Bruce Campell NCSU ST 534 HW 5

Problems 4.13, 4.16, 4.18*, and 4.27 from CH4

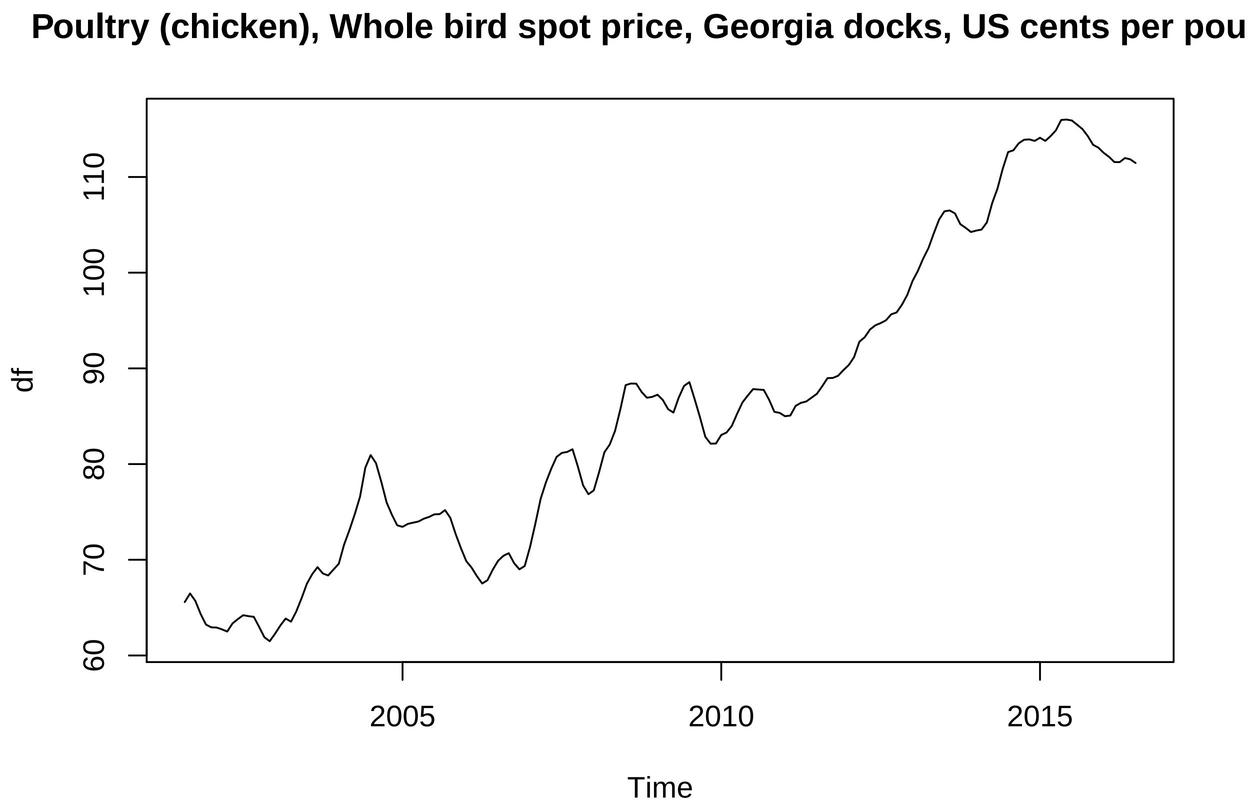
Shumway, Robert H.; Stoffer, David S. Time Series Analysis and Its Applications: With R Examples (Springer Texts in Statistics)

19 November, 2017

4.13 chiken price data.

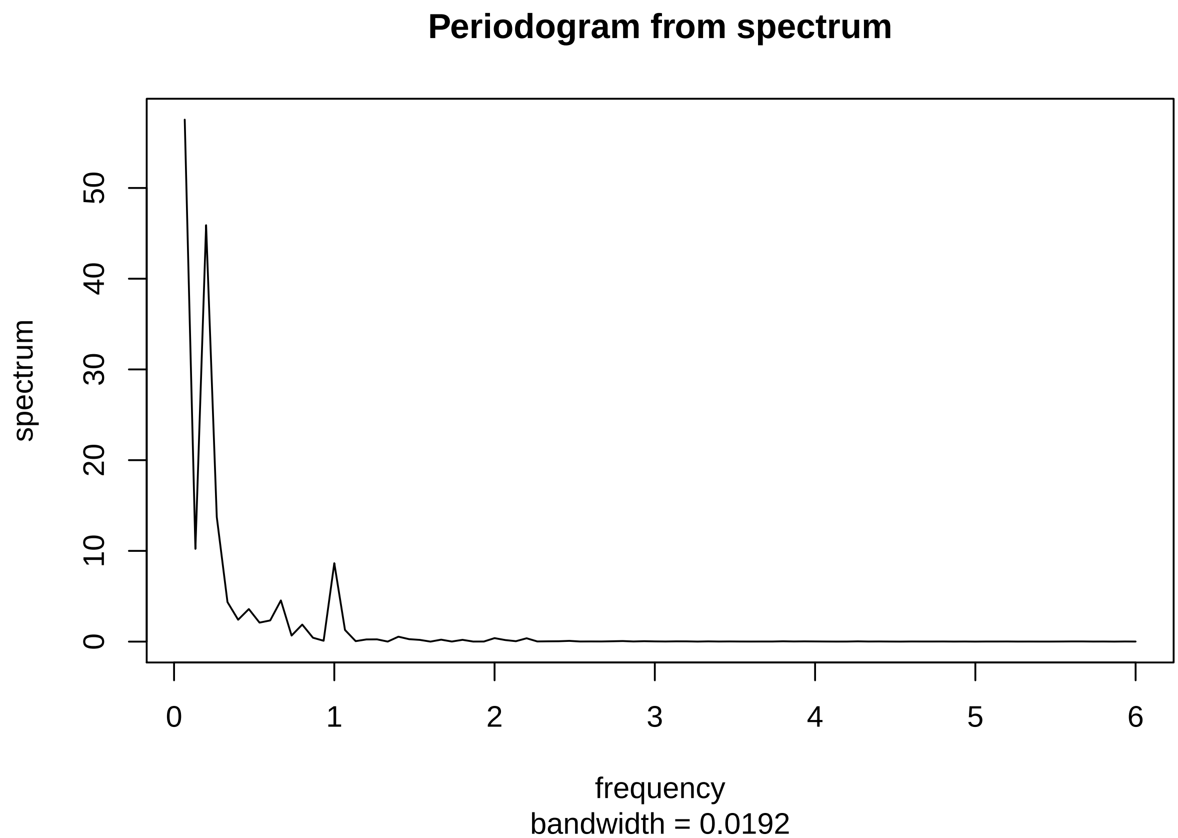
Analyze the chicken price data (chicken) using a nonparametric spectral estimation procedure. Aside from the obvious annual cycle discovered in Example 2.5, what other interesting cycles are revealed?

```
rm(list = ls())
data(chicken, package = "astsa")
df <- chicken
plot(df, main = "Poultry (chicken), Whole bird spot price, Georgia docks, US cents per p
```



Let's calculate the periodogram. There should not be a need to detrend the data since we expect the mean to be represented by the spectrum at $\omega = 0$.

```
df.periodogram.spectrum <- spectrum(df, taper = 0, log = "no", main = "Periodogram from
```




```
max.spec.loc <- which.max(df.periodogram.spectrum$spec)
max.spec <- df.periodogram.spectrum$spec[max.spec.loc]
max.spec.freq <- df.periodogram.spectrum$freq[max.spec.loc]

pander(data.frame(max.spec.freq = max.spec.freq, max.spec = max.spec), caption = "peak f
```

Table 1: peak from spectrum command We see a long term cycle around 15 months in periodicity There is another peak in the spectrum around around 1

max.spec.freq	max.spec
0.06667	57.53

```
max.spec.loc <- which.max(df.periodogram.spectrum$spec)
max.spec <- df.periodogram.spectrum$spec[15]
max.spec.freq <- df.periodogram.spectrum$freq[15]

pander(data.frame(max.spec.freq = max.spec.freq, max.spec = max.spec), caption = "minor
```

