Stat 153 - Homework 1

Nicholas Lai

September 13, 2018

Question 1

(a)

```
library(gtrendsR)
library(reshape2)

#gtrendsR importing data
google.trends = gtrends(c("microsoft"), gprop = "web", time = "all")[[1]]
google.trends = dcast(google.trends, date ~ keyword + geo, value.var = "hits")

#cleaning
rownames(google.trends) = google.trends$date
google.trends$date = NULL
```

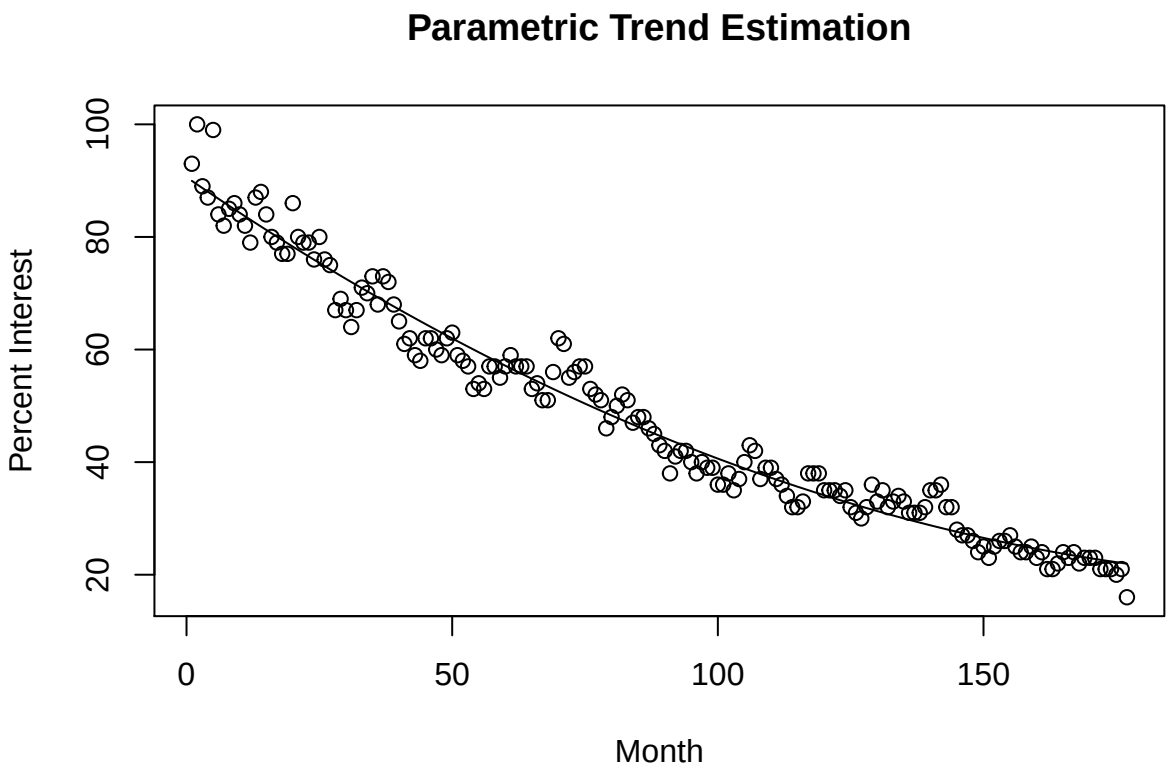
```
#data cleaning
dat <- google.trends$microsoft_world
dat <- data.frame(dat, 1:length(dat))
colnames(dat) <- c('percent_interest', 'month')

#Quadratic Regression
dat$month2 <- (dat$month)^2
quad_reg <- lm(dat$percent_interest~dat$month +dat$month2)
```

