8. Smoothing in the time and frequency domains

Edward Ionides

2/7/2018

Contents

Objectives

- <mark>Estimating a nonparametric trend from a time series is known as smoothing</mark>. We will review some standard smoothing methods.

- We can also smooth the periodogram to estimate a spectral density.

- Many smoothers can be represented as linear filters. We will see that the statistical properties of linear filters for dependent (time-domain) stationary models can be conveniently studied in the frequency domain.

8.1 A motivating example

- The economy fluctuates between periods of rapid expansion and periods of slower growth or contraction.

- High unemployment is one of the most visible signs of a dysfunctional economy, in which labor is under-utilized, leading to hardships for many individuals and communities.

- Economists, politicians, businesspeople and the general public therefore have an interest in understanding fluctuations in unemployment.

- Economists try to distinguish between fundamental structural changes in the economy and the shorter-term cyclical booms and busts that appear to be a natural part of capitalist business activity.

- Monthly unemployment figures for the USA are published by the Bureau of Labor Statistics. Measuring unemployment has subtleties, which should be acknowledged but are not the focus of our current exploration.

```
system("head unadjusted_unemployment.csv",intern=TRUE)
```

```
##  [1] "# Data extracted on: February 4, 2016 (10:06:56 AM)"
##  [2] "# from http://data.bls.gov/timeseries/LNU04000000"
##  [3] "# Labor Force Statistics from the Current Population Survey"
##  [4] "# Not Seasonally Adjusted"
##  [5] "# Series title:        (Unadj) Unemployment Rate"
##  [6] "# Labor force status:  Unemployment rate"
##  [7] "# Type of data:        Percent or rate"
```

```
##  [8] "# Age:                   16 years and over"
##  [9] "Year,Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec"
## [10] "1948,4.0,4.7,4.5,4.0,3.4,3.9,3.9,3.6,3.4,2.9,3.3,3.6"
```

```r
U1 <- read.table(file="unadjusted_unemployment.csv",sep=",",header=TRUE)
head(U1)
```

```
##   Year Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1 1948 4.0 4.7 4.5 4.0 3.4 3.9 3.9 3.6 3.4 2.9 3.3 3.6
## 2 1949 5.0 5.8 5.6 5.4 5.7 6.4 7.0 6.3 5.9 6.1 5.7 6.0
## 3 1950 7.6 7.9 7.1 6.0 5.3 5.6 5.3 4.1 4.0 3.3 3.8 3.9
## 4 1951 4.4 4.2 3.8 3.2 2.9 3.4 3.3 2.9 3.0 2.8 3.2 2.9
## 5 1952 3.7 3.8 3.3 3.0 2.9 3.2 3.3 3.1 2.7 2.4 2.5 2.5
## 6 1953 3.4 3.2 2.9 2.8 2.5 2.7 2.7 2.4 2.6 2.5 3.2 4.2
```