Table 2: minor peak from spectrum command

max.spec.freq	max.spec
1	8.643

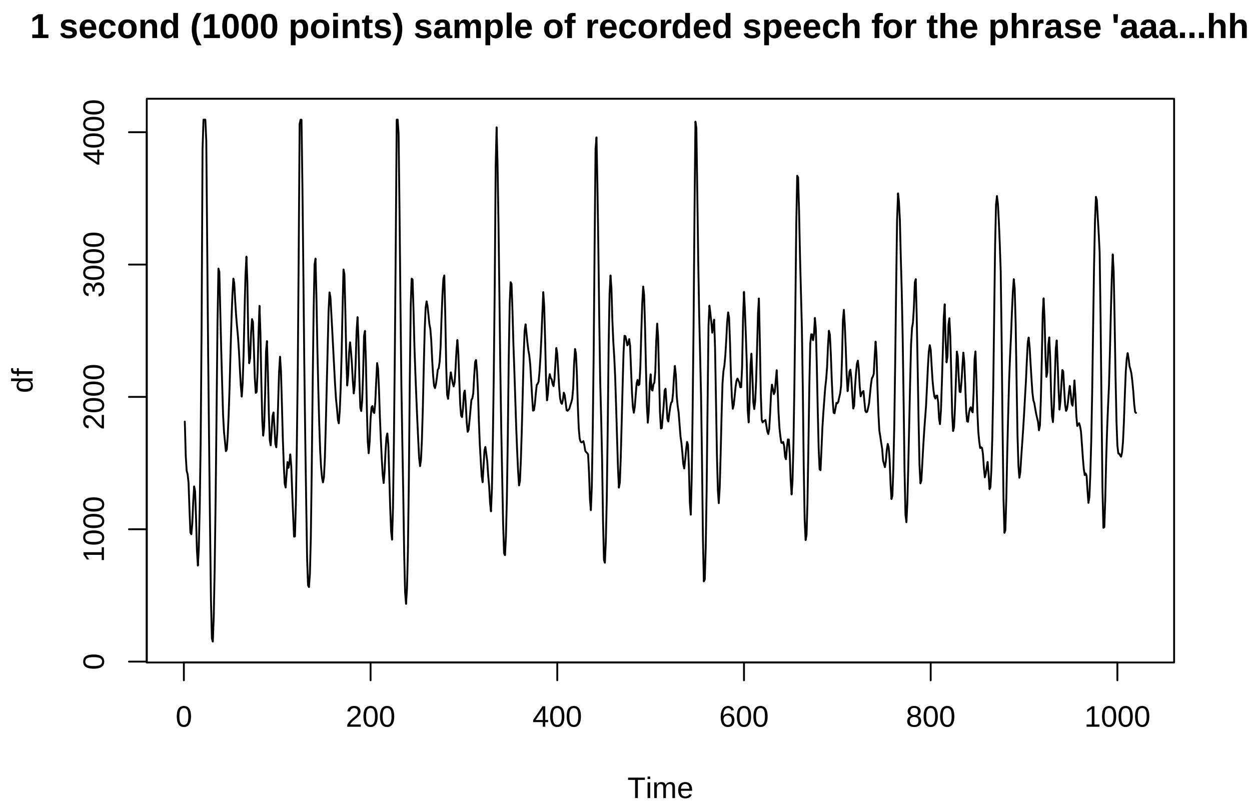
4.16 Cepstral Analysis.

The periodic behavior of a time series induced by echoes can also be observed in the spectrum of the series; this fact can be seen from the results stated in Problem 4.7. Using the notation of that problem, suppose we observe $x_t = s_t + As_{t-D} + n_t$, which implies the spectra satisfy $f_x(\omega) = [1 + A2 + 2Acos(2\pi\ \omega\ D)]f_s(\omega) + f_n(\omega)$. If the noise is negligible $(fn(\omega) \approx 0)$ then $log f_x(\omega)$ is approximately the sum of a periodic component, $log[1+A2+2Acos(2\pi\ \omega\ D)]$, and $log f_s(\omega)$. Bogart et al. [27] proposed treating the detrended log spectrum as a pseudo time series and calculating its spectrum, or cepstrum, which should show a peak at a quefrency corresponding to 1/D. The cepstrum can be plotted as a function of quefrency, from which the delay D can be estimated. For the speech series presented in Example 1.3, estimate the pitch period using cepstral analysis as follows. The data are in speech.

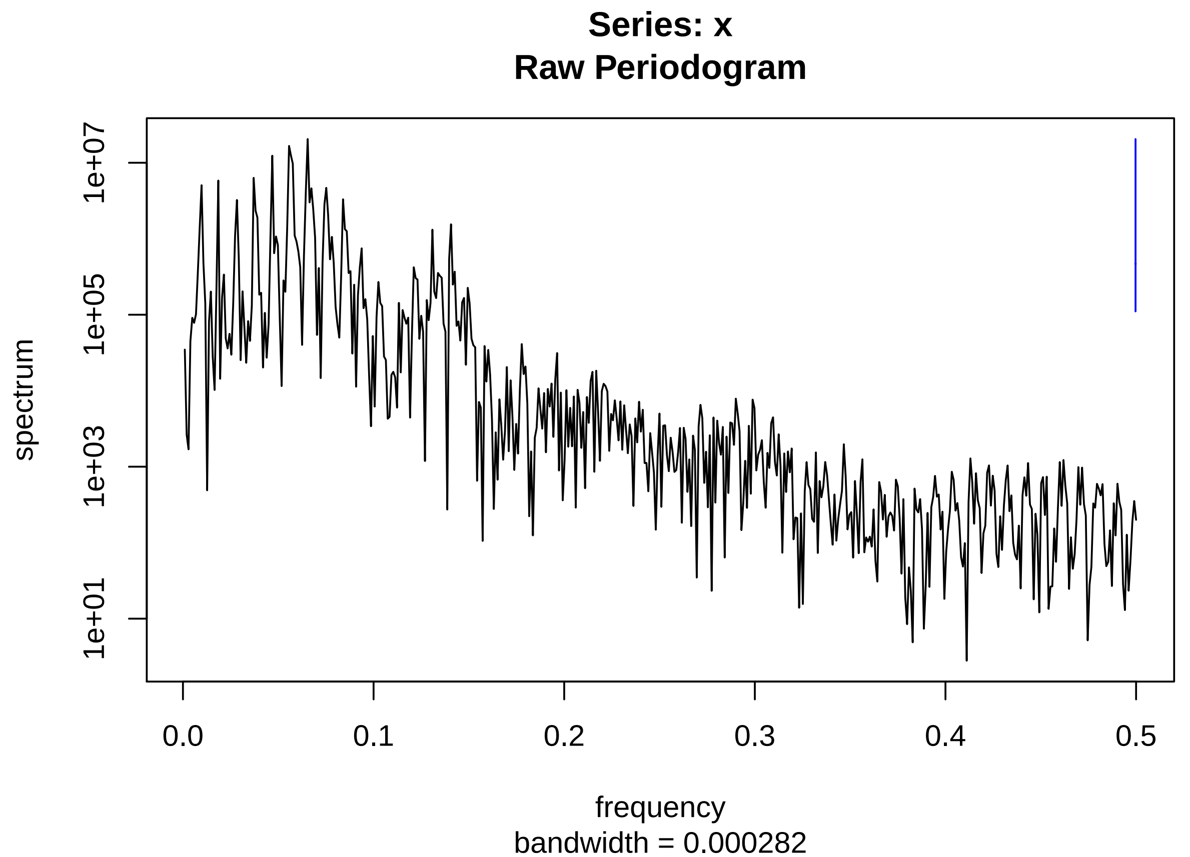
(a)

Calculate and display the log-periodogram of the data. Is the periodogram periodic,as predicted?

```
rm(list = ls())
data(speech, package = "astsa")
df <- speech
plot(df, main = "1 second (1000 points) sample of recorded speech for the phrase 'aaa...
```