Here, a Quadratic trend of the form

$$m_t = ax^2 + bx + c$$

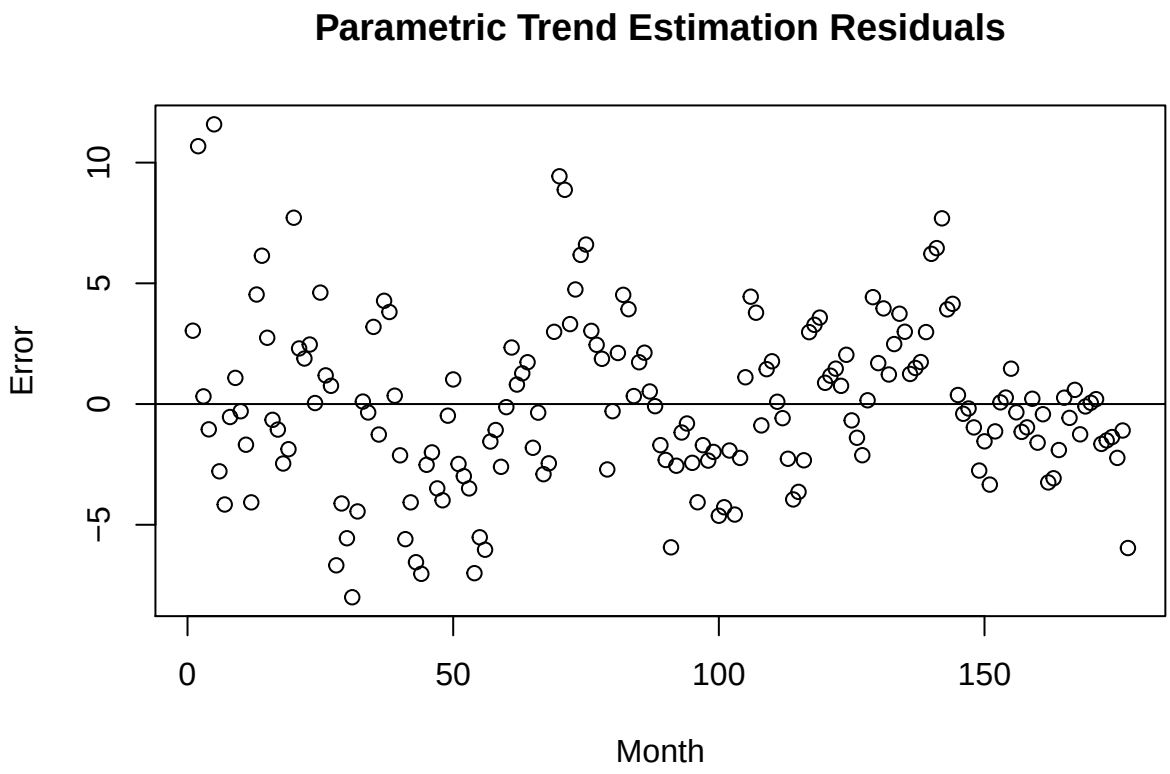
is chosen for the data.

```
plot(dat$month, dat$percent_interest, main = "Parametric Trend Estimation"
,xlab = "Month", ylab = "Percent Interest")
lines(predict(quad_reg))
```



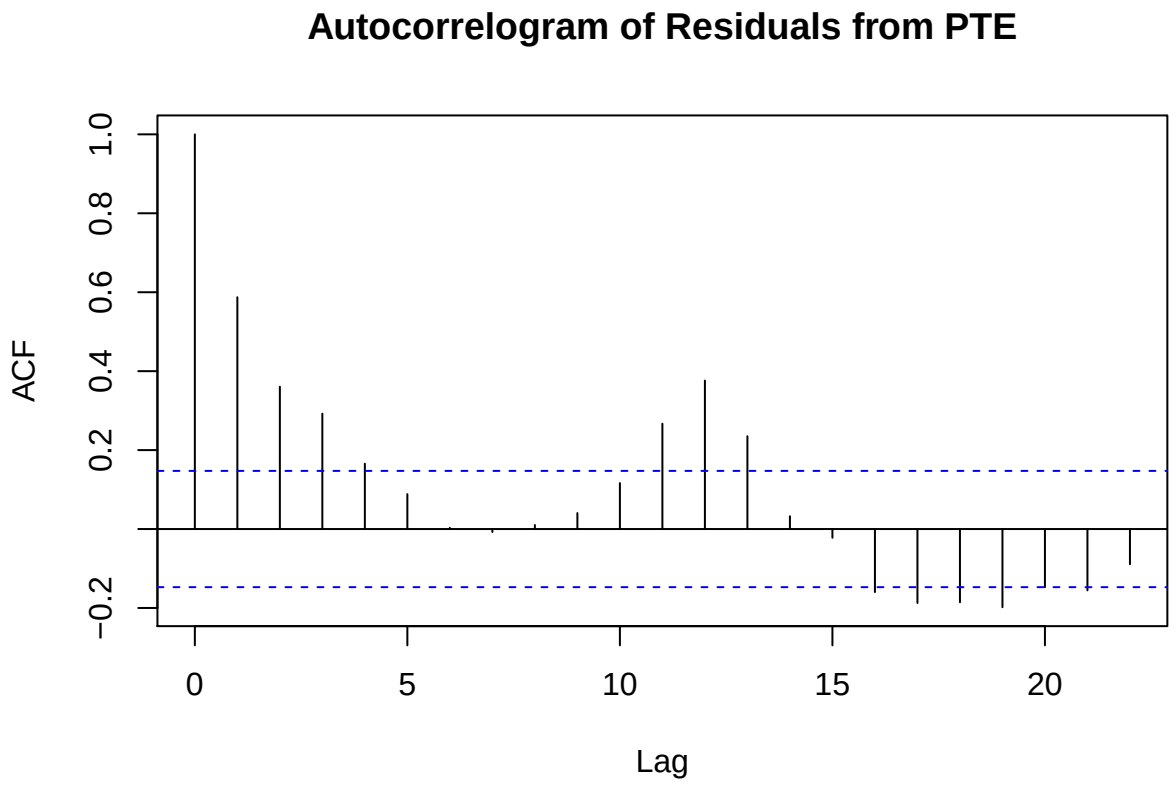
The curve seems to fit the data reasonably well, but seems not to capture any of the data's seasonality.

```
dat$residuals <- residuals(quad_reg)
plot(dat$month, dat$residuals, main = "Parametric Trend Estimation Residuals"
, xlab = "Month", ylab = "Error")
abline(h = 0)
```