- The data are in a table, and we want a time series. Here's one way to do that.

```r
u1 <- t(as.matrix(U1[2:13]))
dim(u1) <- NULL
date <- seq(from=1948,length=length(u1),by=1/12)
plot(date,u1,type="l",ylab="Percent unemployment (unadjusted)")
```
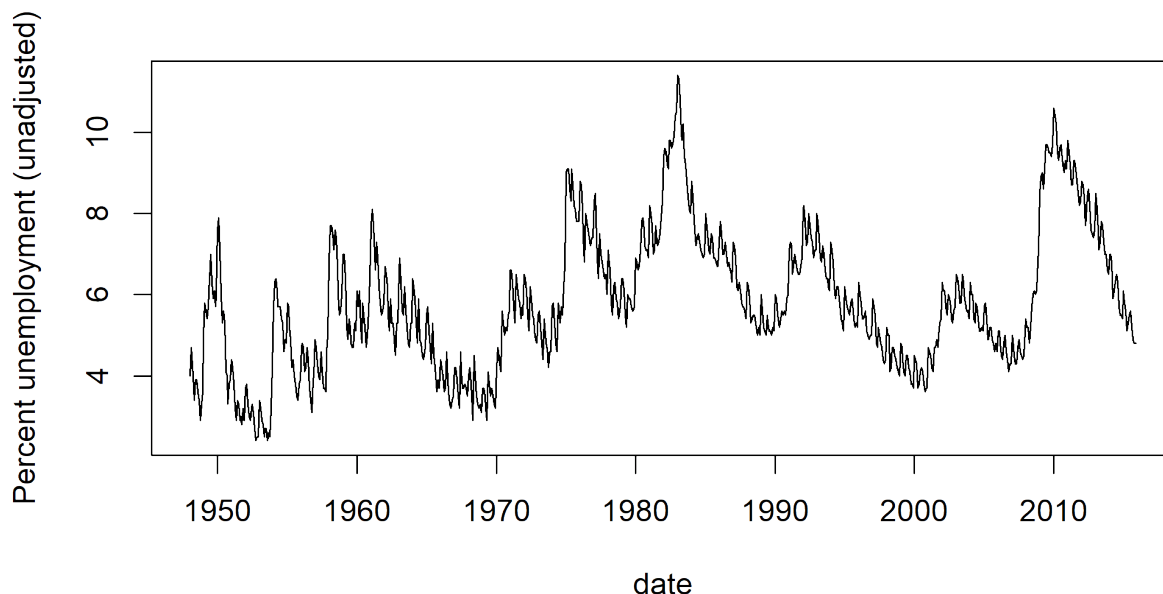

- We see seasonal variation, and perhaps we see business cycles on top of a slower trend.

- The seasonal variation looks like an additive effect, say an annual fluctuation with amplitude around 1 percentage point. For many purposes, we may prefer to look at a measure of monthly seasonally adjusted unemployment, which the Bureau of Labor Statistics also provides.

```r
U2 <- read.table(file="adjusted_unemployment.csv",sep=",",header=TRUE)
u2 <- t(as.matrix(U2[2:13]))
dim(u2) <- NULL
plot(date,u1,type="l",ylab="percent",col="black")
```
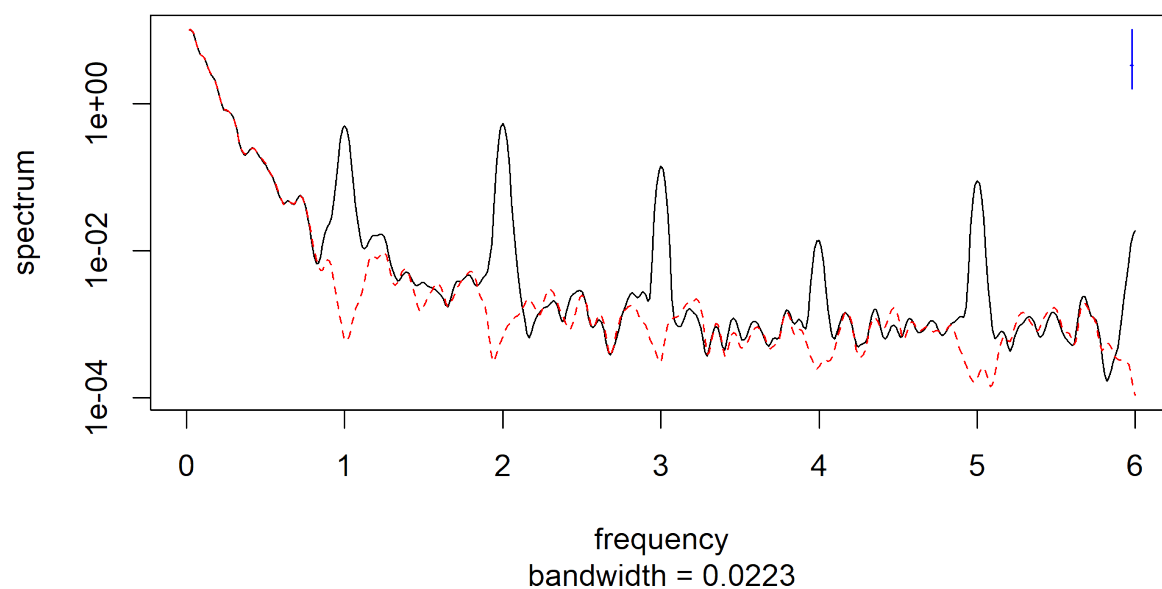
2

```
lines(date,u2,type="l",col="red")
title("Unemployment. Raw (black) and seasonally adjusted (red)")
```