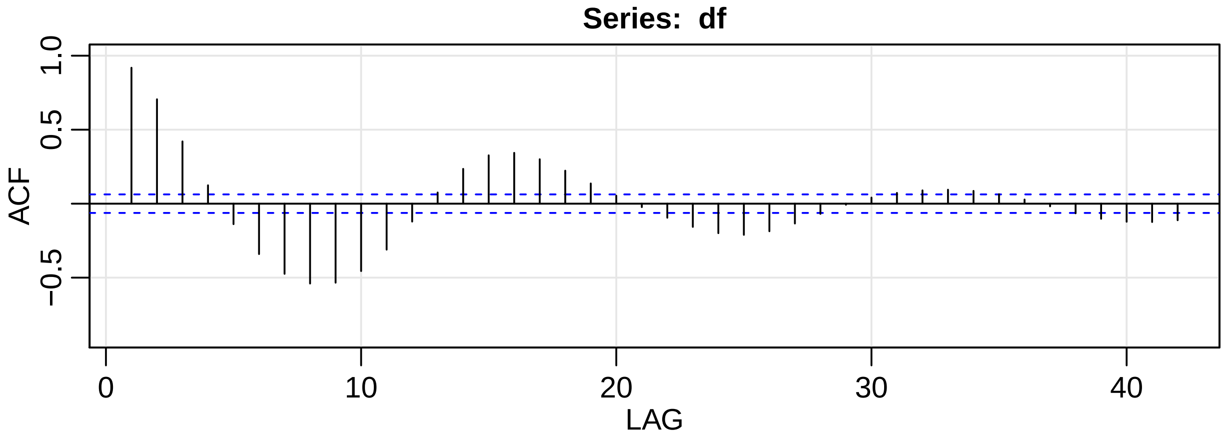


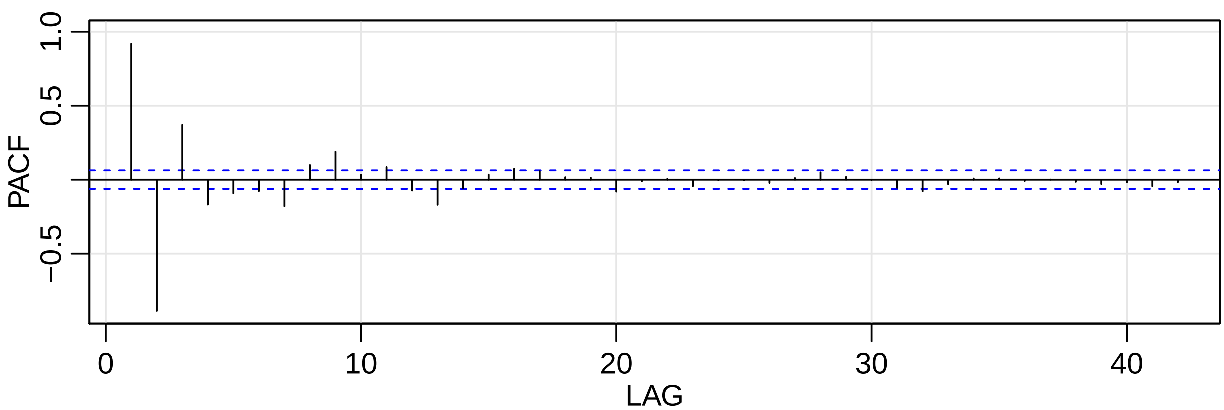

```
df.spec <- spectrum(df, log = "yes")
```



The log periodogram is periodic. We can see this structure in the ACF as well.

```r
library(astsa)
invisible(acf2(df))
```





(b)

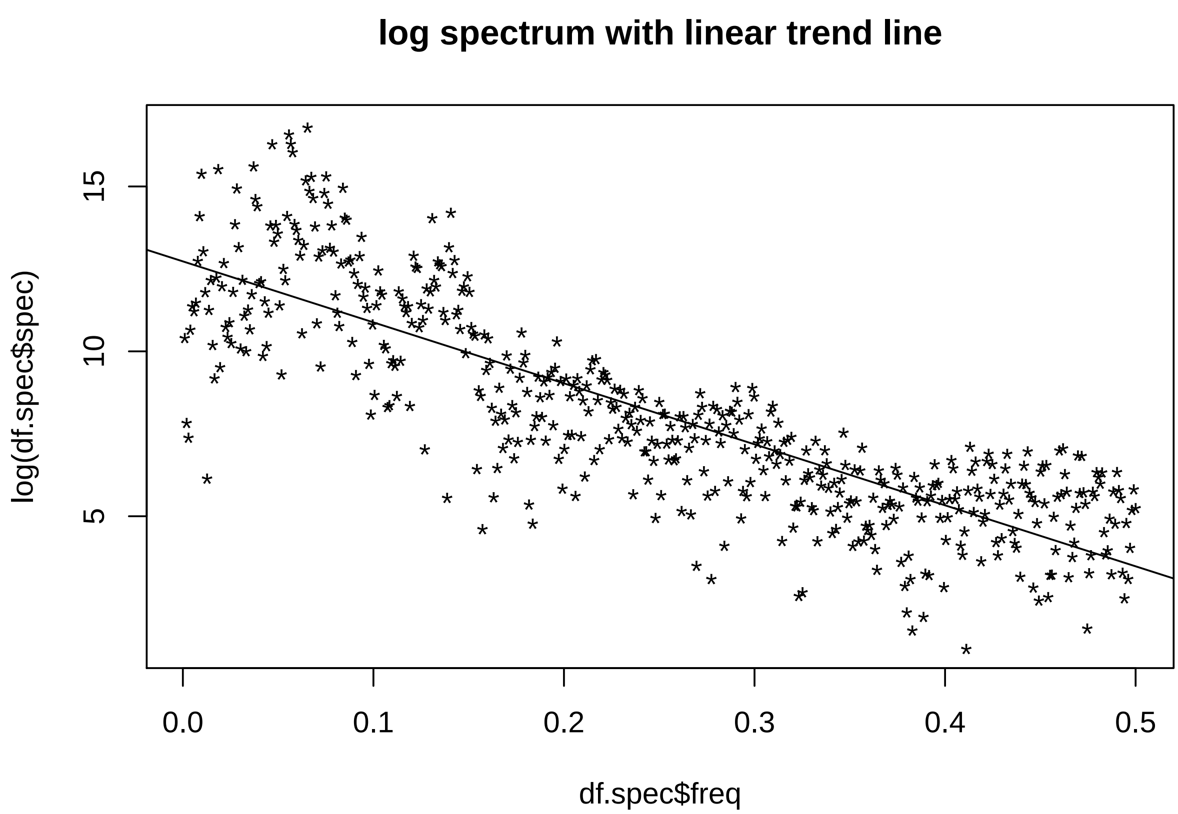
Perform a cepstral (spectral) analysis on the detrended logged periodogram, and use the results to estimate the delay D. How does your answer compare with the analysis of Example 1.27, which was based on the ACF?

```r
df.fs <- data.frame(spec = log(df.spec$spec), freq = df.spec$freq)

lm.fit <- lm(spec ~ freq, df.fs)
summary(lm.fit)
```

```
##
## Call:
## lm(formula = spec ~ freq, data = df.fs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3014 -1.1462  0.0687  1.0819  5.3180
##
## Coefficients:
```

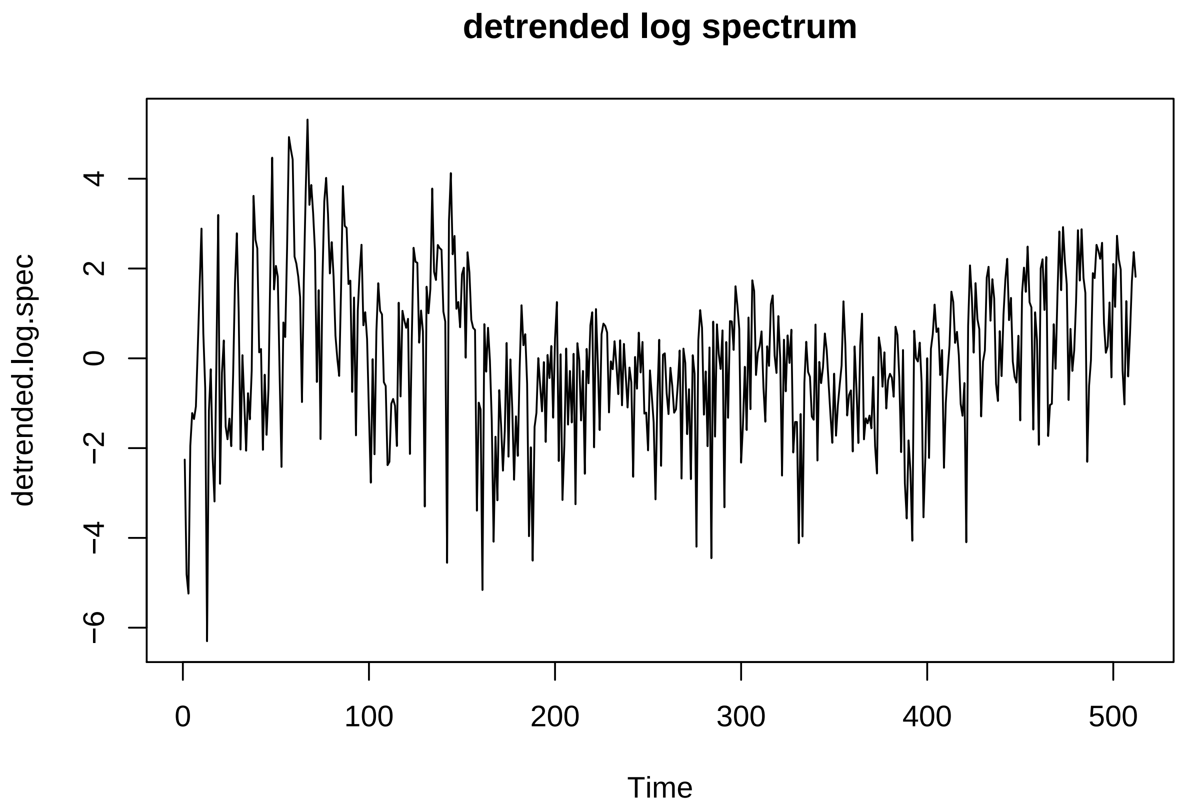
```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.7285     0.1551   82.06   <2e-16 ***
## freq        -18.4939     0.5365  -34.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.752 on 510 degrees of freedom
## Multiple R-squared:  0.6997, Adjusted R-squared:  0.6991
## F-statistic:  1188 on 1 and 510 DF,  p-value: < 2.2e-16
```

```r
df.detrended <- df.spec

plot(df.spec$freq, log(df.spec$spec), pch = "*", main = "log spectrum with linear trend
abline(lm.fit)
```

