2) + I(tc^3) + q)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.258771	-0.067726	-0.008221	0.058294	0.229158

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.067e+00	2.621e-02	40.696	< 2e-16 ***
I(tc)	1.920e-01	4.684e-03	40.986	< 2e-16 ***
I(tc^2)	-3.739e-04	3.541e-04	-1.056	0.29434
I(tc^3)	-3.687e-04	6.526e-05	-5.650	2.58e-07 ***
q2	3.210e-02	3.220e-02	0.997	0.32207
q3	1.063e-01	3.224e-02	3.296	0.00148 **
q4	-1.583e-01	3.230e-02	-4.902	5.14e-06 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1043 on 77 degrees of freedom
## Multiple R-squared:  0.9905, Adjusted R-squared:  0.9897
## F-statistic: 1334 on 6 and 77 DF,  p-value: < 2.2e-16
```

```
model.matrix(fit)
```

	(Intercept)	I(tc)	I(tc^2)	I(tc^3)	q2	q3	q4
## 1	1	-10.00	100.0000	-1000.000000	0	0	0
## 2	1	-9.75	95.0625	-926.859375	1	0	0
## 3	1	-9.50	90.2500	-857.375000	0	1	0
## 4	1	-9.25	85.5625	-791.453125	0	0	1
## 5	1	-9.00	81.0000	-729.000000	0	0	0
## 6	1	-8.75	76.5625	-669.921875	1	0	0
## 7	1	-8.50	72.2500	-614.125000	0	1	0
## 8	1	-8.25	68.0625	-561.515625	0	0	1
## 9	1	-8.00	64.0000	-512.000000	0	0	0
## 10	1	-7.75	60.0625	-465.484375	1	0	0

## 11	1	-7.50	56.2500	-421.875000	0	1	0
## 12	1	-7.25	52.5625	-381.078125	0	0	1
## 13	1	-7.00	49.0000	-343.000000	0	0	0
## 14	1	-6.75	45.5625	-307.546875	1	0	0
## 15	1	-6.50	42.2500	-274.625000	0	1	0
## 16	1	-6.25	39.0625	-244.140625	0	0	1
## 17	1	-6.00	36.0000	-216.000000	0	0	0
## 18	1	-5.75	33.0625	-190.109375	1	0	0
## 19	1	-5.50	30.2500	-166.375000	0	1	0
## 20	1	-5.25	27.5625	-144.703125	0	0	1
## 21	1	-5.00	25.0000	-125.000000	0	0	0
## 22	1	-4.75	22.5625	-107.171875	1	0	0
## 23	1	-4.50	20.2500	-91.125000	0	1	0
## 24	1	-4.25	18.0625	-76.765625	0	0	1
## 25	1	-4.00	16.0000	-64.000000	0	0	0
## 26	1	-3.75	14.0625	-52.734375	1	0	0
## 27	1	-3.50	12.2500	-42.875000	0	1	0
## 28	1	-3.25	10.5625	-34.328125	0	0	1
## 29	1	-3.00	9.0000	-27.000000	0	0	0
## 30	1	-2.75	7.5625	-20.796875	1	0	0
## 31	1	-2.50	6.2500	-15.625000	0	1	0
## 32	1	-2.25	5.0625	-11.390625	0	0	1
## 33	1	-2.00	4.0000	-8.000000	0	0	0
## 34	1	-1.75	3.0625	-5.359375	1	0	0
## 35	1	-1.50	2.2500	-3.375000	0	1	0
## 36	1	-1.25	1.5625	-1.953125	0	0	1
## 37	1	-1.00	1.0000	-1.000000	0	0	0
## 38	1	-0.75	0.5625	-0.421875	1	0	0
## 39	1	-0.50	0.2500	-0.125000	0	1	0
## 40	1	-0.25	0.0625	-0.015625	0	0	1
## 41	1	0.00	0.0000	0.000000	0	0	0
## 42	1	0.25	0.0625	0.015625	1	0	0
## 43	1	0.50	0.2500	0.125000	0	1	0
## 44	1	0.75	0.5625	0.421875	0	0	1
## 45	1	1.00	1.0000	1.000000	0	0	0
## 46	1	1.25	1.5625	1.953125	1	0	0
## 47	1	1.50	2.2500	3.375000	0	1	0
## 48	1	1.75	3.0625	5.359375	0	0	1
## 49	1	2.00	4.0000	8.000000	0	0	0
## 50	1	2.25	5.0625	11.390625	1	0	0
## 51	1	2.50	6.2500	15.625000	0	1	0
## 52	1	2.75	7.5625	20.796875	0	0	1
## 53	1	3.00	9.0000	27.000000	0	0	0
## 54	1	3.25	10.5625	34.328125	1	0	0
## 55	1	3.50	12.2500	42.875000	0	1	0
## 56	1	3.75	14.0625	52.734375	0	0	1
## 57	1	4.00	16.0000	64.000000	0	0	0
## 58	1	4.25	18.0625	76.765625	1	0	0
## 59	1	4.50	20.2500	91.125000	0	1	0
## 60	1	4.75	22.5625	107.171875	0	0	1
## 61	1	5.00	25.0000	125.000000	0	0	0
## 62	1	5.25	27.5625	144.703125	1	0	0
## 63	1	5.50	30.2500	166.375000	0	1	0
## 64	1	5.75	33.0625	190.109375	0	0	1

