Homework 9 Solutions

Due: Wednesday 4/29/20 by 5:00pm

Spectral Methods

This week, we're going to work apply what we learned about spectral analysis. We'll use the scaled periodogram function we discussed in class, which returns the scaled periodogram, and the design matrix and estimated regression coefficients used to compute it.

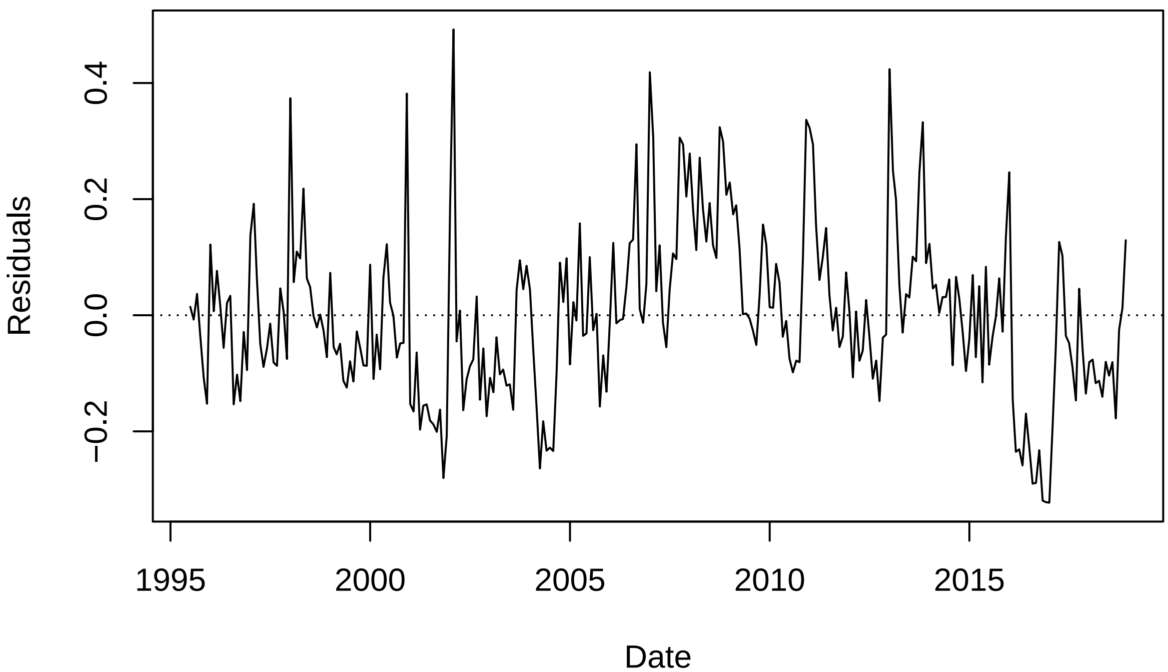
```
# Let's review how we can compute the periodogram
# We have the function we used in the homework
scaled.periodogram <- function(y) {
  n <- length(y)
  # Get number of columns in our design matrix
  Z <- matrix(nrow = n, ncol = n)

  # First column is always the intercept!
  Z[, 1] <- 1
  for (i in 2:n) {
    if (i%%2 == 0) {
      Z[, i] <- cos(2*pi*floor(i/2)*1:n/n)
    } else {
      Z[, i] <- sin(2*pi*floor(i/2)*1:n/n)
    }
  }
  linmod <- lm(y~Z-1)

  # Let's record the coef magnitudes
  m <- ifelse(n%%2 == 0, n/2, (n - 1)/2 + 1)
  coef.mags <- numeric(m)
  for (i in 1:length(coef.mags)) {
    if (i == 1) {
      coef.mags[i] <- coef(linmod)[1]^2
    } else if (i == length(coef.mags) & n%%2 == 0) {
      coef.mags[i] <- coef(linmod)[length(coef(linmod))]^2
    } else {
      coef.mags[i] <- sum(coef(linmod)[1 + 2*(i - 2) + 1:2]^2)
    }
  }
  return(list("coef.mags" = coef.mags, "freqs" = 0:(m - 1)/n, "Z" = Z,
              "coefs" = linmod$coefficients))
}
```