If the model were a good fit for the data, we would expect the residuals to have no pattern. The residuals show a periodic pattern, so it seems that the model needs to be improved because it has systemic error.

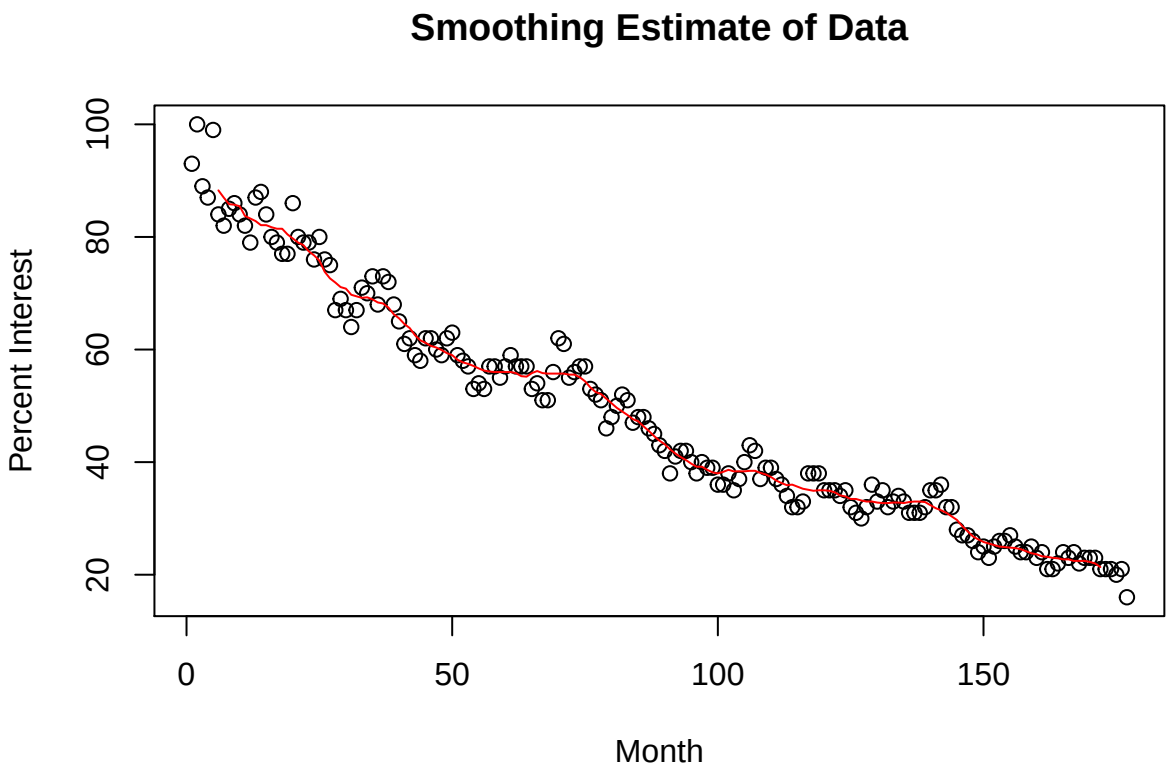
```
acf(dat$residuals, main = "Autocorrelogram of Residuals from PTE")
```



Furthermore, if the model were a good fit for the data, the autocorrelation function of the residuals should resemble those of white noise, W_t, which are i.i.d. normally distributed. The ACF of these residuals are clearly not independent, so this is further evidence of the model's shortcoming.