Unemployment. Raw (black) and seasonally adjusted (red)
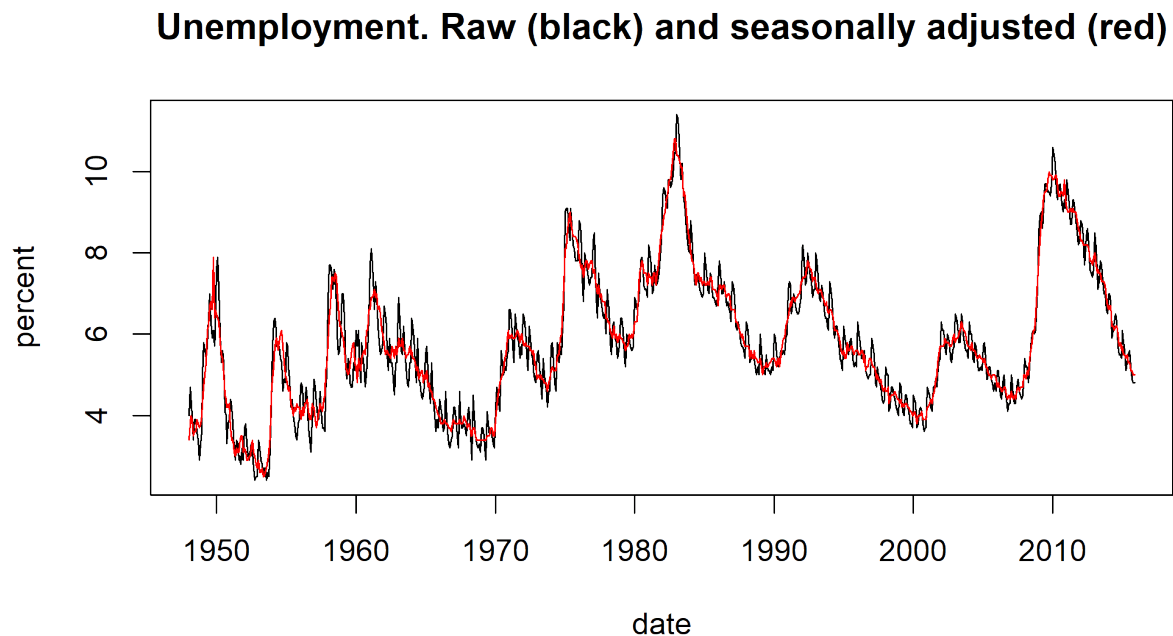
- As statisticians, we may be curious about how the Bureau of Labor Statistics adjusts the data, and whether this might introduce any artifacts that a careful statistician should be aware of.

- Let's look at what the adjustment does to the smoothed periodogram.

- To help R figure out units for plotting the spectrum, we're going to put our time series in the `ts` class.

```
u1_ts <- ts(u1,start=1948,frequency=12)
u2_ts <- ts(u2,start=1948,frequency=12)
spectrum(ts.union(u1_ts,u2_ts),spans=c(3,5,3),main="Unemployment. Raw (black) and seasonally adjusted (:
```

Unemployment. Raw (black) and seasonally adjusted (red)

frequency
bandwidth = 0.0223

> Why peaks at 2, 3, 4, 5, 6 cycles per year? A peak at only 1 cycle per year is sinusoidal seasonality. The 2-cycle might be the idea that there's less unemployment during summer and winter (Christmas time).


8.1.1 Question: What are the x-axis units?

> Cycles-per-year (rather than cycles-per-month). This more natural unit of frequency came about from the *ts* function in *R*, and setting the frequency to 12.