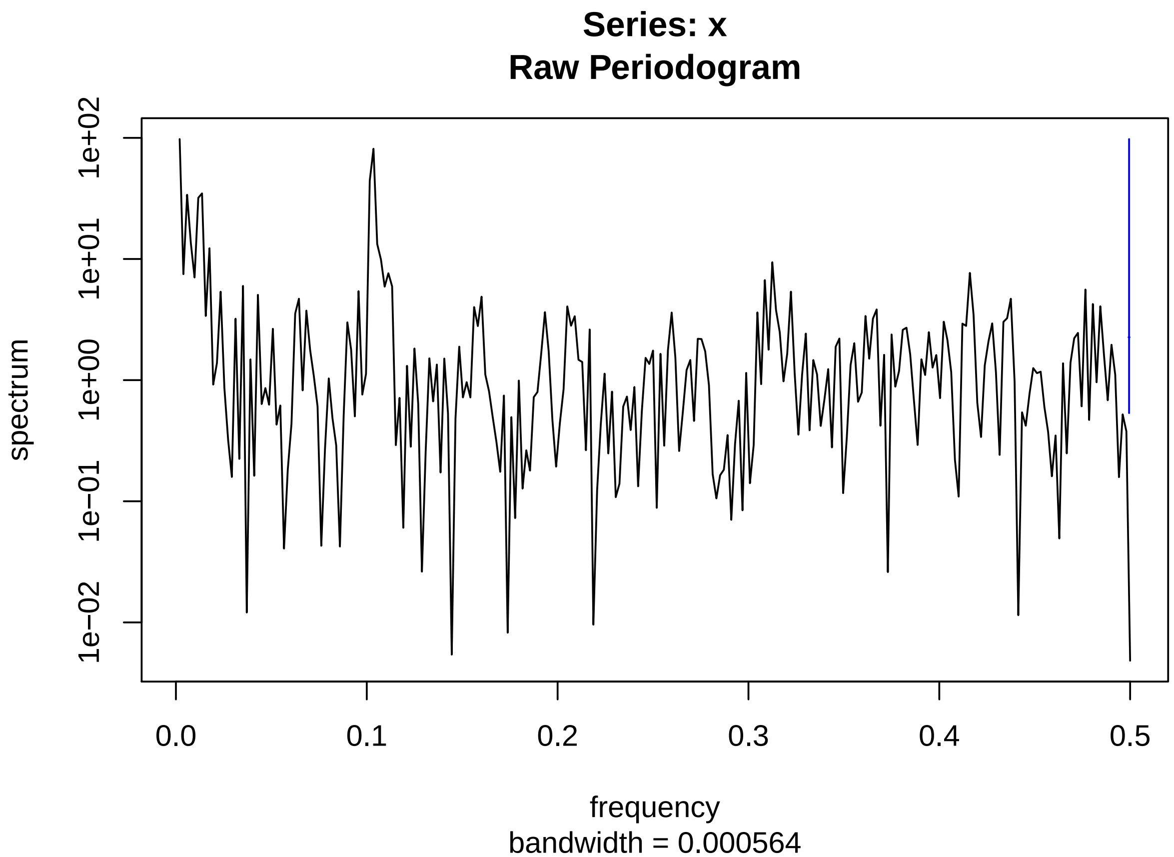


Now we subtract the trend and calculate the spectrum of the detrended log spectrum.

```r
detrended.log.spec <- ts(log(df.spec$spec) - predict(newdata = df.fs, lm.fit))
plot(detrended.log.spec, main = "detrended log spectrum")
```




```
spc.detrended.log.spec <- spectrum(detrended.log.spec, log = "yes")
```



We see the pitch period D located around 0.1, or 10Hz.

```
pander(data.frame(omega = spc.detrended.log.spec$freq[53], power = spc.detrended.log.spe
    caption = "Frequency and Power of Pitch Period D")
```

Table 3: Frequency and Power of Pitch Period D

omega	power
0.1035	81.27

4.18 Consider two time series

$x_t = w_t - w_{t-1}$, $y_t = \frac{1}{2}(w_t + w_{t-1})$, formed from the white noise series w_t with variance $\sigma^2 = 1$.

(a)

Are x_t and y_t jointly stationary? Recall the cross-covariance function must also be a function only of the lag h and cannot depend on time.

Each series is stationary and the cross covariance function is a function of the lag only so these time series are jointly stationary. We omit the computations here, but the cross covariance is given by

$$\gamma_{x\,y}(h) = \begin{cases} \frac{1}{2}\sigma^2 & |h = -1 \\ 0 & |h = 0 \\ -\frac{1}{2}\sigma^2 & |h = 1 \\ 0 & |h \neq 0, -1, 1 \end{cases}$$

(b)

Compute the spectra $f_y(\omega)$ and $f_x(\omega)$, and comment on the difference between the two results.

Using our expression for the spectra of an ARMA process and keeping in mind that $\theta_x(z) = 1 - z$ and $\theta_y(z) = \frac{1}{2} + \frac{1}{2}z$we have

$$f_x(\omega) = \sigma^2|\theta_x(e^{-2\pi i\omega})|^2 = \sigma^2|1 - e^{-2\pi i\omega}|^2 = 2\sigma^2(1 - cos(2\pi i\omega))$$

$$f_y(\omega) = \sigma^2|\theta_x(e^{-2\pi i\omega})| = \frac{\sigma^2}{2}|1 + e^{-2\pi i\,\omega}| = \sigma^2(1 + cos(2\pi i\omega))$$

(c)

Suppose sample spectral estimators $\bar{f}_y(.10)$ are computed for the series using $L = 3$. Find a and b such that

$$P\{a \leq \bar{f}_y(0.10) \leq b\} = 0.90$$

This expression gives two points that will contain 90% of the sample spectral values. Put 5% of the area in each tail.

Appendix C.2 is pretty technical. If you know some analysis we can show that

$$2L\bar{f}_y(\omega) \sim \chi^2_{2L}$$

This at least makes intuitive sense since we're summing squares of normal random variable to estimate the periodogram. Now the pivot can be used to form a probability statement for $\bar{f}_y(\omega)$

$$P[\frac{2L}{\chi^2_{2L}(1-\frac{\alpha}{2})} \leq \bar{f}_y(\omega) \leq \frac{2L}{\chi^2_{2L}(\frac{\alpha}{2})}] = 1-\alpha$$

We calculate this in the code below

```
omega <- 0.1
L <- 3
ch.right.tail <- qchisq(1 - 0.05, 2 * L)
ch.left.tail <- qchisq(0.05, 2 * L)

a <- 2 * L/ch.right.tail

b <- 2 * L/ch.left.tail

pander(data.frame(a = a, b = b), caption = "Probability Mass Limits")
```

Table 4: Probability Mass Limits

a	b
0.4765	3.669

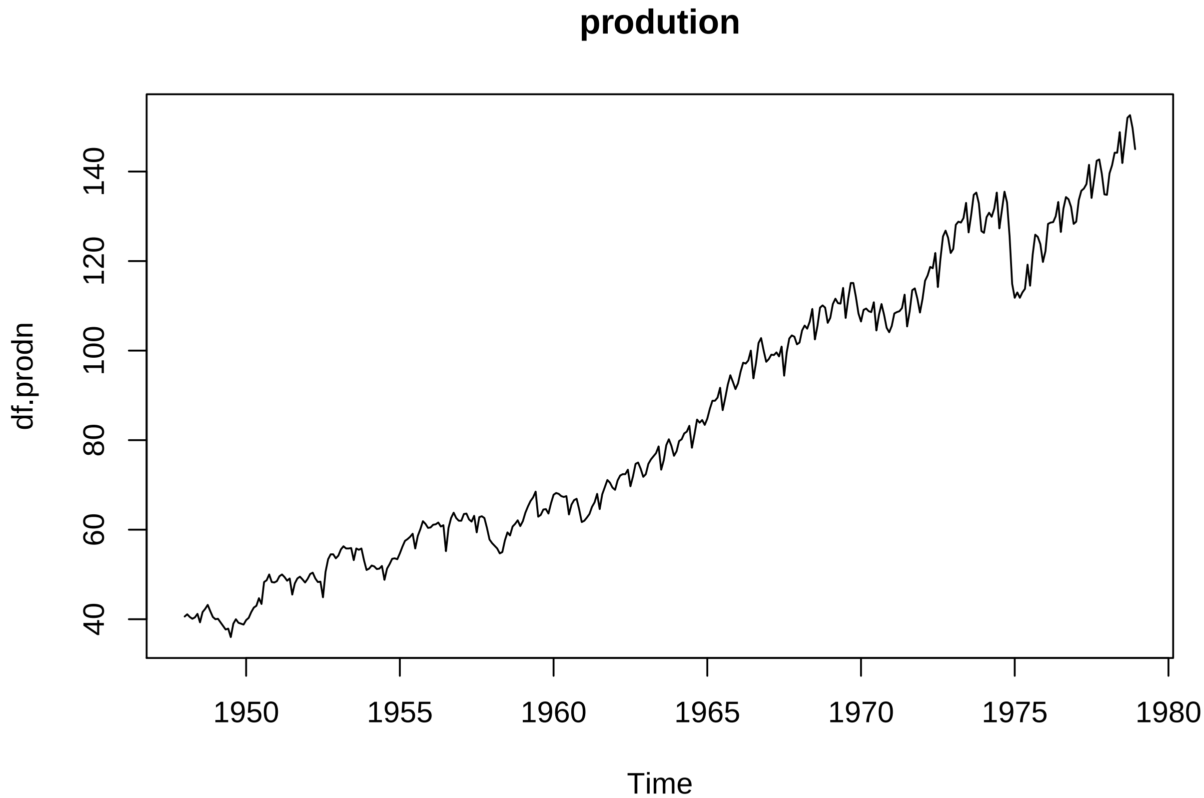
4.27 Production vs Unemployment Analysis

Consider the bivariate time series records containing monthly U.S. production (prodn) as measured by the Federal Reserve Board Production Index and the monthly unemployment series (unemp).