```
## 65      1  6.00 36.0000 216.000000 0 0 0
## 66      1  6.25 39.0625 244.140625 1 0 0
## 67      1  6.50 42.2500 274.625000 0 1 0
## 68      1  6.75 45.5625 307.546875 0 0 1
## 69      1  7.00 49.0000 343.000000 0 0 0
## 70      1  7.25 52.5625 381.078125 1 0 0
## 71      1  7.50 56.2500 421.875000 0 1 0
## 72      1  7.75 60.0625 465.484375 0 0 1
## 73      1  8.00 64.0000 512.000000 0 0 0
## 74      1  8.25 68.0625 561.515625 1 0 0
## 75      1  8.50 72.2500 614.125000 0 1 0
## 76      1  8.75 76.5625 669.921875 0 0 1
## 77      1  9.00 81.0000 729.000000 0 0 0
## 78      1  9.25 85.5625 791.453125 1 0 0
## 79      1  9.50 90.2500 857.375000 0 1 0
## 80      1  9.75 95.0625 926.859375 0 0 1
## 81      1 10.00 100.0000 1000.000000 0 0 0
## 82      1 10.25 105.0625 1076.890625 1 0 0
## 83      1 10.50 110.2500 1157.625000 0 1 0
## 84      1 10.75 115.5625 1242.296875 0 0 1
## attr("assign")
## [1] 0 1 2 3 4 4 4
## attr("contrasts")
## attr("contrasts")$q
## [1] "contr.treatment"
```

c)

Assuming the model is correct the logged earnings changed from the third quarter to the fourth quarter by -0.2646. The quarter with the highest average logged earnings is quarter 3, and the lowest is quarter 4. The intercept term is the estimate for the first quarter and the coefficients for quarters 2-4 are estimates of the difference in average logged earnings compared to quarter 1. Quarters 2 and 3 have average logged earnings of 0.032 and 0.106 larger than quarter 1, respectively, while quarter 4 has an average logged earnings of 0.158 less than quarter 3.

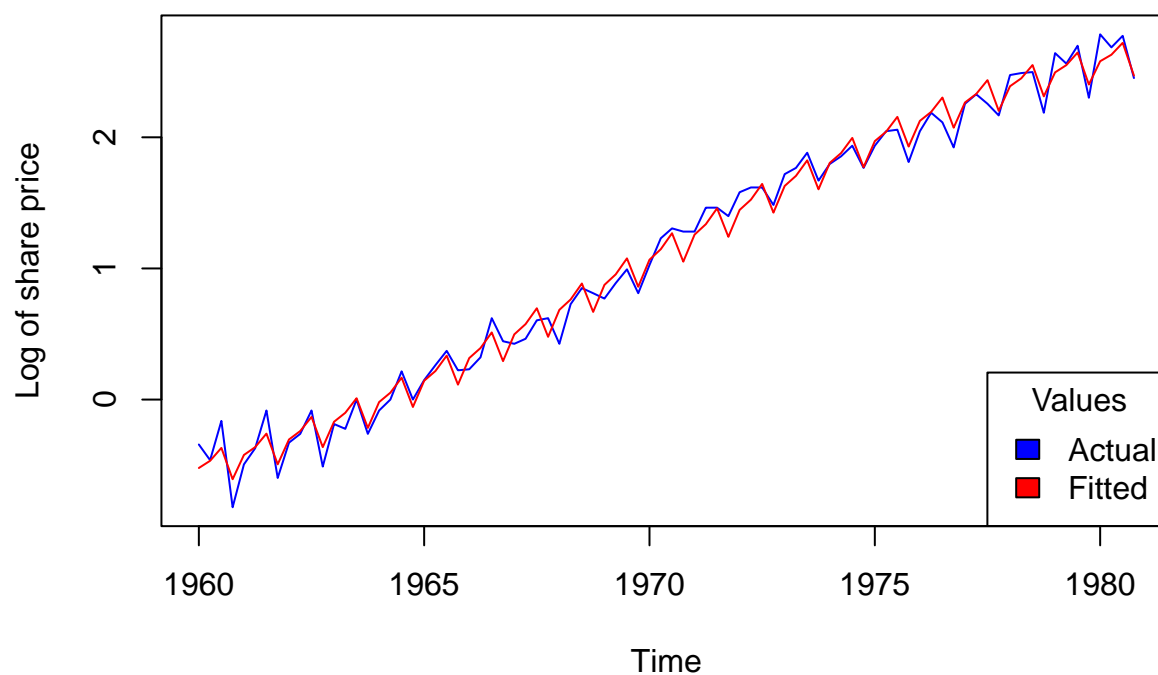
d)

If  $\alpha_1 Q_1(t)$  was included in the model, it would be perfectly collinear with the intercept term since both would estimate the average logged earnings for J&J if the time is 1970 in quarter 1. Therefore, this term is omitted from the model since it adds redundant information to the model.

e)

