Stat 510, Lab 10.5

April 27, 2007

1 In Class

1.1 Smoothing

Let's look at the Australian labour data which was seen in previous labs. We would like to estimate the spectral density of the data. We may generate the raw periodogram using the following code:

```
labour=scan("labour.dat")
labour=labour[1:144]
tlabour= diff(diff(labour),12)
labourpgram=spec.pgram(tlabour,taper=0, log="no")
```

It is pretty hard to tell what's going on with the unsmoothed periodogram. We will start by adding some smoothing with the following command:

```
spec.pgram(tlabour, spans=3, taper=0, log="no")
```

The argument "spans" corresponds to the "L" that we've been using in class which means that this will be the total window size used in the argument. By default the modified Daniell kernel is used. If we would like to use one expansion of this kernel, then we can use the command

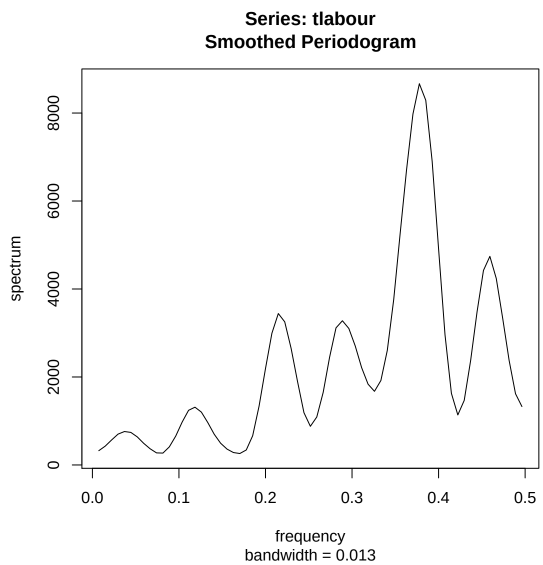
```
spec.pgram(tlabour, spans=c(3,3), taper=0, log="no")
```

This yields a total number of five coefficients, so the smoothing window is five. We can obtain which coefficients are being used with the following command:

```
spec.pgram(tlabour, spans=c(3,3), taper=0, log="no")$kernel
```

A practical suggestion would be to keep the numbers the same, and try different odd sized windows. What happens with values of 3, 5, 7, 9, and 11.

Figure 1:



1.2 taper

We can see in this example that `spans=c(5,5)` gives a reasonable smoothing for this periodogram (See figure 1). What are the effects of tapering on this estimate?

What happens if we put in just a bit of tapering? In order to do 20% tapering, I will put in the following command

```
\spec.pgram(tlabour, spans=c(5,5), taper=0.1, log="no")
```

The argument in the R command uses `taper=0.1` to do 20% tapering. Notice in figure 3 with 40% tapering that there is a somewhat strong effect here. Is there something particular about this data that would cause tapering to yield such a dramatic effect?

Now let's look at full tapering in figure 4. The effect compared to no tapering is stark. Given how dynamic the smoothed periodogram appears and the substantial changes generated by various levels of tapering the data, this data requires a substantial amount of tapering for accuracy. Remember, however, that too much tapering causes an increase in the variance of our estimate. We may want to stick with a substantial but not a full taper.

Figure 2:

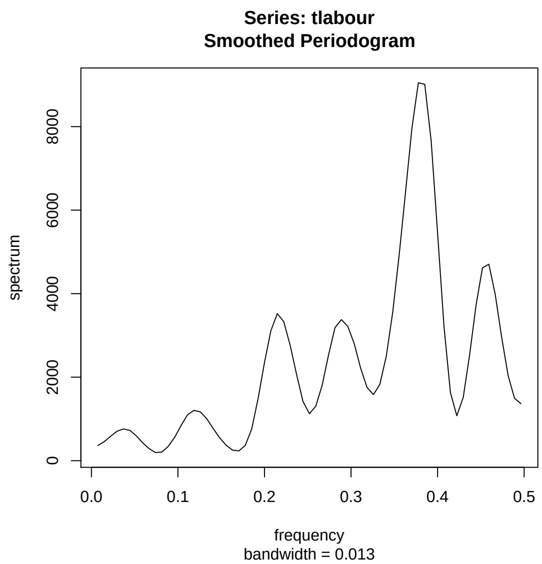


Figure 3:

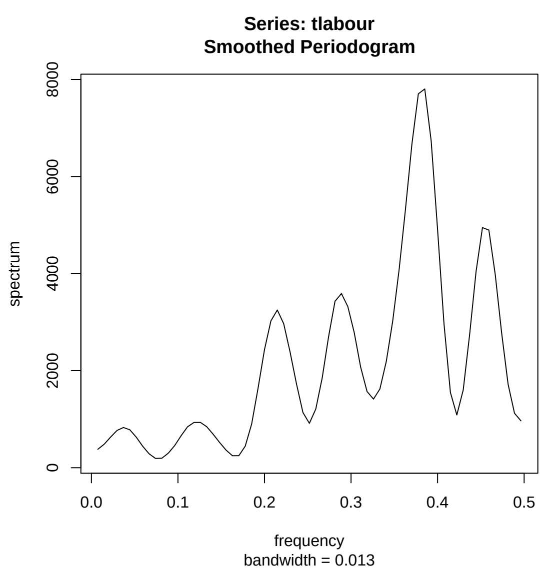
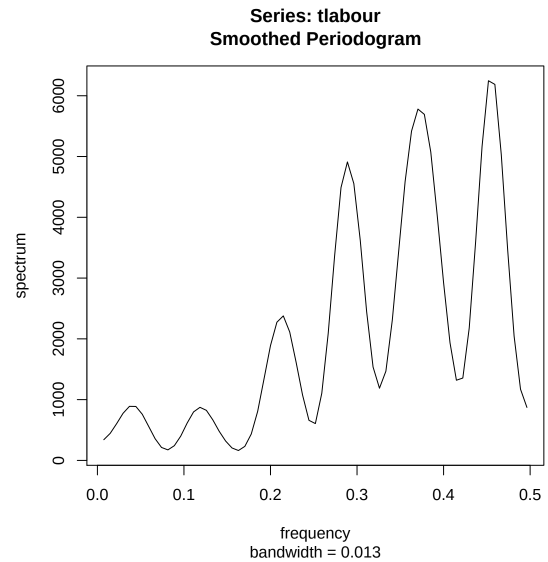


Figure 4:



2 Homework

The first three problems are to be done with NO tapering.

1. In this exercise, we will simulate time series that follows an AR model with $\phi_1 = 0.8$ (with $\sigma^2 = 1$). We can use the following command to simulate observations from this model:

   ```
   x=arima.sim(n = 200, list(ar = c(0.8), ma = 0), sd = 1)
   ```

   There is a closed form representation for the spectral density of ARMA models. So, we will try to determine how much smoothing is necessary to obtain a good representation for the spectral density. Load the functions found in "lab10function.R". We may now use the following command to obtain a plot of the periodogram (with some smoothing) along with the TRUE spectral density:

   ```
   specplotter( x,spans=c(3,3),phi=c(0.8))
   ```

   This R function when we know the true underlying model. With real data, we will not have the ability to look at the real spectral density.

(a) Introduce smoothing into this setup. How much smoothing is required to obtain a fairly good estimate of the spectral density without over smoothing? Are there certain features of the true spectral density that are difficult to obtain?

(b) Simulate a new time series with the same model but with 400 data points. Repeat attempts to smooth the periodogram. Are there differences here with more data?