We're going to keep working with the `broc` data for a little while longer. It is boring to keep doing so, but it's nice to have a constant baseline as we learn. Again, it is posted on the course website, which contains the average price of one pound of broccoli in urban areas each month, from July 1995 through December 2019. In this problem, we'll regress out a polynomial time trend obtained from fitting a linear regression model to all but the last 12 months of data (we won't be doing any forecasting, so leaving out the last 12 months of data is inconsequential here, but we'll continue to leave the last 12 months out for simplicity). Let y_t refer to the corresponding residuals.

```
load("~/Dropbox/Teaching/TimeSeries2020/stat697/content/data/broc.RData")
set.seed(1)
broc$fdate <- as.Date(broc$date, "%Y-%m-%d")
broc$month <- format(broc$fdate, "%m")
broc$dayssincestart <- as.numeric(broc$fdate) - min(as.numeric(broc$fdate))
n <- nrow(broc)
m <- nrow(broc) - 12
```

(a) First, construct the residuals by regressing a linear time trend. Do the residuals appear stationary?

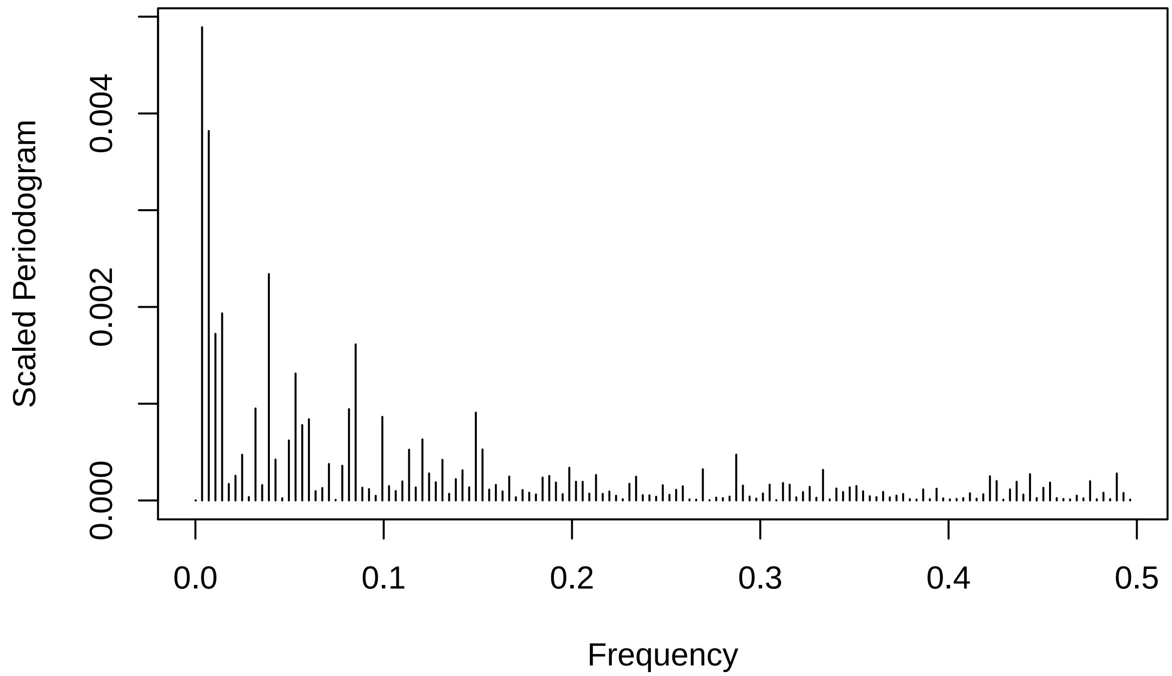
```
linmod <- lm(price~poly(dayssincestart, 1), data = broc,
             subset = 1:m)
pred <- predict(linmod, broc)
y <- (broc$price - pred)[1:m]
plot(broc$fdate[1:m], linmod$residuals,
     xlab = "Date", ylab = "Residuals", type = "l")
abline(h = 0, lty = 3)
```



The residuals do not appear stationary - there is evidence of a slowly moving time trend.

(b) Using the `scaled.periodogram` function, obtain the scaled periodogram of the residuals from (a). Plot the scaled periodogram as a function of the frequency, and identify the frequency that has the highest scaled periodogram value.