```
plot(data[,2], main = "Log of J & J Quarterly and fitted values of earnings over time", ylab = "Log of :
lines(t, fit$fitted.values, col = "red")
legend("bottomright", legend = c("Actual", "Fitted"), fill = c("blue", "red"), title = "Values")
```

## Log of J & J Quarterly and fitted values of earnings over time

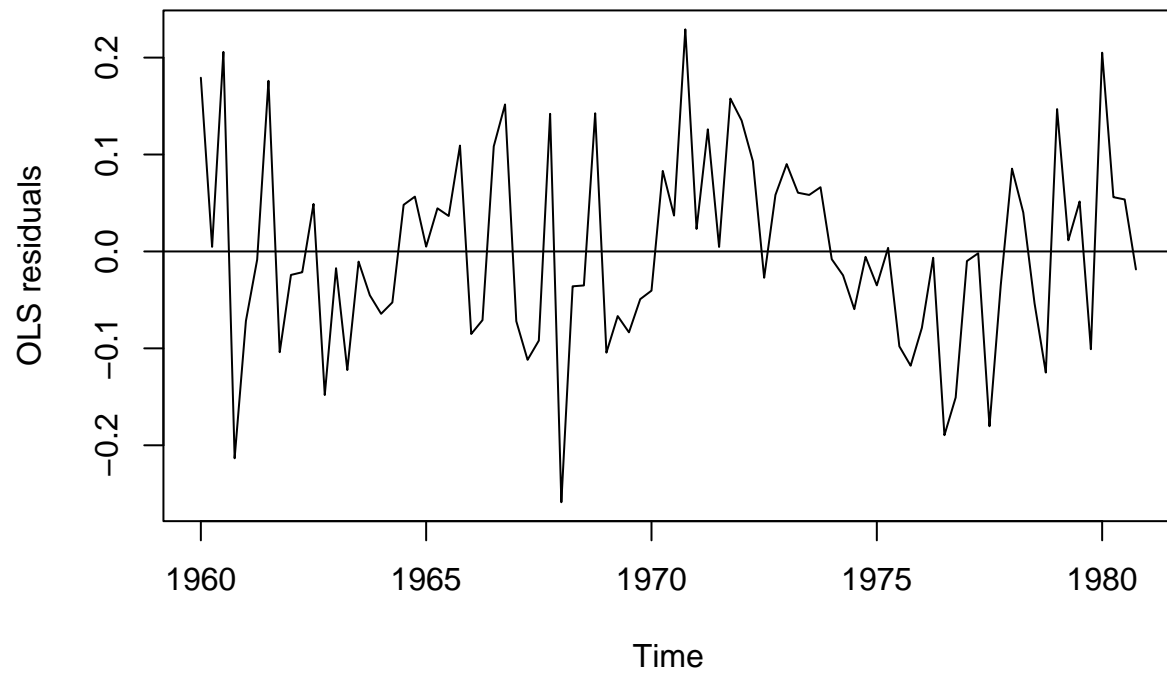


Comparing the fitted values to the actual values, the data appears to fit the data well for the majority of the data, with only slight deviations at the beginning of 1960, between 1970 and 1975 and around 1980.

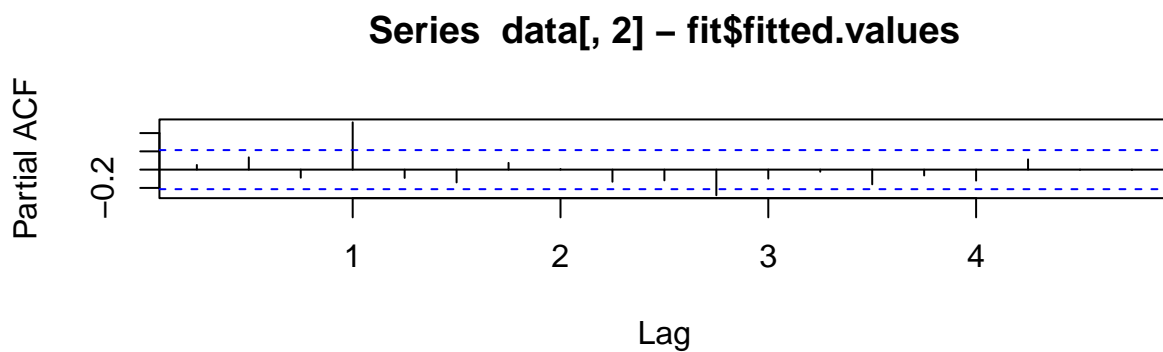
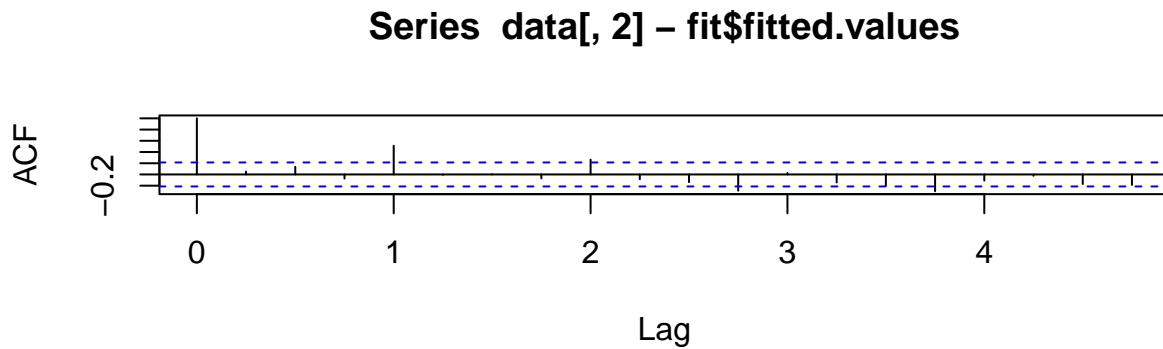
f)