(b)

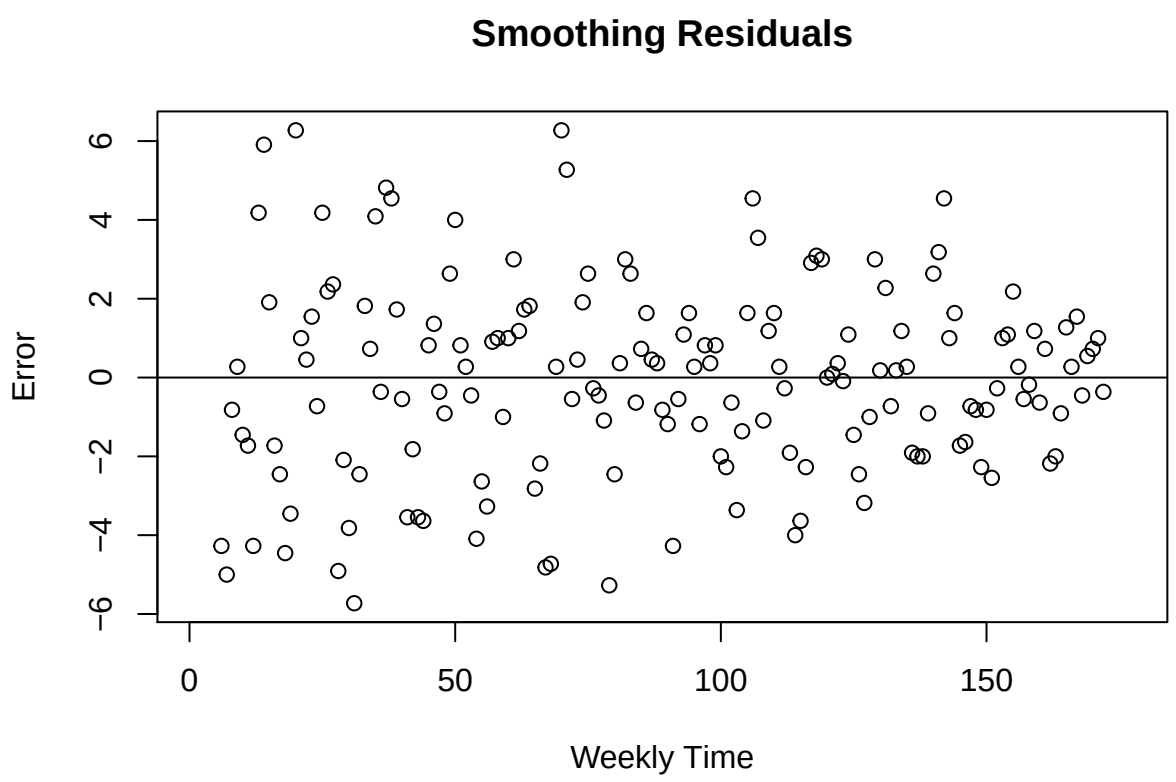
```
ts_dat = dat$percent_interest
qish = 11 #(1+2q)

smoothed = filter(ts_dat, rep(1, qish)/qish)
time = 1:length(smoothed)

plot(ts_dat, ylab = "Percent Interest", xlab = "Month",
main = "Smoothing Estimate of Data")
points(time, smoothed, type = "l", col = "red")
```



Here, the smoothing parameter $q = 5$ was chosen because too large a q would lose a significant amount of data points (here we lose 10/177) and would create a large bias in the estimate. Too small a q risks variance in the estimate, making it impractical.