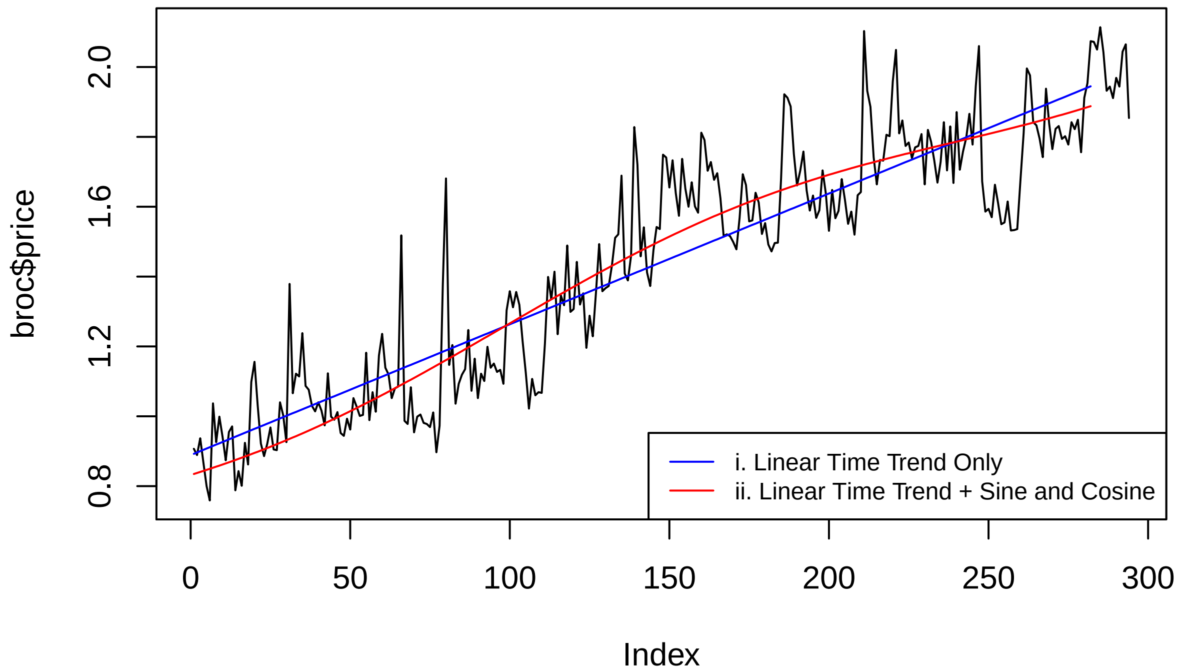
```
sp <- scaled.periodogram(y)
plot(sp$freqs, sp$coef.mags, type = "h",
     xlab = "Frequency", ylab = "Scaled Periodogram")
```



The frequency that has the highest scaled periodogram value is 0.004.

(c) Recall that the scaled periodogram at any frequency is made up of the squared coefficients for sine and cosine terms that oscillate at the same frequency. Using the output from the `scaled.periodogram` function - specifically the design matrix `Z` and the regression coefficients `coefs`, plot the first m broccoli prices. Add (i) the fitted values obtained from regressing the broccoli prices on a linear time trend and (ii) the sum of the fitted values obtained from regressing the broccoli prices on a linear time trend and the fitted values obtained by regressing the residuals on an intercept and the sine and cosine terms that oscillate at the frequency you identified in (b).

```
plot(broc$price, type = "l")
lines(pred[1:m], col = "blue")
lines(pred[1:m] + sp$Z[, 1:3]%*%sp$coefs[1:3], col = "red")
legend("bottomright", lty = c(1, 1), col = c("blue", "red"),
       legend = c("i. Linear Time Trend Only", "ii. Linear Time Trend + Sine and Cosine"),
       cex = 0.75)
```