```
plot(data[,2] - fit$fitted.values, ylab = "OLS residuals", main = "Residuals of the OLS regression model",  
abline(h=0))
```

## Residuals of the OLS regression model



```
par(mfrow=c(2,1))  
acf(data[,2] - fit$fitted.values)  
pacf(data[,2] - fit$fitted.values)
```



```
par(mfrow=c(1,1))
```

The plot of the residuals do not seem to be stationary because while the mean is constant around 0, the variance is not constant across time and there is evidence of a cyclical trend. The spikes also suggest evidence of covariance over time with nearby points.

Examining the ACF and PACF plots, we see that there are lagged ACF values in both plots that exceed the error bars, which implies that we cannot treat the residuals as white noise. Given that the residuals plot illustrates autocorrelation in the residuals and the ACF and PACF plots, it would not be appropriate to treat the residuals as white noise and therefore it would not be appropriate to use the OLS model from part b.

## Question 2

a)

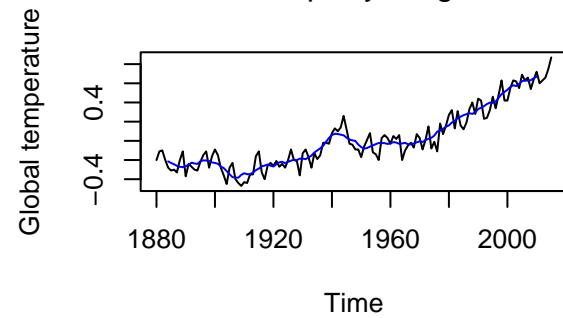
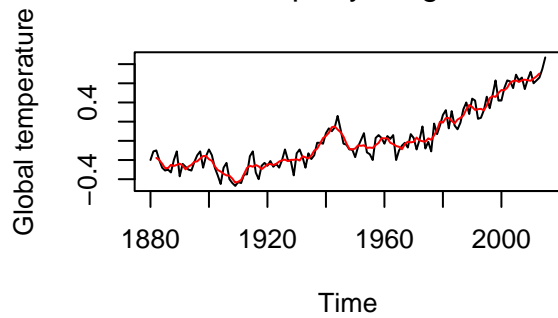
```
data2 <- astsa::globtemp
t2 <- time(data2)
## linear filters
r <- c(5, 10, 20, 50)
cols <- c("red", "blue", "green", "orange")
#plot(data2, main = "Linear filters of raw data", ylab = "Global temperature")
par(mfrow=c(2,2))
for (i in 1:length(r)){
```