8.1.2 Question: Comment on what you learn from comparing these smoothed periodograms.

- Note: the `ts` class can also be useful for helping R choose other plotting options in a way appriate for time series. For example,

```
plot(u1_ts)
```
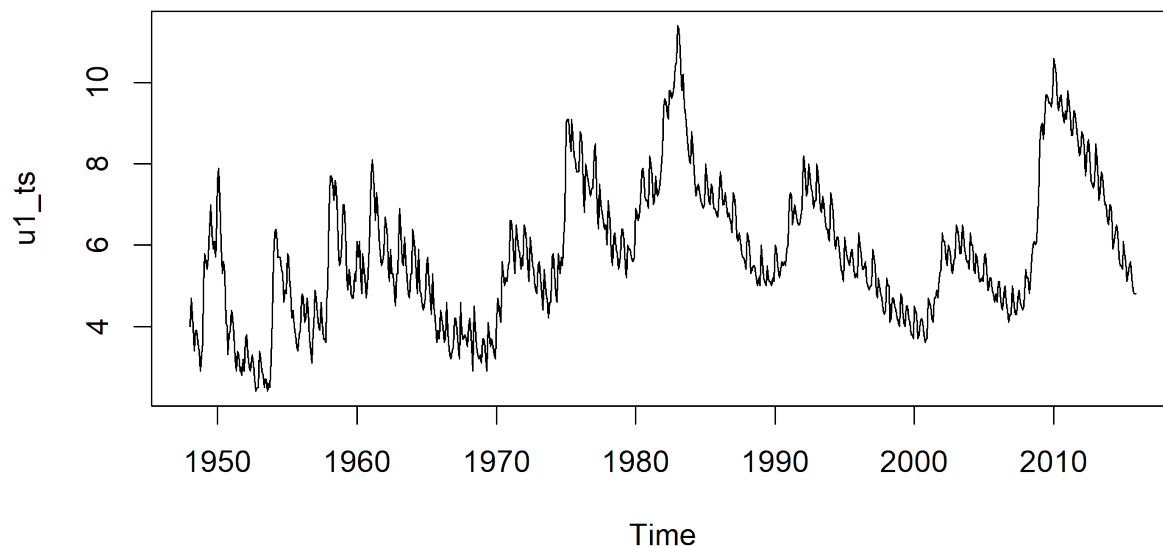
- Note: For a report, we should add units to plots. Also, extra details (like `bandwith` in the periodogram plot) should be explained or removed.

8.2 The transfer function (or frequency response function) of a smoother

- The <mark>ratio of the periodograms of the smoothed and unsmoothed time series is called the **transfer function** or **frequency response function** of the smoother.</mark>
- We can infer the frequency response of the smoother used by Bureau of Labor Statistics to deseasonalize the unemployment data.

```
s <- spectrum(ts.union(u1_ts,u2_ts),plot=FALSE)
```

