Stat 463, Lab 10

December 7, 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:

Series: tlabour
Smoothed Periodogram

frequency
bandwidth = 0.013
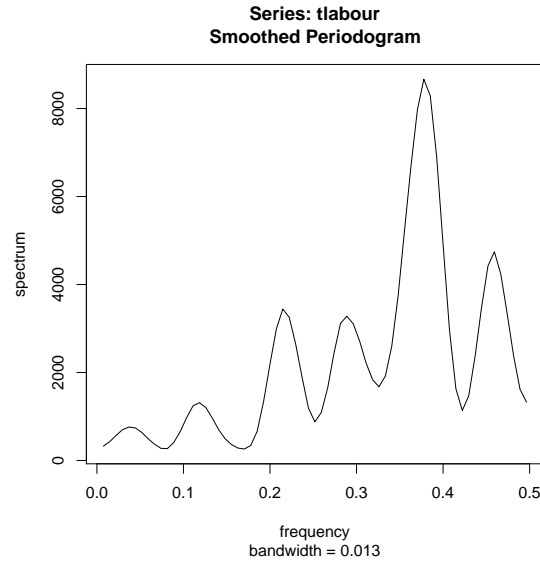
1.2   taper

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.
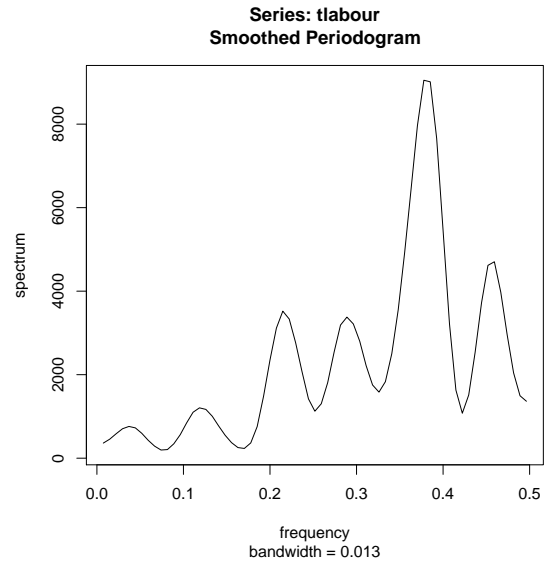
Figure 2:

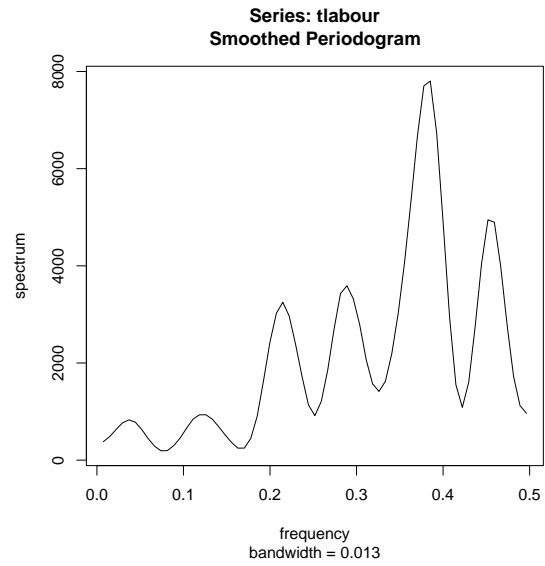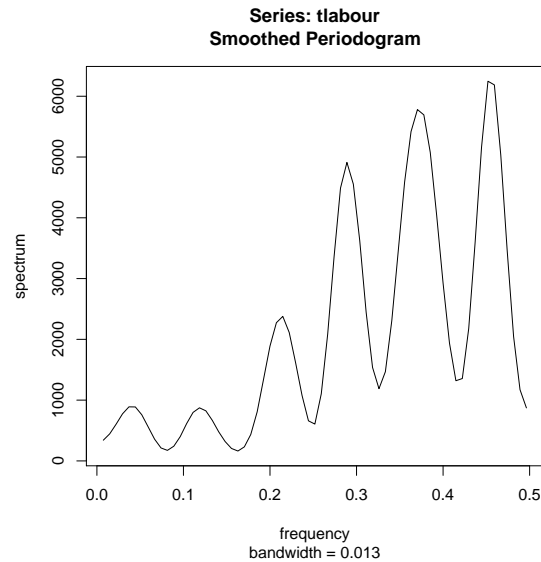**Series: tlabour**
Smoothed Periodogram

frequency
bandwidth = 0.013

Figure 3:

Series: tlabour
Smoothed Periodogram

frequency
bandwidth = 0.013

Figure 4:

Series: tlabour
Smoothed Periodogram

2 Homework

The first three problems are to be done with NO tapering. (You've already done that part.)

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 "lab9function.R". We may now use the following command to obtain a plot of the periodogram (with some smoothing) along with the TRUE spectral density:

   ```
   specplot( spec.pgram(x, spans=c(3,3), taper=0, log="no"),phi=c(0.8))
   ```

4

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.

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))
```

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)
```

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

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

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

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.