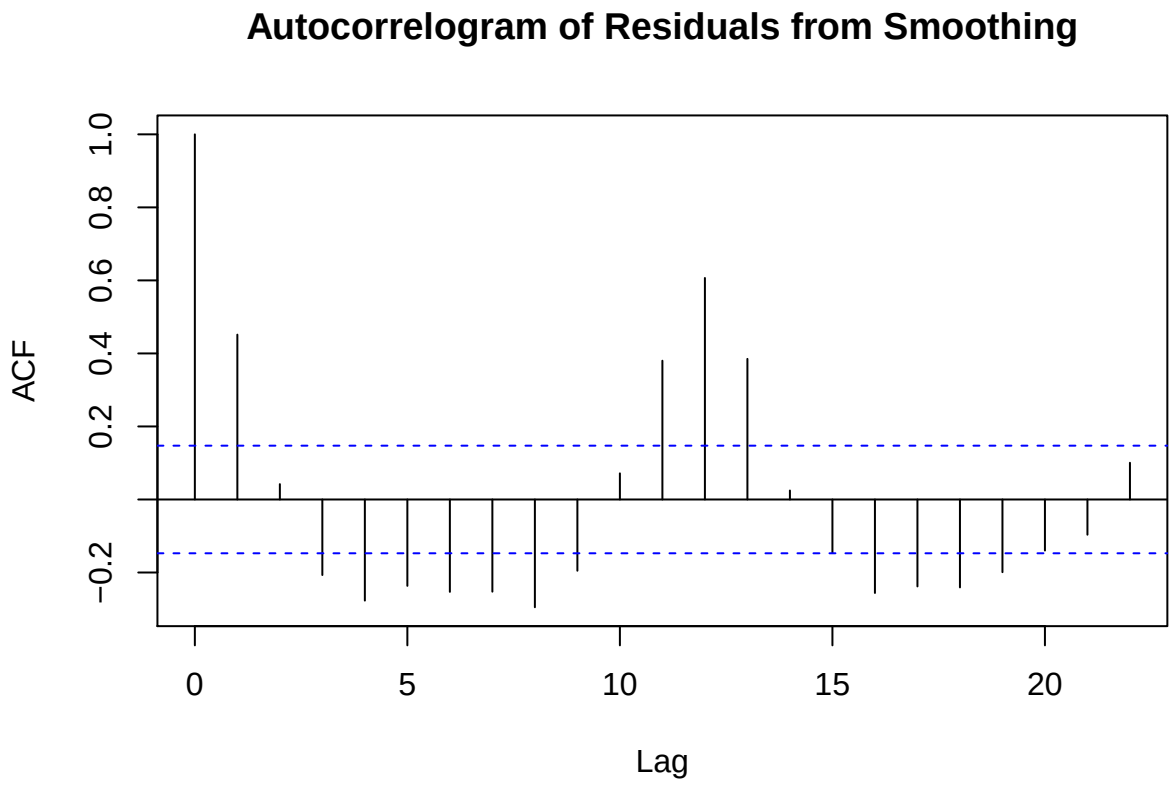
Unfortunately, varying the smoothing parameter upwards very quickly seems to erase the capacity to capture the seasonality in the data, making the estimate biased.

```
sm.res = ts_dat - smoothed
plot(time, sm.res, xlab = "Weekly Time", ylab = "Error", main ="Smoothing Residuals")
abline(h=0)
```



Here, the residuals should resemble white noise, and the residual graph gives at best weak evidence against that hypothesis.

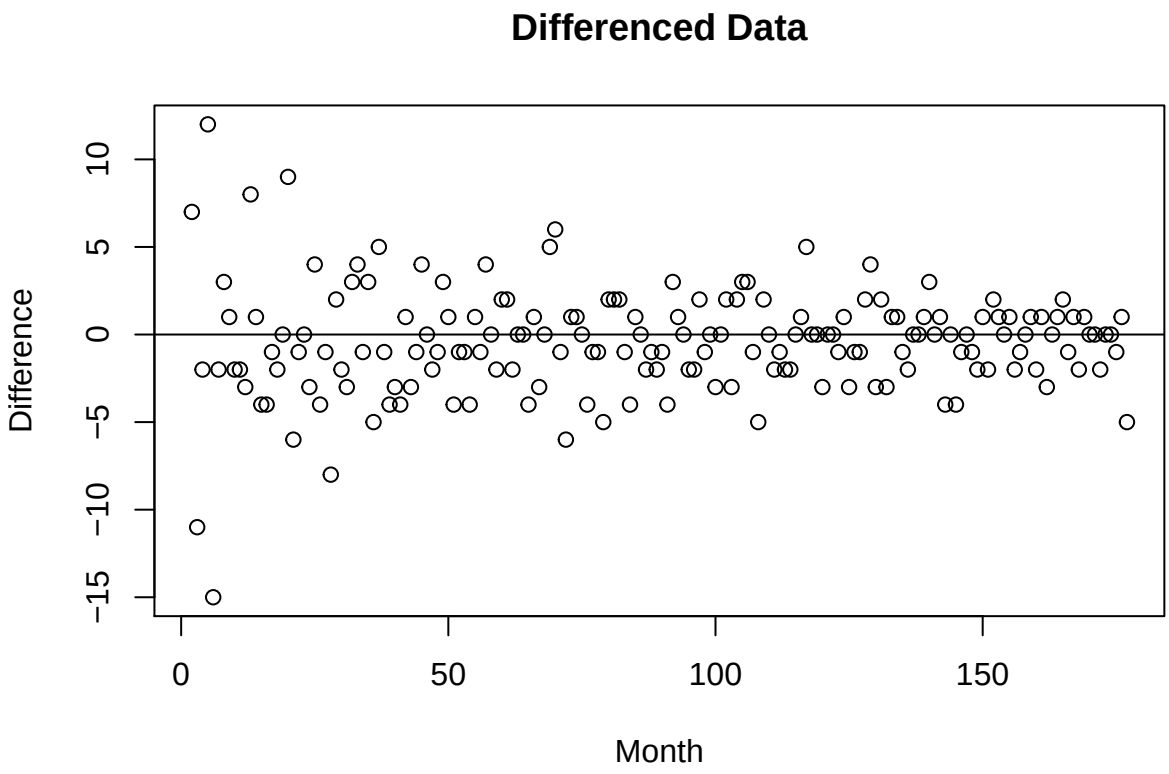
```
acf(sm.res, na.action=na.pass,
main = "Autocorrelogram of Residuals from Smoothing")
```



However, quite clearly the graph of ACFs shows that the residuals are not independent, and therefore the model is still biased.