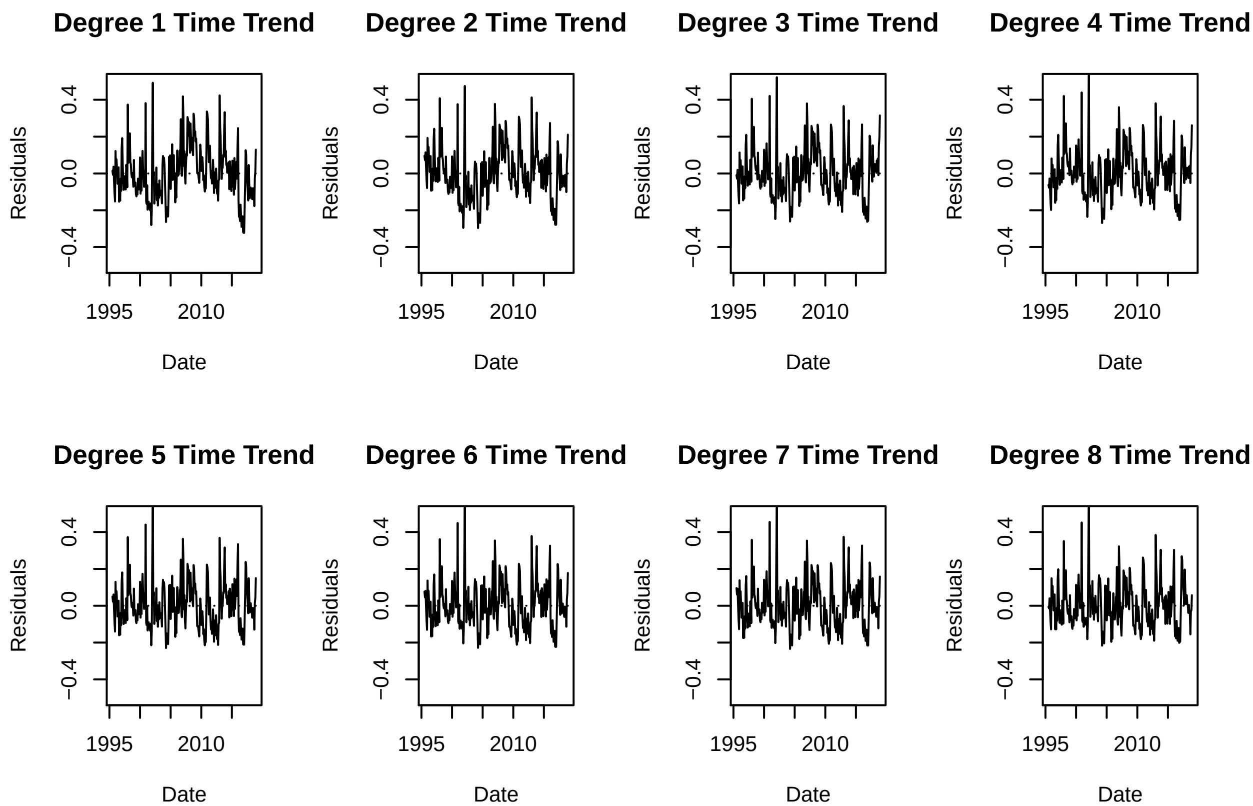
(d) Based on what you observe in (c), do you think a linear time trend is appropriate for this data?

Based on what I observed in (c), I do not think a linear time trend is appropriate for this data. It looks like there an additional slowly varying time trend is needed to explain the variability in the observed data.

(e) Now, construct the residuals by regressing out polynomial time trend from the broccoli price data, choosing the degree to be the smallest integer that produces residuals that appear stationary.