```

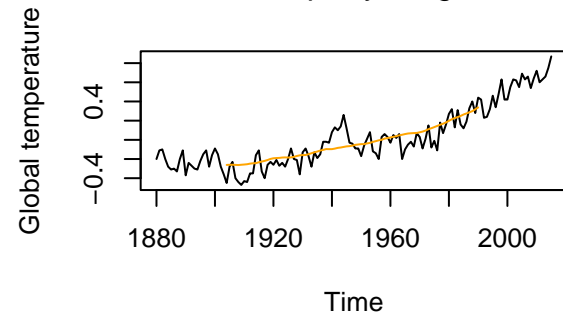
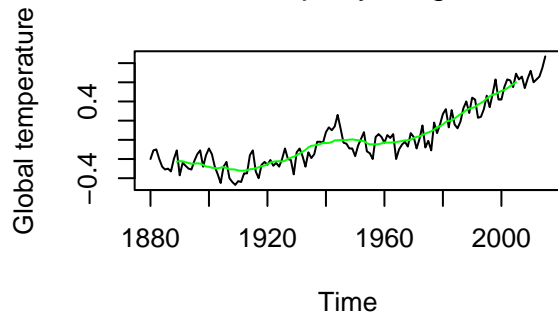
plot(data2,
      main = substitute(paste("Linear filter with ", r, " equally weighted observations")), list(r = r[
      ylab = "Global temperature")
      lines(stats::filter(data2, filter = rep(1/r[i], r[i])), col = cols[i])
}

```

Linear filter with 5 equally weighted observations      Linear filter with 10 equally weighted observations



Linear filter with 20 equally weighted observations      Linear filter with 50 equally weighted observations



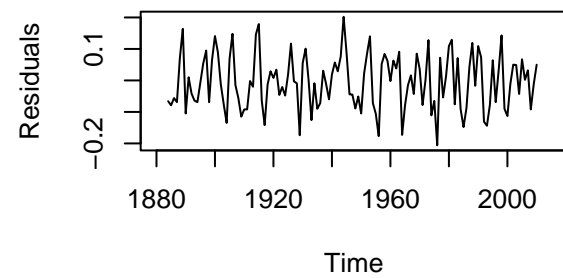
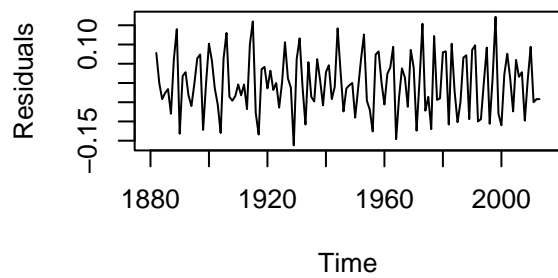
```

for (i in 1:length(r)){
  plot(data2 - stats::filter(data2, filter = rep(1/r[i], r[i])),
        main = substitute(paste("Linear filter residuals with ", r, " equally weighted observations")), 1.
        ylab = "Residuals")
}