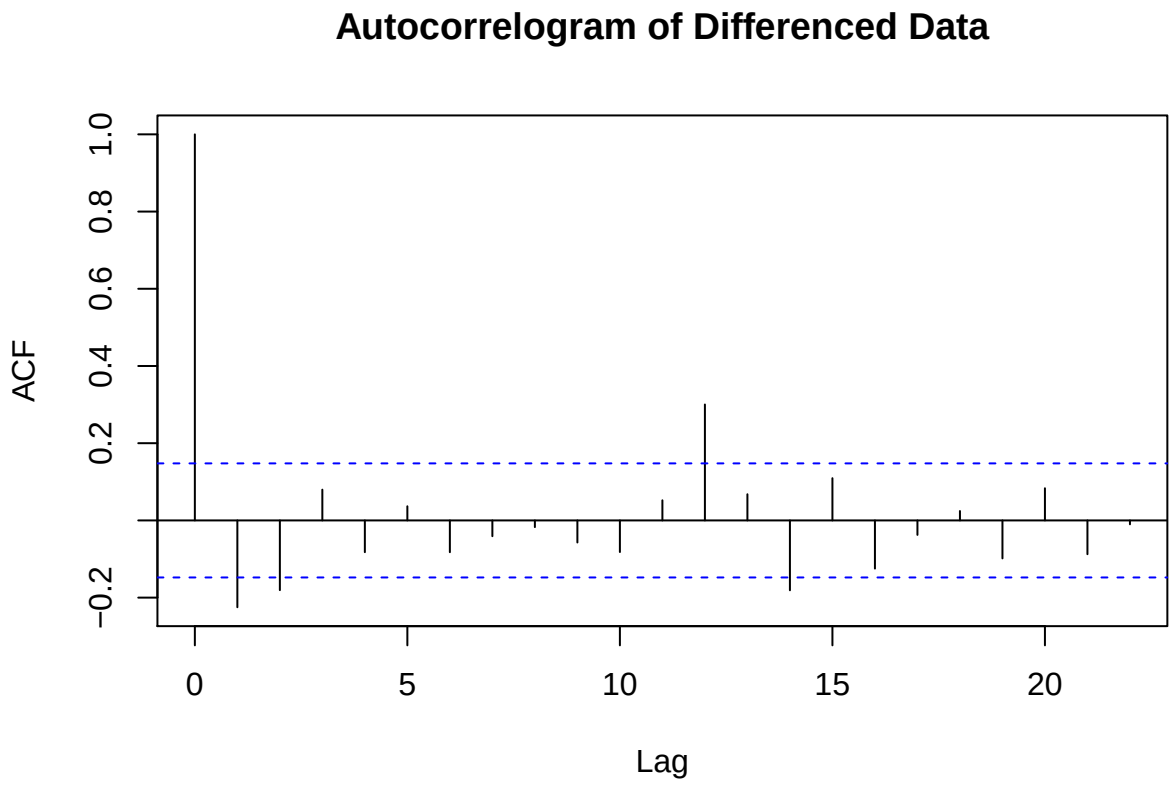
(c)

```
diff_ts = diff(ts_dat)
plot (time[-1], diff_ts, main = "Differenced Data",
xlab = 'Month', ylab = "Difference")
abline(h = 0)
```



If there is a trend left in the data after one differencing, it is very weak.