(c) Now simulate a data set of 400 points from an ARMA model with coefficients $\phi_1 = 0.6$,$\phi_2 = -0.2$, and $\theta_1 = -0.7$. How much smoothing is necessary to obtain a good estimate of the spectral density.

solutions:

(a) For the AR model with 200 points. The periodogram plots associated with my simulation are below in figures 5 and 6. These will differ depending on each person's simulation. It is especially difficult to estimate the low frequencies because of the asymptote in the low frequencies. This feature is difficult to obtain. The plots represent the use of the option `spans=c(i,i)` where i is 3, 5,...,17 going from left to right then top to bottom. In my opinion, the sufficient amount of smoothing is around 11 (the top left plot of figure 6). There will be some variation in the opinions

(b) There is not a great deal of difference with more data and the same level of smoothing. This is because with increased data we have more points but if the smoothing window is fixed then the variance is fixed. I think the plot with 9 as an argument to the spans seems reasonable.

(c) The spectral density for this model has a somewhat pronounced maximum. It is more difficult (requires more smoothing) to fit the spectral density here. The plots are found in figures 9, 10, and 11. Things don't start to look reasonable until the final panel this represents span levels of 19,21, 23, and 25.

2. Estimate the spectral density airline data. Use the following to transform the data before beginning:

```
airline=scan("airline.dat")
tairline=diff(diff(log(airline),12))
```

Figure 5:

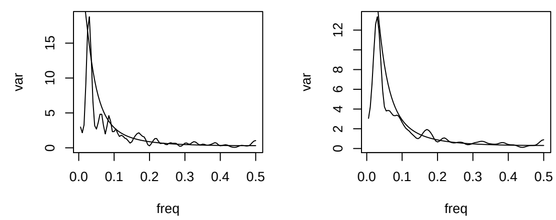


Figure 6:

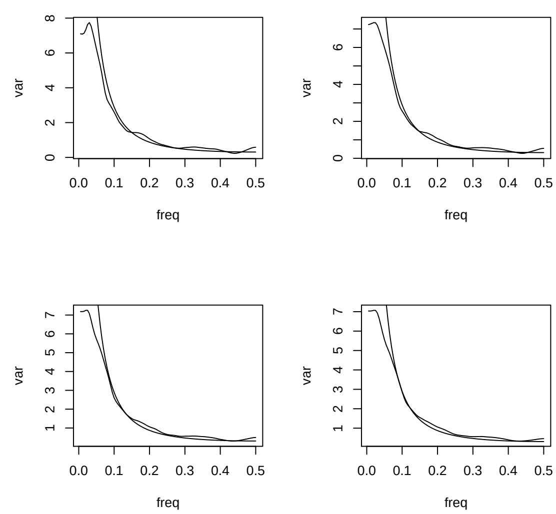


Figure 7:

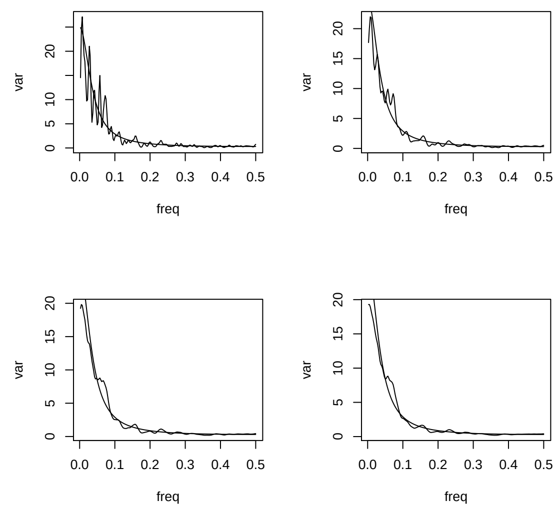


Figure 8:

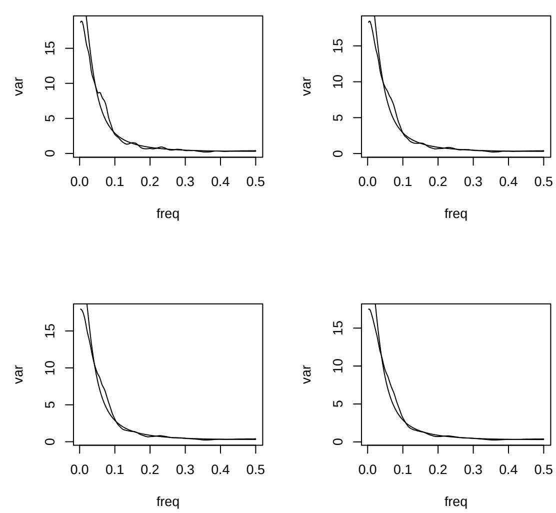


Figure 9:

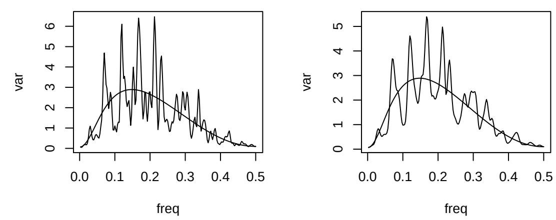


Figure 10:

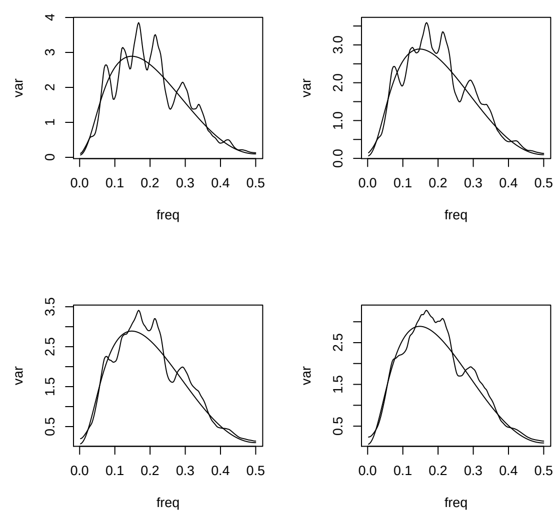
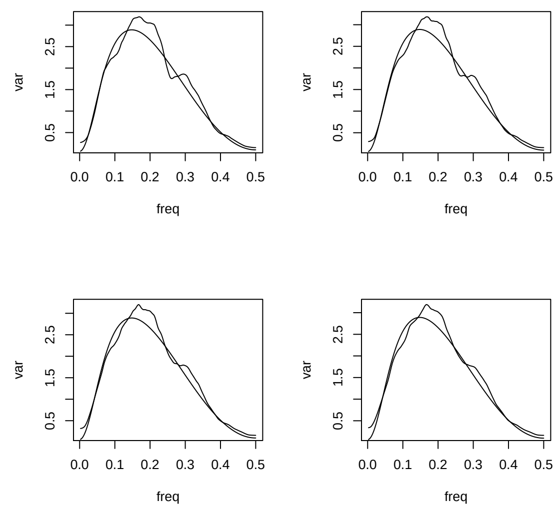


Figure 11:



solution: The smoothing of the periodogram is seen in figures 12 and 13. An argument could be made that spans of c(5,5) (upper right of 12 is sufficient, but I think that spans c(9,9) (lower right of 12is superior. The pattern of a single maximum is sustained over a fairly broad spectrum of smoothing.

3. Estimate the spectral density of the ozone data. Use the following to transform the data before beginning:

```
ozone=scan("ozone.dat")
tozone=diff(diff(log(ozone)),12)
```

   solution: Again an argument could be made that the spans of c(5,5) is sufficient. However, I think c(7,7) seems to take care of some of the smaller spurious features. The plots are found in figures 14 and 15

4. What effect does tapering have on the estimate of the spectral density you have found for number 2 above?

   solution: Selecting the smoothing of c(9,9). I have shown the effect of tapering in plot 16. The plot that is shown includes 0, 20%, 40%

Figure 12:

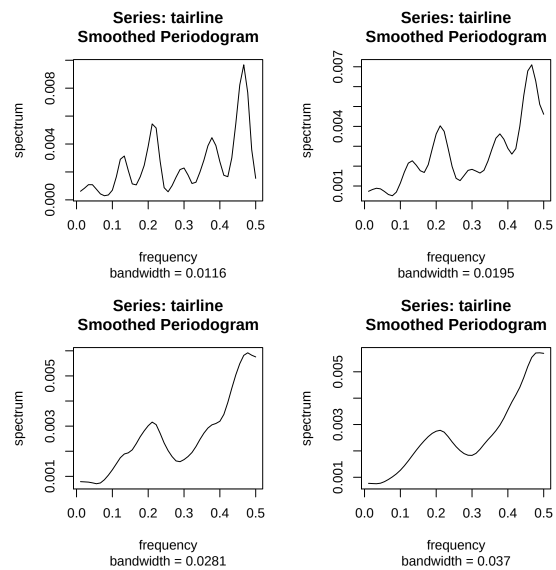


Figure 13:

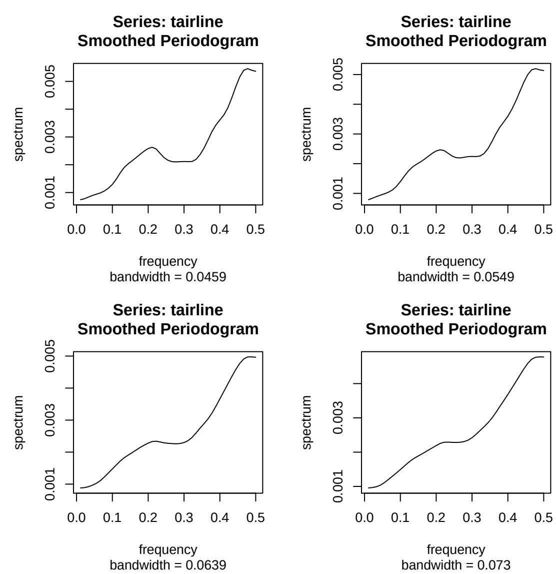


Figure 14:

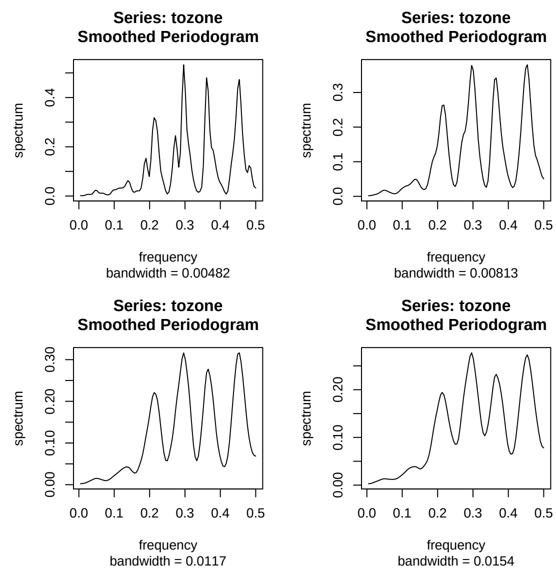


Figure 15:

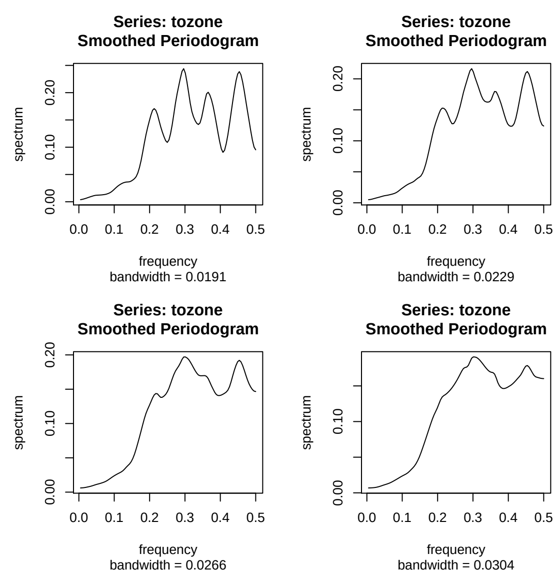
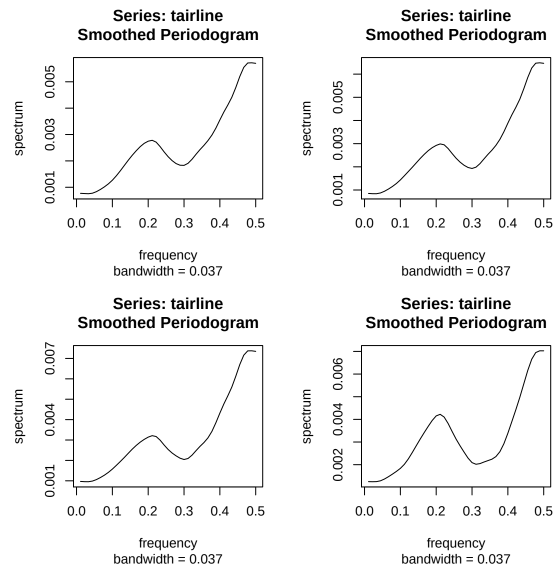