- We need to figure out how to extract the bits we need from `s`

```
names(s)
```
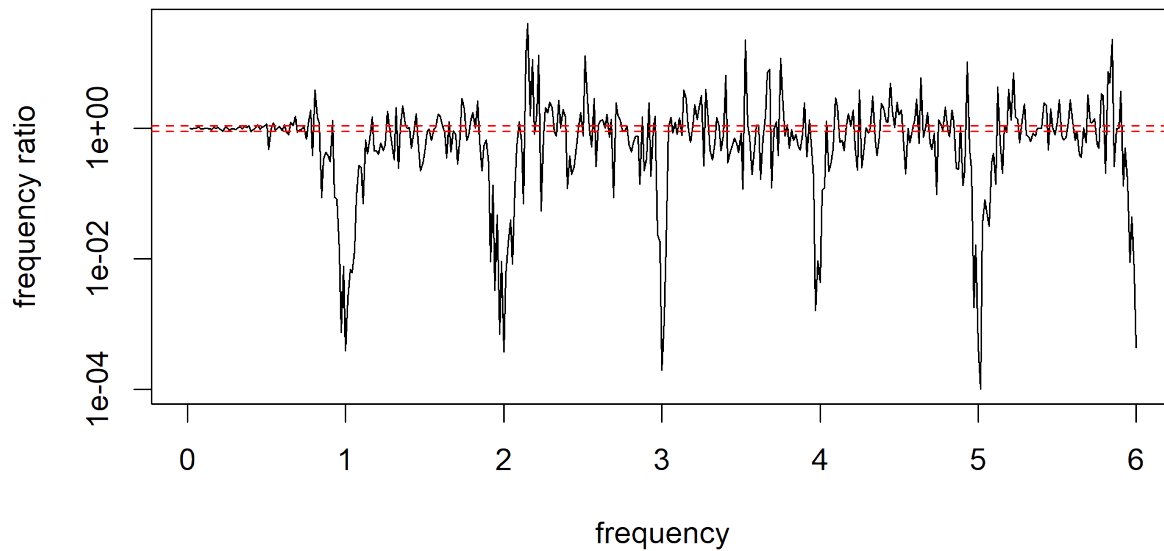
```
##  [1] "freq"      "spec"      "coh"      "phase"    "kernel"
##  [6] "df"        "bandwidth" "n.used"   "orig.n"   "series"
## [11] "snames"    "method"    "taper"    "pad"      "detrend"
## [16] "demean"
```

```
dim(s$spec)
```

```
## [1] 432   2
```

```
plot(s$freq,s$spec[,2]/s$spec[,1],type="l",log="y",
  ylab="frequency ratio", xlab="frequency",
  main="frequency response (dashed lines at 0.9 and 1.1)")
abline(h=c(0.9,1.1),lty="dashed",col="red")
```

frequency response (dashed lines at 0.9 and 1.1)

8.2.1 Question: What do you learn from this frequency response plot?

8.2.2 Loess smoothing

- Loess is a **Local linear regression** approach (perhaps an acronym for LOcal Estimation by Smoothing?)

- The basic idea is quite simple: at each point in time, we carry out a linear regression (e.g., fit a constant, linear or quadratic polynomial) using only points close in time. Thus, we can imagine a moving window of points included in the regression.

- `loess` is an R implementation, with the fraction of points included in the moving window being scaled by the `span` argument.

- Let's choose a value of the span that visually separates long term trend from business cycle.

```
u1_loess <- loess(u1~date,span=0.5)
plot(date,u1,type="l",col="red")
lines(u1_loess$x,u1_loess$fitted,type="l")
```
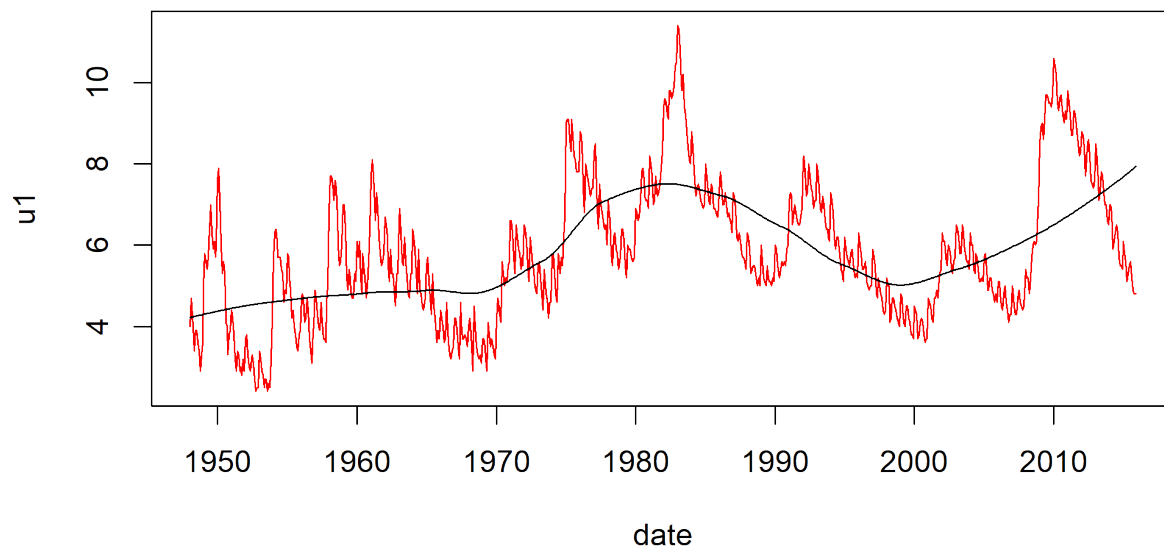
- Now, we can compute the frequency response function for what we have done.

```
s2 <- spectrum(ts.union(
  u1_ts,ts(u1_loess$fitted,start=1948,frequency=12)),
  plot=FALSE)
plot(s2$freq,s2$spec[,2]/s$spec[,1],type="l",log="y",
  ylab="frequency ratio", xlab="frequency", xlim=c(0,1.5),
  main="frequency response (dashed line at 1.0)")
abline(h=1,lty="dashed",col="red")
```