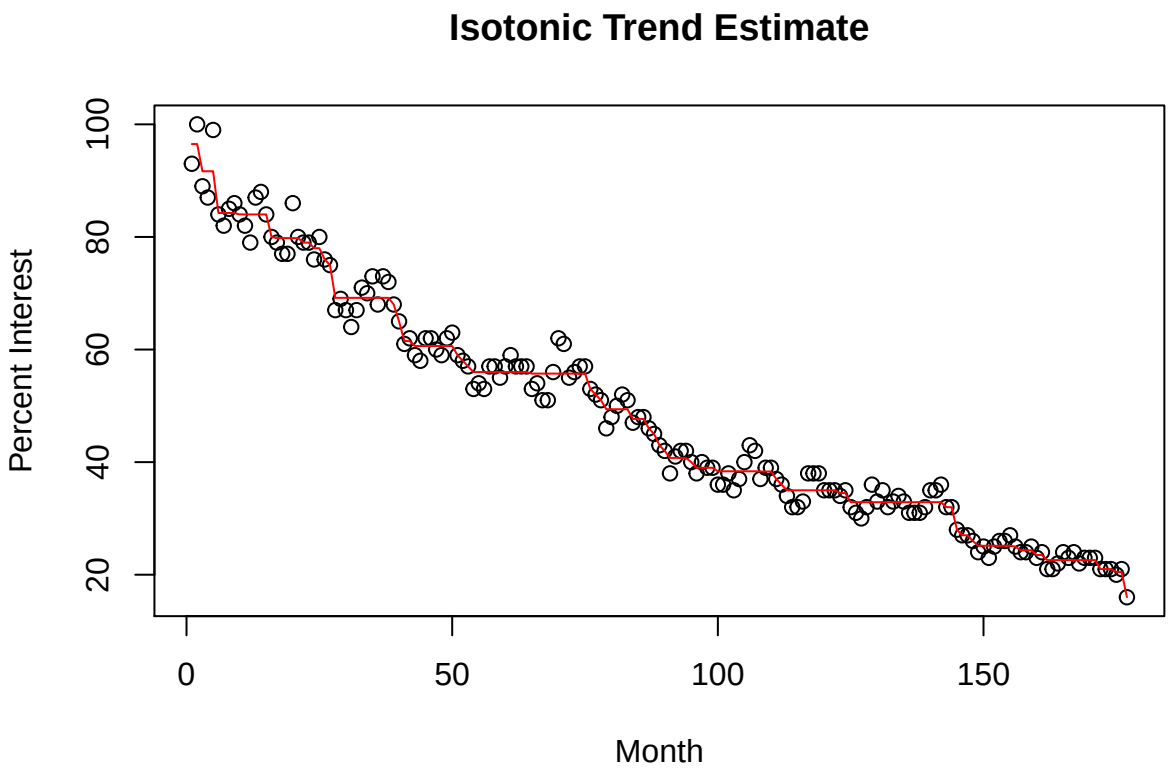
```
acf(diff_ts, main = 'Autocorrelogram of Differenced Data')
```



The ACF graph of the differenced data looks like i.i.d. normal random variables (Most inside blue band, no patterns), so there is good evidence that only white noise is left after differencing the data (no trend remaining).

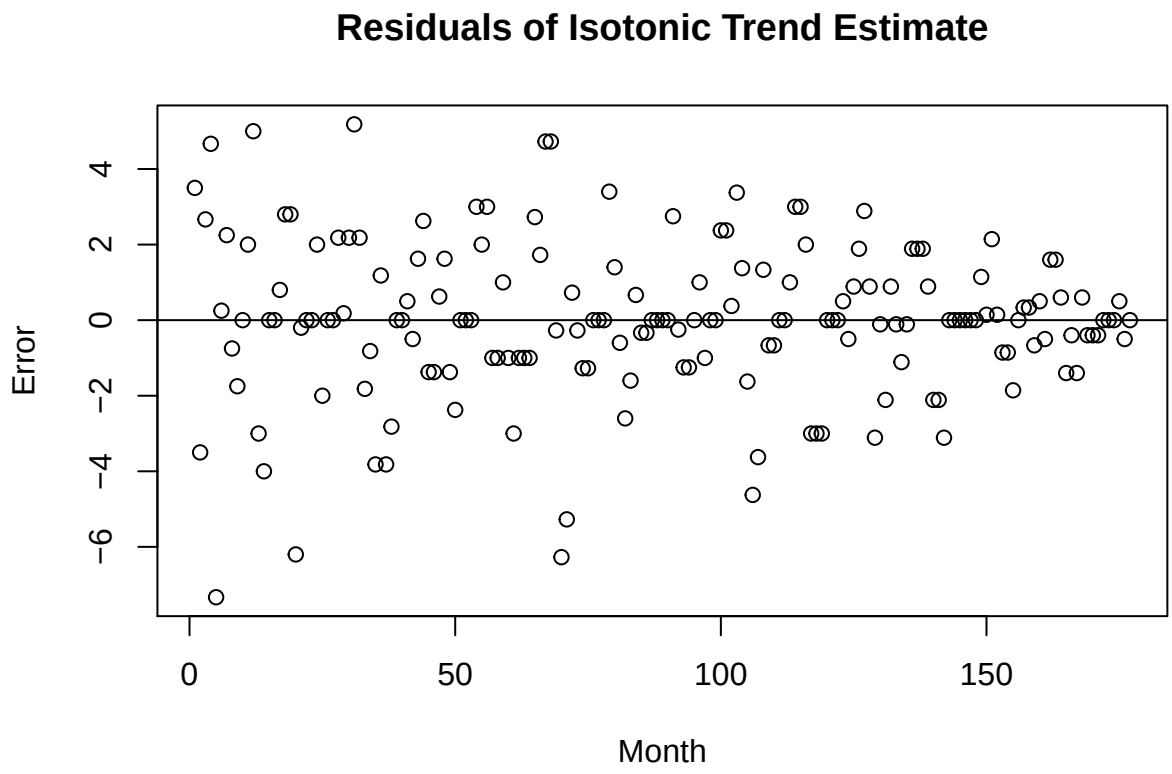
(d)

```
#isoreg is for monotone increasing, so reverse function used to reverse the data
iso_ts <-  isoreg(time, rev(dat$percent_interest))
plot(iso_ts$x, rev(iso_ts$y),
main = "Isotonic Trend Estimate", xlab = "Month", ylab = "Percent Interest")
points(iso_ts$x, rev(iso_ts$yf), type = 'l', col = 'red')
```