Figure 16:



and full tapering. The effect of tapering is very mild. This is to be expected in that the spectral density estimate is very mild in terms of local maxima and minima. Only a small amount of tapering is needed.

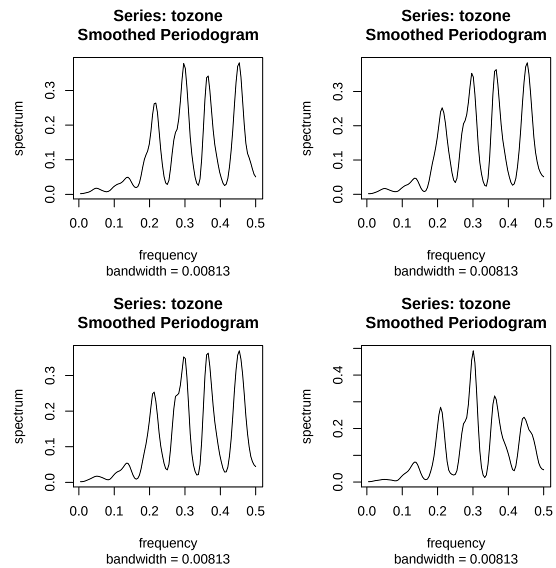
5. What effect does tapering have on the estimate of the spectral density you have found for number 3 above?

   solution: The effect of tapering on this data is rather strong. This is to be expected with such dramatic maxima and minima in the spectral density. A reasonable amount of tapering (perhaps 40%) would be in order. The effect of tapering is seen in 17.

6. What is the difference between the bias introduced by smoothing and the bias that is reduced by doing tapering?

   solution: By smoothing the periodogram, we are attempting to estimate the spectral density at each of the fundamental frequencies ω_j, i.e. $f(\omega_j)$. Smoothing introduces bias when we take a moving average over the neighbors of ω_j. In other words, we are averaging over $I(\omega_j)$ and it's neighboring periodogram values. Even if $E(I(\omega_j))$ were $f(\omega_j)$, the expectaion of the neighboring periodogram points will not

Figure 17:



be $f(\omega_j)$ (unless the spectral density if perfectly flat). The bias that is reduced by tapering is due to the fact that $E(I(\omega_j)) \neq f(\omega_j)$. The expected value of the periodogram is the integral of the spectral window over the spectral density–the spectral window introduces the bias. Tapering helps fix this bias.

7. Give a BROAD outline of what R needs to do to implement the following command assuming x is a stationary time series :

```
spec.pgram(x,taper=0.1,spans=3, log="no")
```

By broad, I mean you do not need to write down formulas. You may say things like “take the Fourier transform” etc.

solution: Here are the basic steps that R takes:

(a) Multiply the data by the appropriate transformation, h_t, to taper the data.

(b) Take the Fourier transform of the data.

(c) Calculate the periodogram by taking the modulus squared of the Fourier transform of the data.

(d) Take the moving average of the resulting periodogram to obtain the estimate of the spectral density.