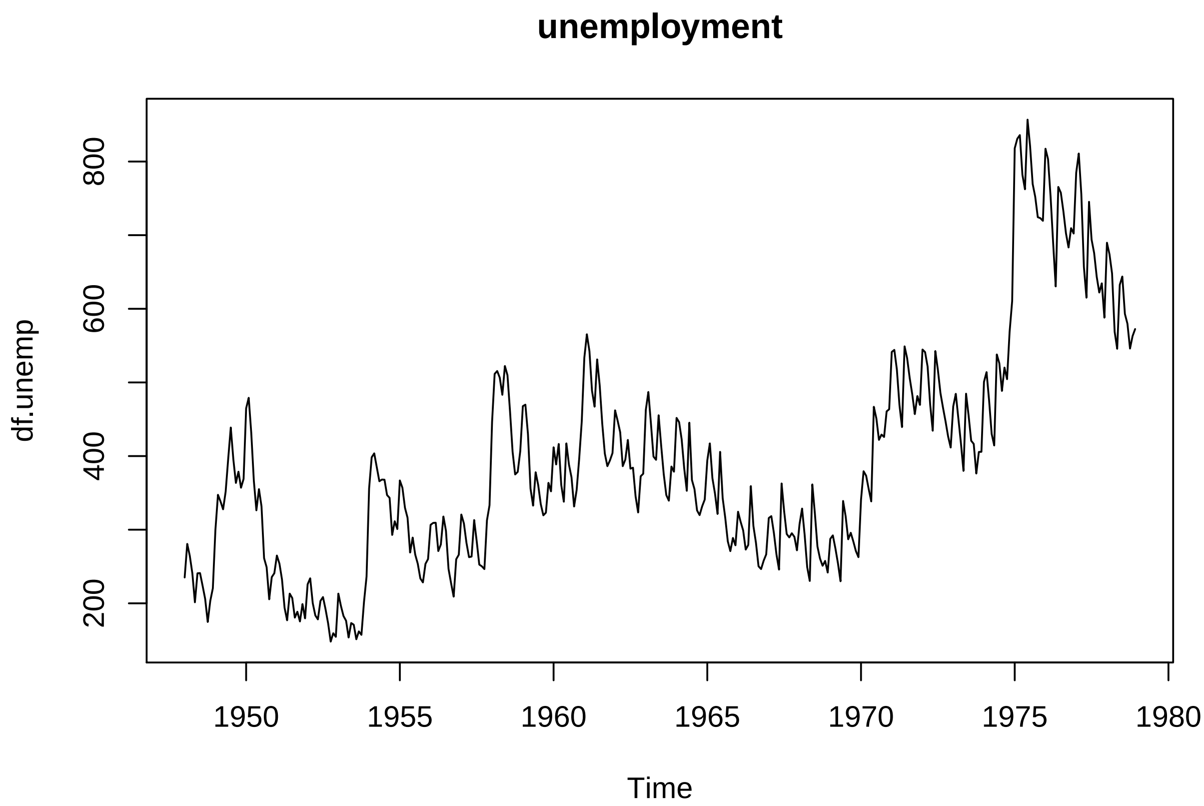
(a)

Compute the spectrum and the log spectrum for each series, and identify statistically significant peaks. Explain what might be generating the peaks. Compute the coherence, and explain what is meant when a high coherence is observed at a particular frequency.

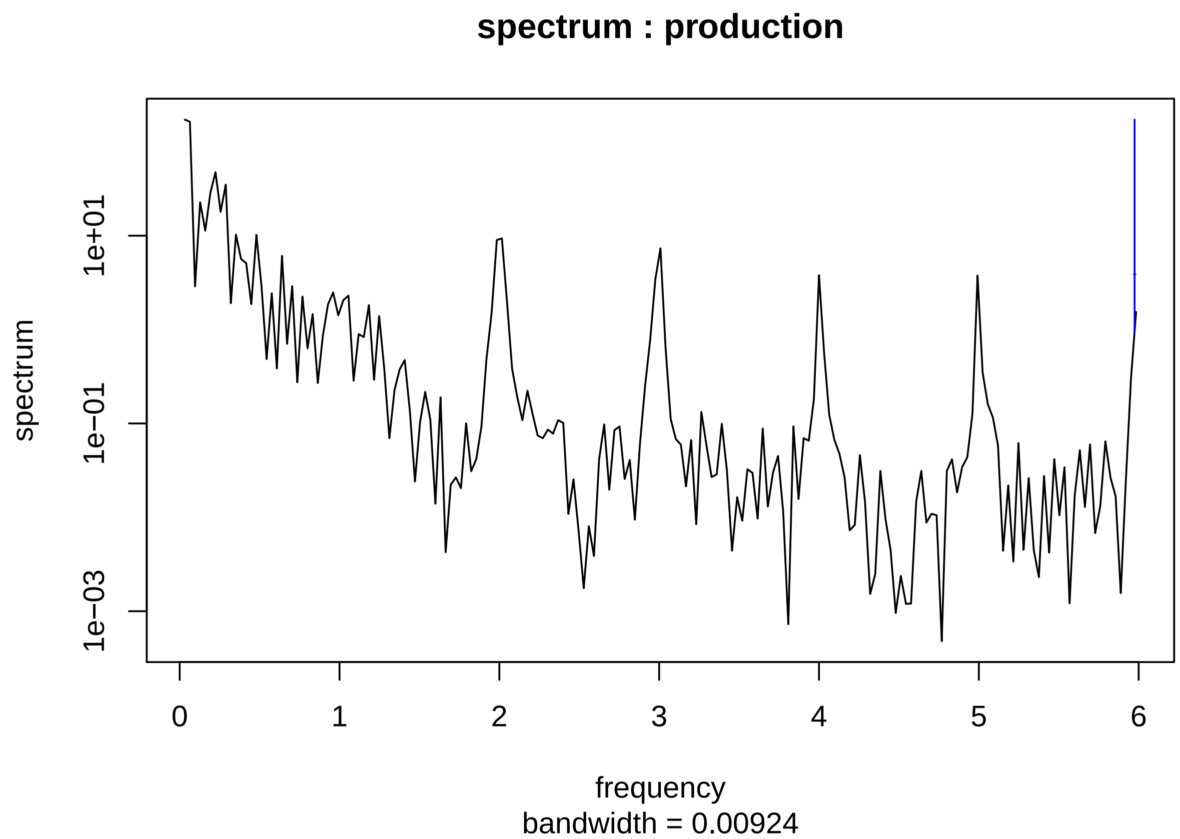
```
rm(list = ls())
library(astsa)
data(prodn, package = "astsa")
df.prodn <- prodn
data(unemp, package = "astsa")
df.unemp <- unemp
plot(df.prodn, main = "prodution")
```