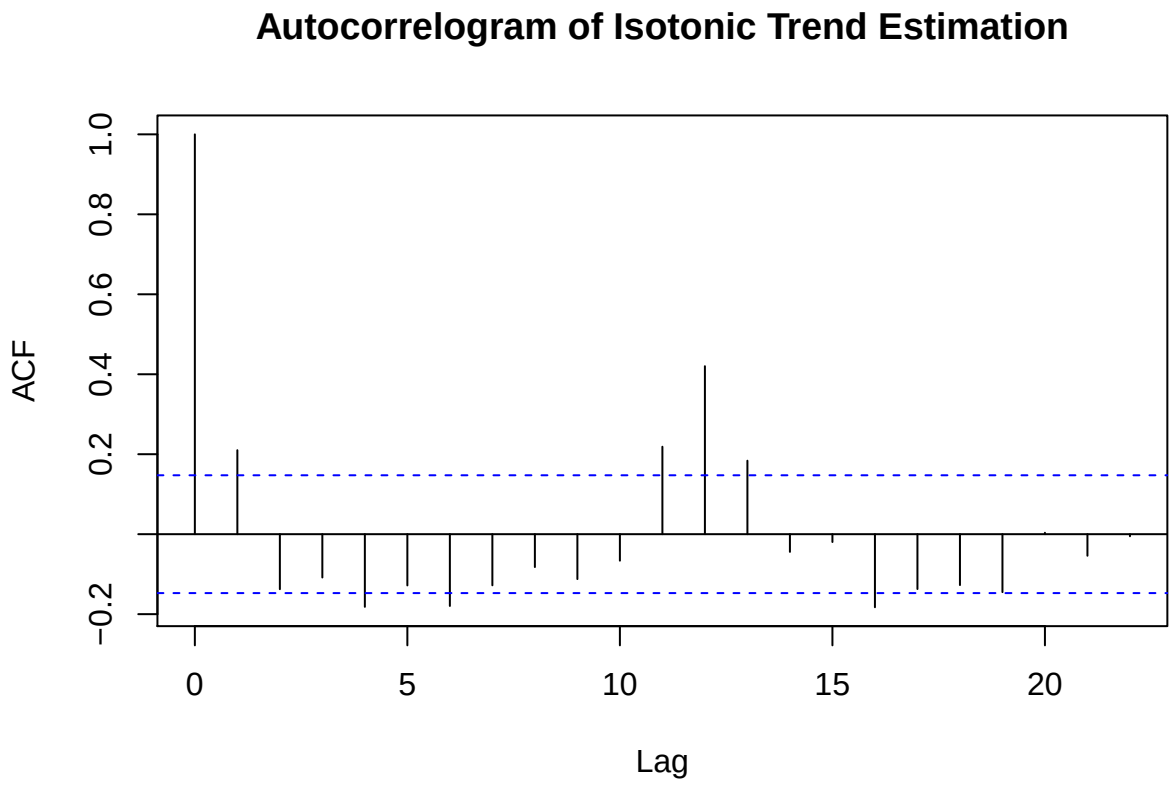
The isotone decreasing restriction of the model seems not to be true in the data, as there is some seasonality in the data. Therefore, the model's performance is in doubt.

```
residual_ts <- rev(iso_ts$yf) - rev(iso_ts$y)
plot(residual_ts, type = ,
main = "Residuals of Isotonic Trend Estimate", ylab = "Error", xlab = "Month")
abline(h=0)
```



The Residuals show weak evidence against white noise.

```
acf(residual_ts,
main = "Autocorrelogram of Isotonic Trend Estimation")
```



The ACF graph of the Isotonic Trend Estimate shows a clear periodic pattern, suggesting that the model is biased in similar ways as the models in (a) and (b) were shown to be biased.

References:

Example code for Lectures 1, 2

https://datascienceplus.com/analyzing-google-trends-data-in-r/ - Guide to basic usage of gtrendsR package