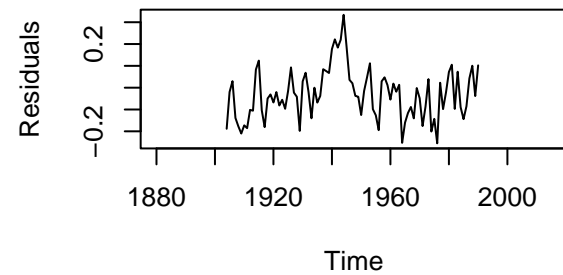
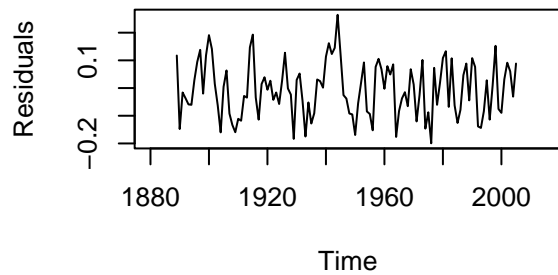
```



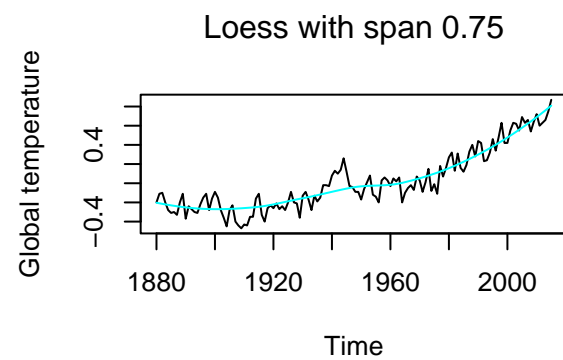
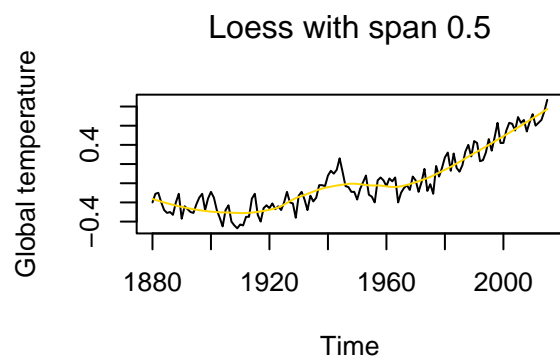
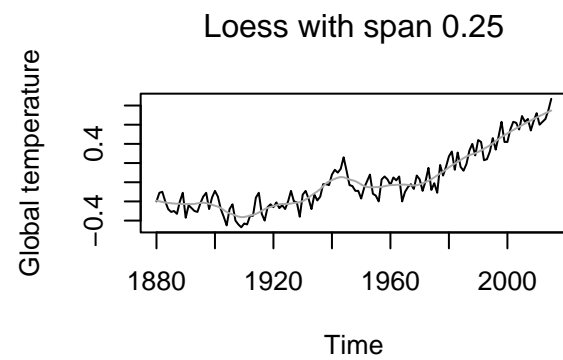
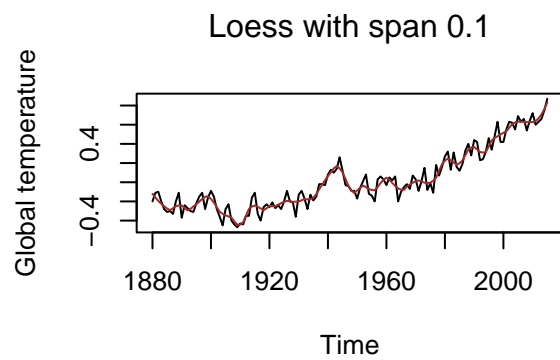
ar filter residuals with 5 equally weighted obsr filter residuals with 10 equally weighted ob:



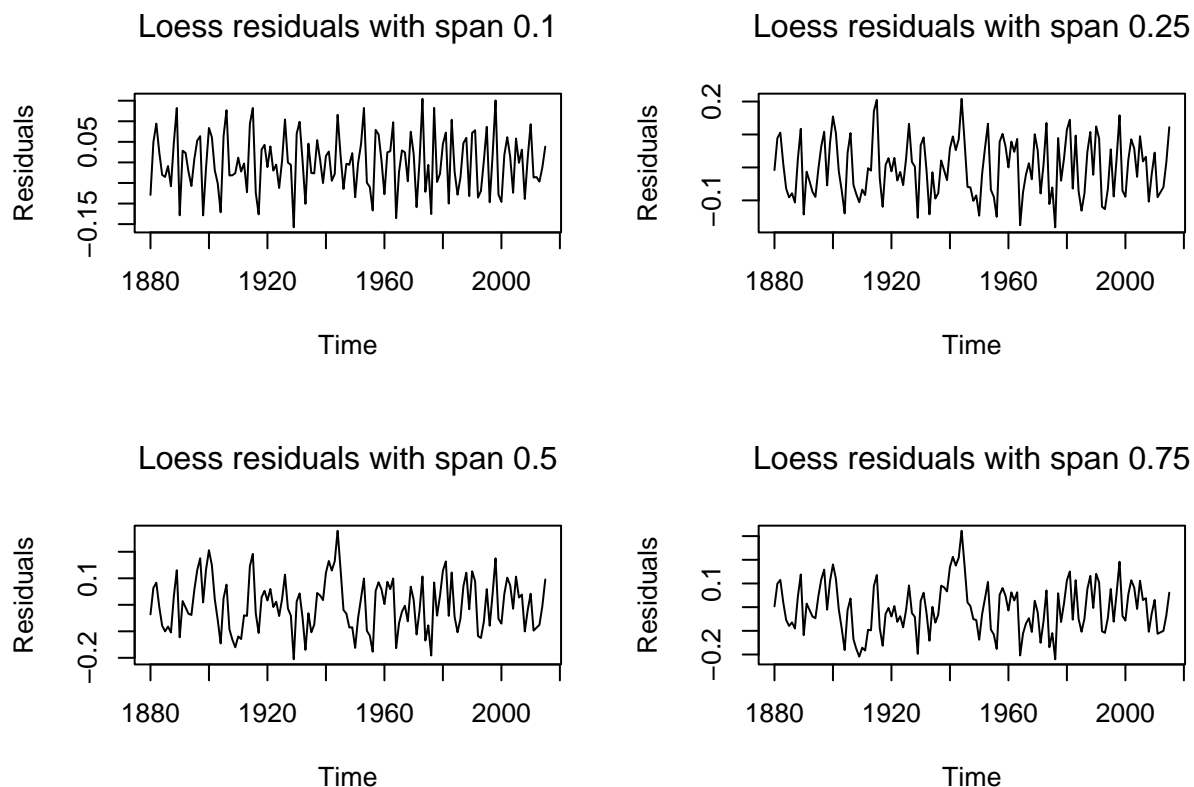
r filter residuals with 20 equally weighted ob:r filter residuals with 50 equally weighted ob:



```
# loess
s <- c(0.1, 0.25, 0.5, 0.75)
cols2 <- c("brown", "darkgrey", "gold", "cyan")
for (j in 1:length(s)){
  plot(data2,
        main = substitute(paste("Loess with span ", s), list(s = s[j])),
        ylab = "Global temperature")
  lines(as.numeric(t2), fitted(loess(data2~t2, span = s[j])), col = cols2[j])
}
```



```
for (j in 1:length(s)){
  plot(data2 - fitted(loess(data2~t2, span = s[j])),
       main = substitute(paste("Loess residuals with span ", s), list(s = s[j])),
       ylab = "Residuals")
}
```



```
par(mfrow=c(1,1))
```

For linear filters, as we increase the window of the moving average, the trend becomes smoother and less sensitive to random fluctuations in the data. However, even at the highest level of smoothing there are still small spikes in the smoothed trend. The window of the moving average filter can range from 1 to the length of the data set. However, this also removes observations at the tails of the data, so a larger span for the moving average filter implies more lost observations and therefore a shorter trend line relative to the data.

For loess, as the span of the window increases, the trend becomes smoother and less sensitive to random fluctuations in the data. The span parameter takes on values between 0 and 1. Unlike the moving average filter, loess does not remove observations at the tails of the data and therefore can provide estimates of the trends at the boundaries regardless of the smoothing parameter. The loess outputs smoother trends between data points since it relies on a local polynomial regression function to plot the data. This polynomial fitting also better job of balancing the bias variance trade-off than linear filters in that it is less sensitive to the noise in the data.

From the plots, I would choose a linear filter with a window of 10 and a loess span parameter value of 0.25. From these two, I would prefer to use loess for plotting smooth lines compared to linear filters because of the boundary behavior, the smoother fit to the data, and the fact that the model better balances the bias-variance trade-off than linear filtering.

### Question 3

See end of document for a and b.

c)

```
set.seed(36618)
ARMAacf(ma=c(0.6,-0.3), lag.max = 3) %>% round(3)

##      0      1      2      3
## 1.000 0.290 -0.207 0.000

for (i in c(100, 1000, 10000, 100000)){
  print(arima.sim(list(ma=c(0.6, -0.3)), n = i) %>%
    acf(na.action = na.omit, plot = F, lag.max = 3))
}