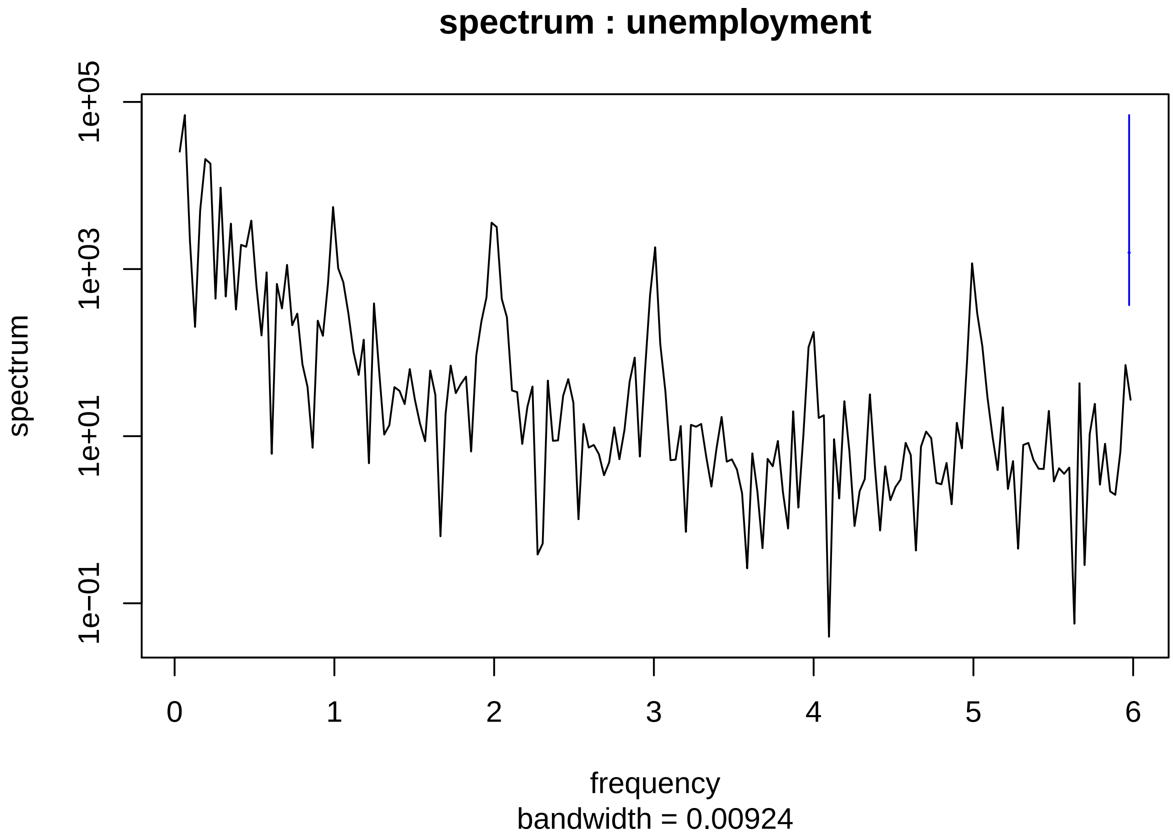

```r
plot(df.unemp, main = "unemployment")
```



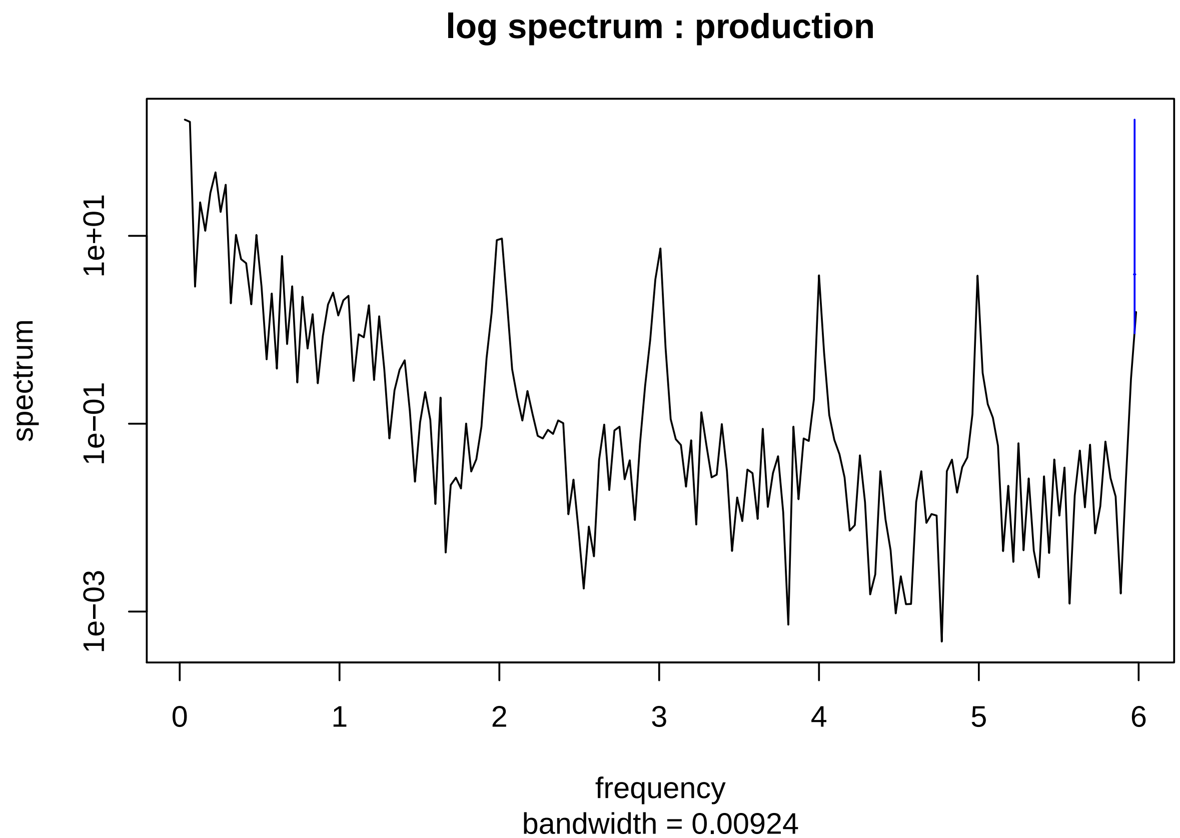

```
spec.prodn <- spectrum(df.prodn, main = "spectrum : production")
```



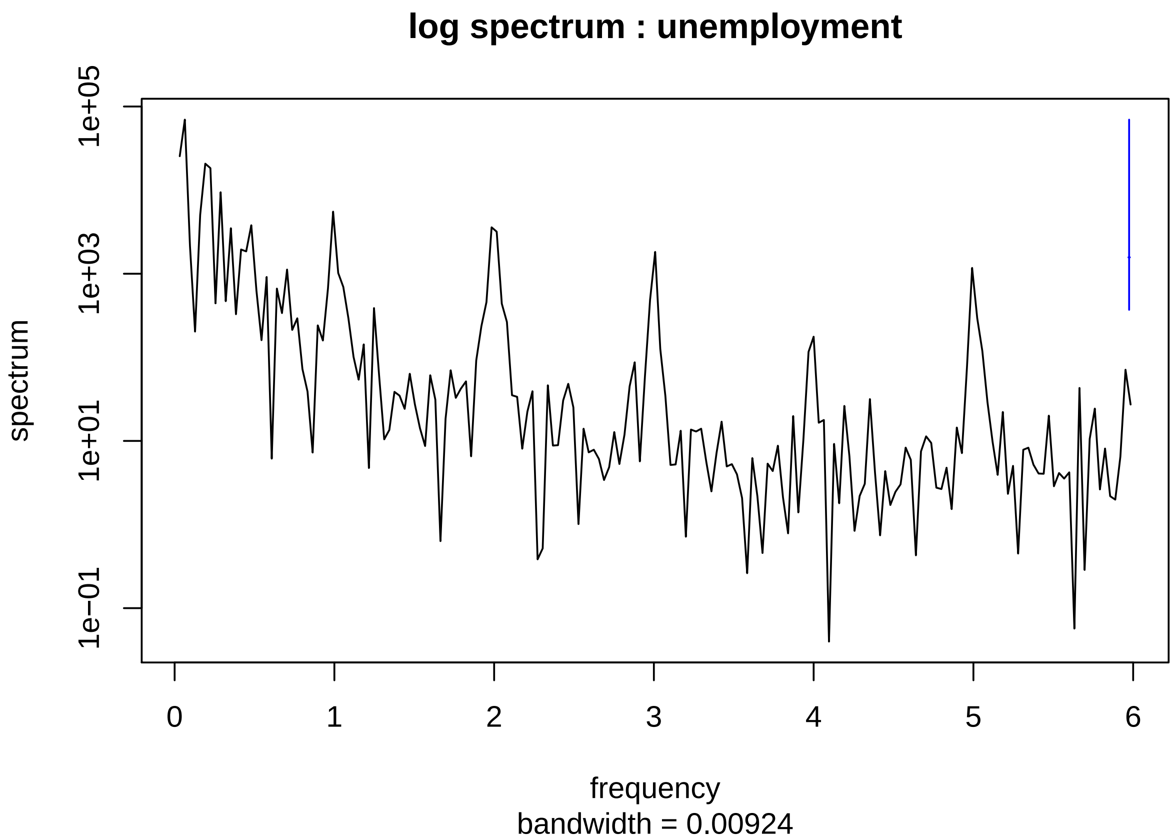

```
spec.unemp <- spectrum(df.unemp, main = "spectrum : unemployment")
```




```
spec.prodn <- spectrum(df.prodn, log = "yes", main = "log spectrum : production")
```




```
spec.unemp <- spectrum(df.unemp, log = "yes", main = "log spectrum : unemployment")
```