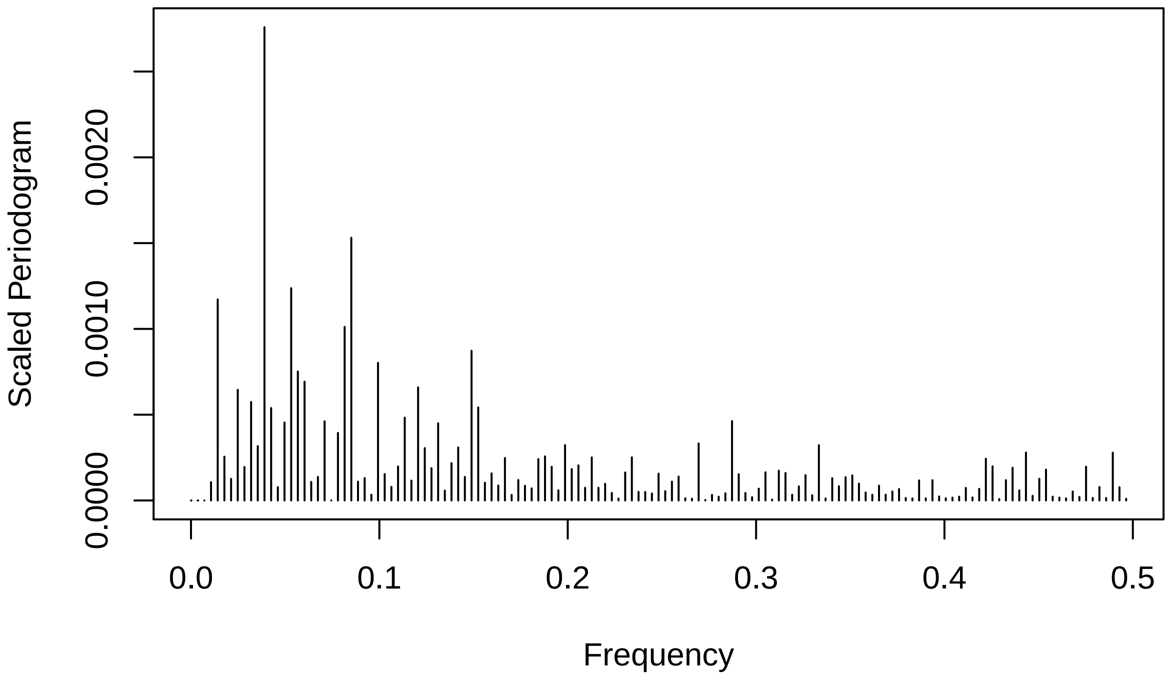
Based on visually inspecting the data, a degree 8 time trend appears to be the smallest integer that produces residuals that appear stationary. Note that this is a pretty ad-hoc choice - we're just trying to remove the slowly varying time trends so we can examine the scaled periodogram of what's left over to see if there are any remaining interpretable cyclical trends that make sense given the context of the problem.

```
par(mfrow = c(2, 4))
for (p in 1:8) {
  linmod <- lm(price~poly(dayssincestart, p), data = broc,
               subset = 1:m)
  plot(broc$fdate[1:m], linmod$residuals,
     xlab = "Date", ylab = "Residuals", type = "l",
     main = paste("Degree ", p, " Time Trend", sep = ""), ylim = c(-0.5, 0.5))
  abline(h = 0, lty = 3)
}
```



(f) Using the `scaled.periodogram` function, obtain the scaled periodogram of the residuals from (e). Plot the scaled periodogram as a function of the frequency.

```
linmod <- lm(price~poly(dayssincestart, 8), data = broc,
             subset = 1:m)
pred <- predict(linmod, broc)
y <- (broc$price - pred)[1:m]
sp <- scaled.periodogram(y)
plot(sp$freqs, sp$coef.mags, type = "h",
     xlab = "Frequency", ylab = "Scaled Periodogram")
```



(g) Identify the five frequencies that have the highest scaled periodogram value. Do any of these frequencies correspond to variability that makes some sense to you, given the time scale? Which observations do they indicate positive or negative correlation between?

Scaled Periodogram Value	Frequency	Inverse Frequency (Period Length)
2.76×10^{-3}	0.039	25.636
1.53×10^{-3}	0.085	11.750
1.24×10^{-3}	0.053	18.800
1.17×10^{-3}	0.014	70.500
1.01×10^{-3}	0.082	12.261

Several of the frequencies that correspond to the highest scaled periodogram values make sense in the context of the problem. In particular, the frequencies that correspond to correlations between observations that are 0.5, 1, 1.5, 2 years apart make sense in terms of this problem, in which the data are observed monthly and have clear seasonal trends.

```
head((1/sp$freq)[order(sp$coef.mags, decreasing = TRUE)])
```

Note - another way to assess which other frequencies contribute to explaining the observed time series would have been to just look at the rest of the frequencies with large periodogram values after just regressing out a linear trend. Frequencies corresponding to 1 and 2 year periods are among the frequencies with the highest scaled periodogram values when the scaled periodogram is constructed from the residuals after regressing out a linear time trend.

```
linmod <- lm(price~poly(dayssincestart, 1), data = broc,
             subset = 1:m)
pred <- predict(linmod, broc)
y <- (broc$price - pred)[1:m]
sp <- scaled.periodogram(y)
head((1/sp$freq)[order(sp$coef.mags, decreasing = TRUE)])
```