##
## Autocorrelations of series '.', by lag
##
##      0      1      2      3
## 1.000 0.315 -0.022 0.001
##
## Autocorrelations of series '.', by lag
##
##      0      1      2      3
## 1.000 0.310 -0.230 -0.113
##
## Autocorrelations of series '.', by lag
##
##      0      1      2      3
## 1.000 0.286 -0.208 0.006
##
## Autocorrelations of series '.', by lag
##
##      0      1      2      3
## 1.000 0.292 -0.201 0.000
```

We see that the theoretical ACF values match our computed values, and that as we simulate more Gaussian white noise variables the empirical autocorrelation at each lag approaches the theoretical autocorrelation values.

# HW2 Q3

$$x_t = w_t + 0.6w_{t-1} - 0.3w_{t-2}, \quad w_t \sim N(0, 2^2)$$

$$\begin{aligned} a) \quad \gamma(h) &= \text{Cov}(x_{t+h}, x_t) = \text{Cov}(w_{t+h} + 0.6w_{t+h-1} - 0.3w_{t+h-2}, w_t + 0.6w_{t-1} - 0.3w_{t-2}) \\ &= \text{Cov}(w_{t+h}, w_t) + \text{Cov}(w_{t+h}, 0.6w_{t-1}) + \text{Cov}(w_{t+h}, -0.3w_{t-2}) \\ &\quad + \text{Cov}(0.6w_{t+h-1}, w_t) + \text{Cov}(0.6w_{t+h-1}, 0.6w_{t-1}) + \text{Cov}(0.6w_{t+h-1}, -0.3w_{t-2}) \\ &\quad + \text{Cov}(-0.3w_{t+h-2}, w_t) + \text{Cov}(-0.3w_{t+h-2}, 0.6w_{t-1}) + \text{Cov}(-0.3w_{t+h-2}, -0.3w_{t-2}) \end{aligned}$$

Note:  $\text{Cov}(w_t, w_{t-k}) = 0$  if  $k \neq 0$  s/c  $\{w_t\}$  is white noise

Choi. 1. f.  $w_t$ ,  $h = 0, 1, 2, 3$

$$h=0$$

$$\begin{aligned} \gamma(0) &= \text{Cov}(w_t, w_t) + \text{Cov}(w_t, 0.6w_{t-1}) + \text{Cov}(w_t, -0.3w_{t-2}) \\ &\quad + \text{Cov}(0.6w_{t-1}, w_t) + \text{Cov}(0.6w_{t-1}, 0.6w_{t-1}) + \text{Cov}(0.6w_{t-1}, -0.3w_{t-2}) \\ &\quad + \text{Cov}(-0.3w_{t-2}, w_t) + \text{Cov}(-0.3w_{t-2}, 0.6w_{t-1}) + \text{Cov}(-0.3w_{t-2}, -0.3w_{t-2}) \\ &= \text{var}(w_t) + 0 + 0 + 0 + \text{var}(0.6w_{t-1}) + 0 + 0 + 0 + \text{var}(-0.3w_{t-2}) \\ &= 4 + (0.6)^2 4 + (-0.3)^2 4 = 5.8 \end{aligned}$$

$$h=1$$

$$\begin{aligned} \gamma(1) &= 0 + 0 + 0 + \text{Cov}(0.6w_t, w_t) + 0 + 0 + 0 + \text{Cov}(-0.3w_{t-1}, 0.6w_{t-1}) \\ &\quad + 0 = (0.6) \text{var}(w_t) + (-0.3 \cdot 0.6) \text{var}(w_{t-1}) = (0.6) 4 - (0.18) 4 = \\ &\quad (0.42) 4 = 1.68 \end{aligned}$$

$$h=2$$

$$\begin{aligned} \gamma(2) &= 0 + 0 + 0 + 0 + 0 + 0 + 0 + -0.3 \text{Cov}(w_t, w_t) + 0 + 0 \\ &= -0.3(4) = -1.2 \end{aligned}$$

$$h=3$$

$$\gamma(3) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$\gamma(h) = \begin{cases} 5.8, & h=0 \\ 1.68, & h=1 \\ -1.2, & h=2 \\ 0, & h=3 \end{cases}$$

$$b) \quad \rho(h), \quad h = 0, 1, 2, 3$$

$$h=0$$

$$\rho(0) = \frac{\gamma(0)}{\gamma(0)} = 1$$

$$h=1$$

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{1.68}{5.8} = 0.290$$

$$h=2$$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = \frac{-1.2}{5.8} = -0.2069$$

$$h=3$$

$$\rho(3) = \frac{\gamma(3)}{\gamma(0)} = 0$$

$$\rho(h) = \begin{cases} 1, & h=0 \\ 0.290, & h=1 \\ -0.207, & h=2 \\ 0, & h=3 \end{cases}$$