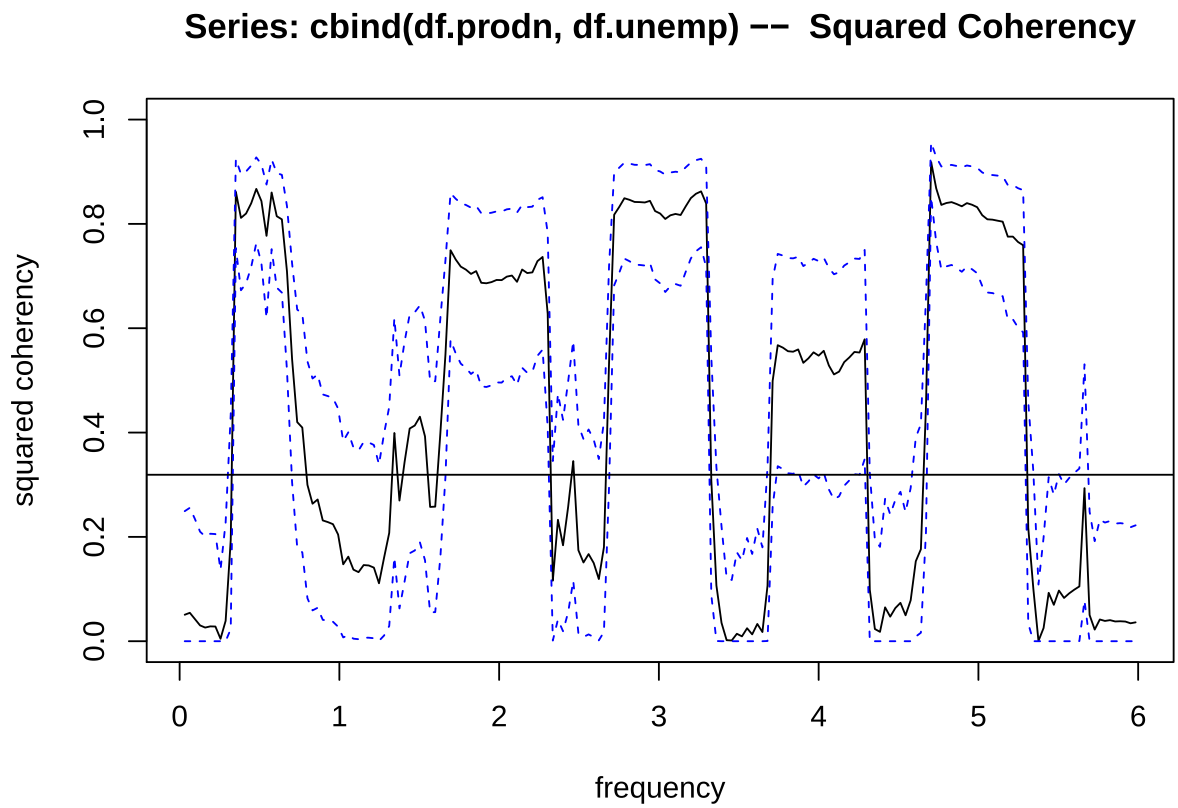


Economic cycles such as recession would be the sources of the strong peaks.

```
sr = spec.pgram(cbind(df.prodn, df.unemp), kernel("daniell", 9), taper = 0,
    plot = FALSE)
sr$df
```

```
## [1] 37.696
```

```
f = qf(0.999, 2, sr$df - 2)
C = f/(18 + f)
plot(sr, plot.type = "coh", ci.lty = 2)
abline(h = C)
```



The areas of large coherence represent frequency ranges where we may predict one series from another via a linear relationship such as a filter. If one series lags another at a particular frequency the phase of the two components are different. The series with the larger phase can be used to predict the other. Keeping mind that we're talking about projecting the series onto the Fourier basis function for a particular ω and using that.

The plot above displays coherence with the line indicating a significance level of $\alpha = 0.001$

(b)

What would be the effect of applying the filter $u_t = x_t - x_{t-1}$ followed by $v_t = u_t - u_{t-12}$ to the series given above? Plot the predicted frequency responses of the simple difference filter and of the seasonal difference of the first difference.

In applying u_t we'd be attenuating the low frequency components in the time series. Applying v_t would attenuate the components around the seasonal signal. If there we're a peak in the spectrum at the ω corresponding to a period 12, the filter v_t would suppress that. We can calculate the frequency response of the filter from

Using the notation from the book, and that u_t has coefficients $a_0 = 1,\ a_1 = -1$ and all other $a_j = 0$ in the linear filter, we have

$$|A_{xu}(\omega)| = \sum_{j=-\infty}^{j=+\infty} a_j e^{-2\pi\omega ij} = (1 - e^{-2\pi\omega i})$$

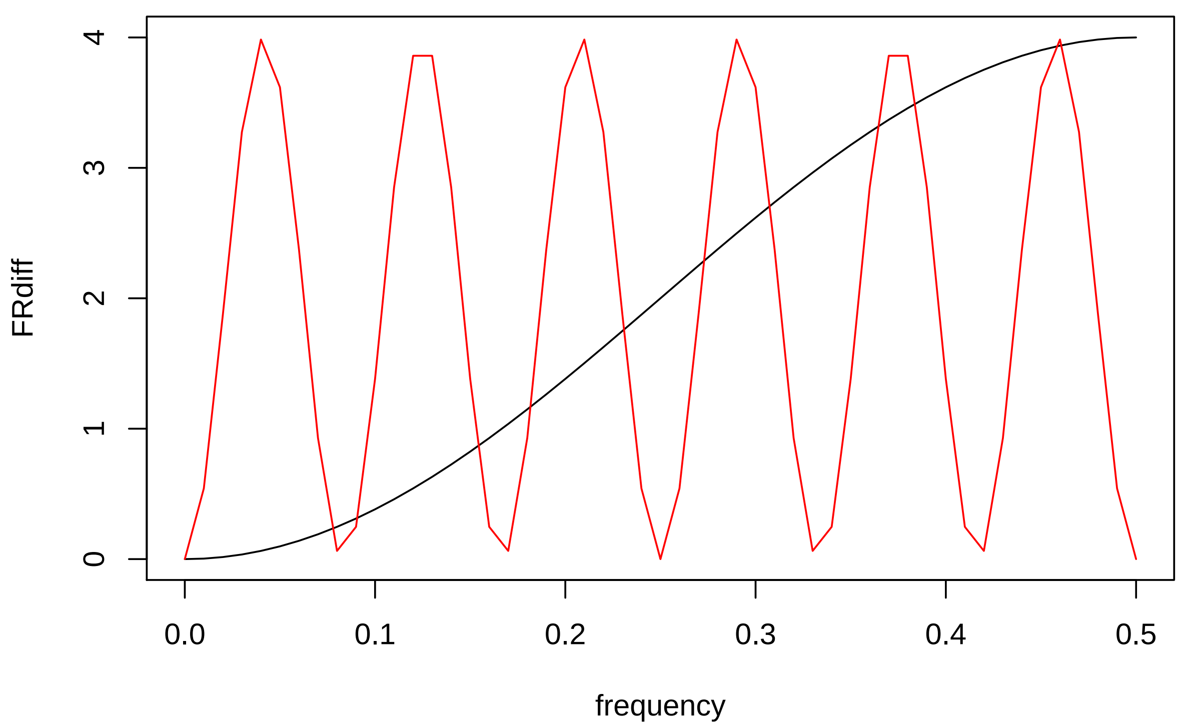
$$|A_{xu}(\omega)|^2 = (1 - e^{-2\pi\omega i}) \times (1 - e^{2\pi\omega i}) = 2(1 - cos(2\pi\omega))$$

Similar arguments for ν_t give

$$|A_{x\nu}(\omega)|^2 = 2(1 - cos(2\pi 12\omega))$$

The frequency responses are displayed below.

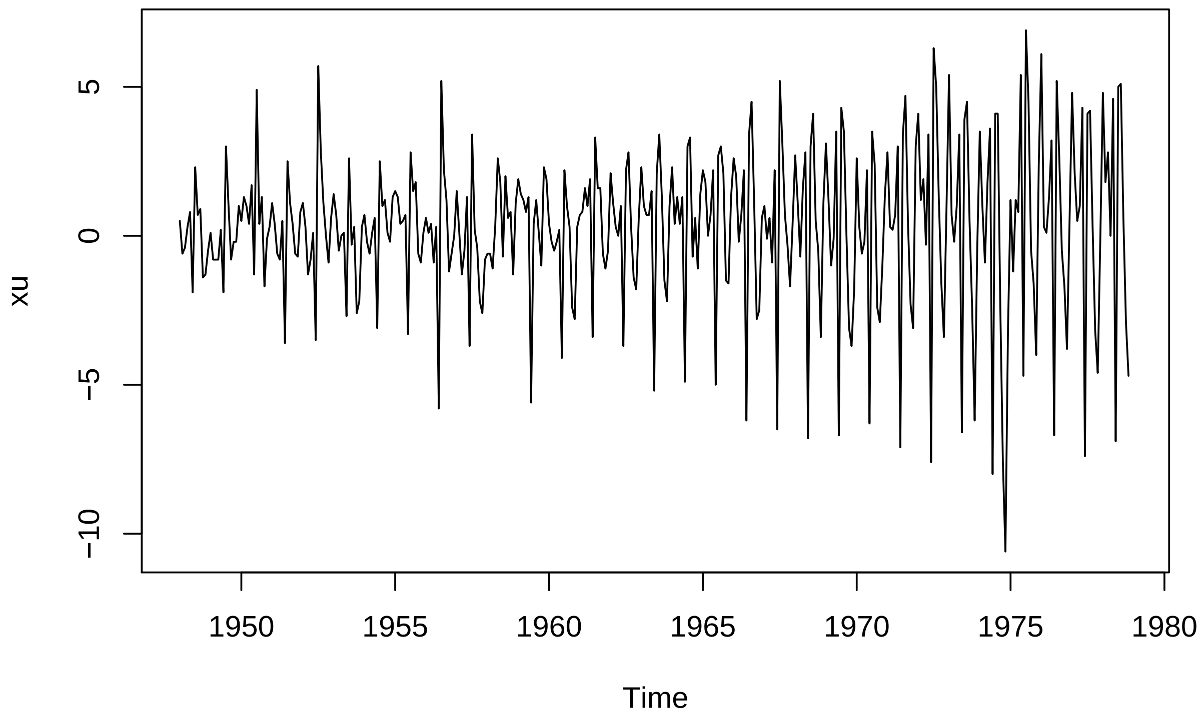
```
w = seq(0, 0.5, by = 0.01)
FRdiff = abs(1 - exp((0+2i) * pi * w))^2
FRdiff12 = abs(1 - exp((0+2i) * pi * 12 * w))^2
plot(w, FRdiff, col = 1, type = "l", xlab = "frequency")
lines(w, FRdiff12, col = 2)
```



(c)

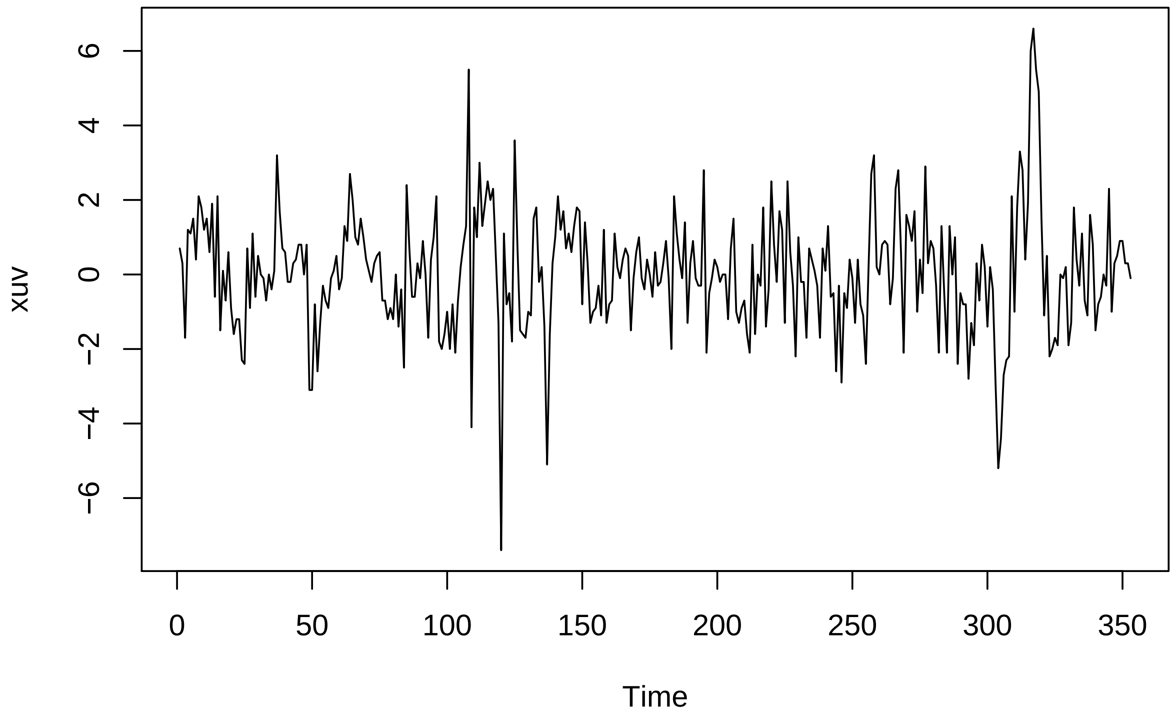
Apply the filters successively to one of the two series and plot the output. Examine the output after taking a first difference and comment on whether stationary is a reasonable assumption. Why or why not? Plot after taking the seasonal difference of the first difference. What can be noticed about the output that is consistent with what you have predicted from the frequency response? Verify by computing the spectrum of the output after filtering

```
xu <- filter(df.prodn, filter = c(1, -1))
plot(xu)
```



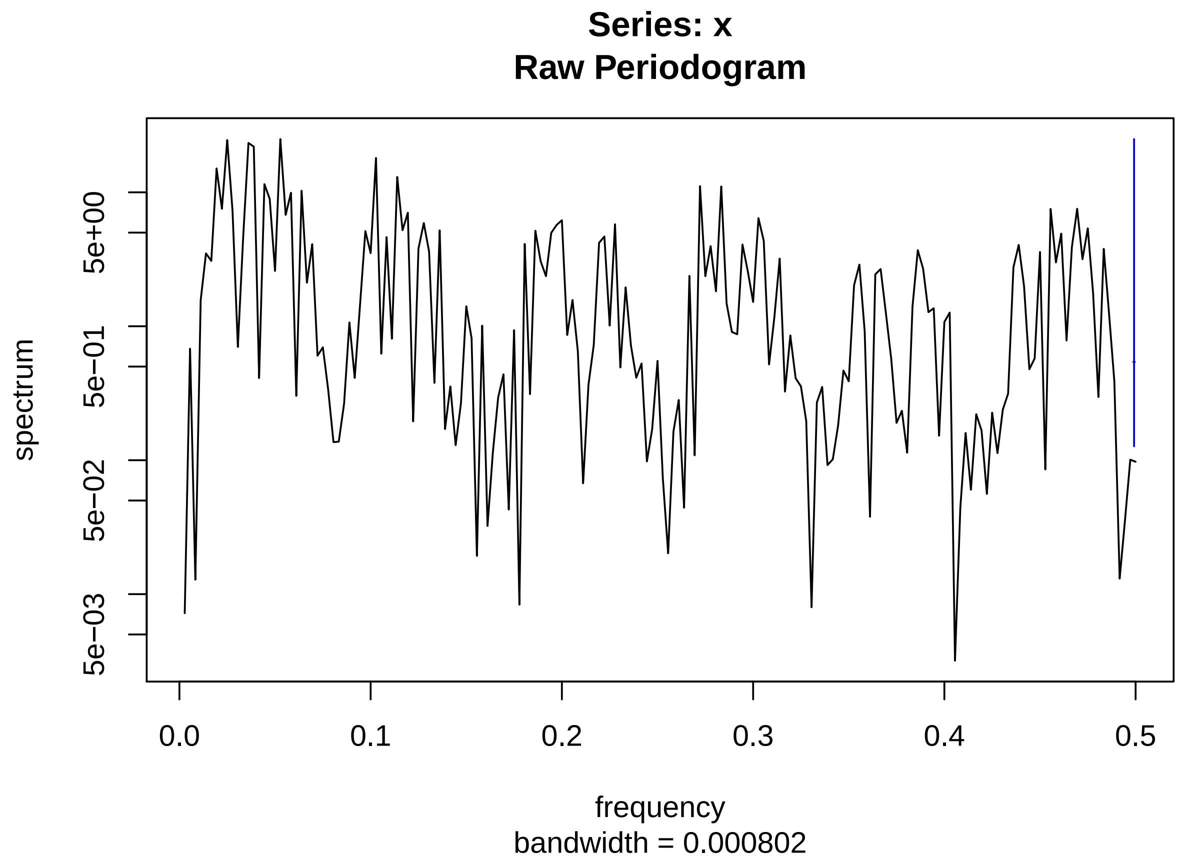
We see the variance is not stable - second order stationarity does not hold for this data. Now we filter the seasonal component.

```
xuv <- filter(xu, filter = c(1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1))
xuv <- xuv[-(1:12)]
xuv <- ts(xuv[-(354:360)])
plot(xuv)
```



We seem to have removed the seasonal component of the series. Let's look at the spectrum to verify this.

```r
xuv.spec <- spectrum(xuv)
```


Series: x
Raw Periodogram
spectrum
5e+00
5e−01
5e−02
5e−03
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.000802