## frequency response (dashed line at 1.0)
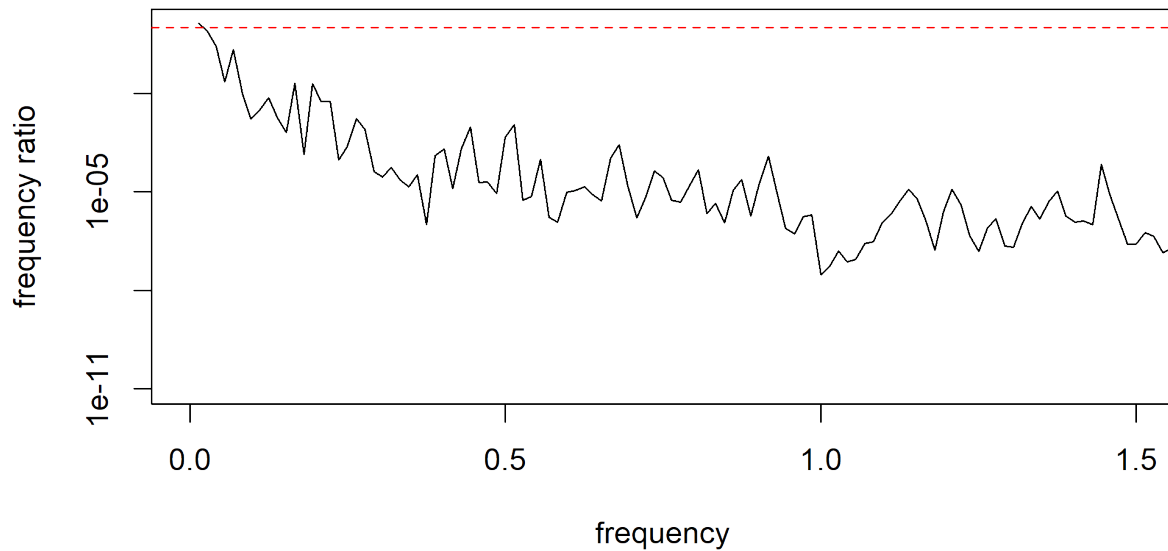
8.2.3 Question: Describe the frequency domain behavior of this filter.

## 8.3 Extracting business cycles: A band pass filter

- For the unemployment data, high frequency variation might be considered "noise" and low frequency variation might be considered trend.

- A band of mid-range frequencies might be considered to correspond to the business cycle.

- Let's build a smoothing operation in the time domain to extract business cycles, and then look at its frequency response function.

```
u_low <- ts(loess(u1~date,span=0.5)$fitted,start=1948,frequency=12)
u_hi <- ts(u1 - loess(u1~date,span=0.1)$fitted,start=1948,frequency=12)
u_cycles <- u1 - u_hi - u_low
plot(ts.union(u1, u_low,u_hi,u_cycles),
  main="Decomposition of unemployment as trend + noise + cycles")
```

## Decomposition of unemployment as trend + noise + cycles
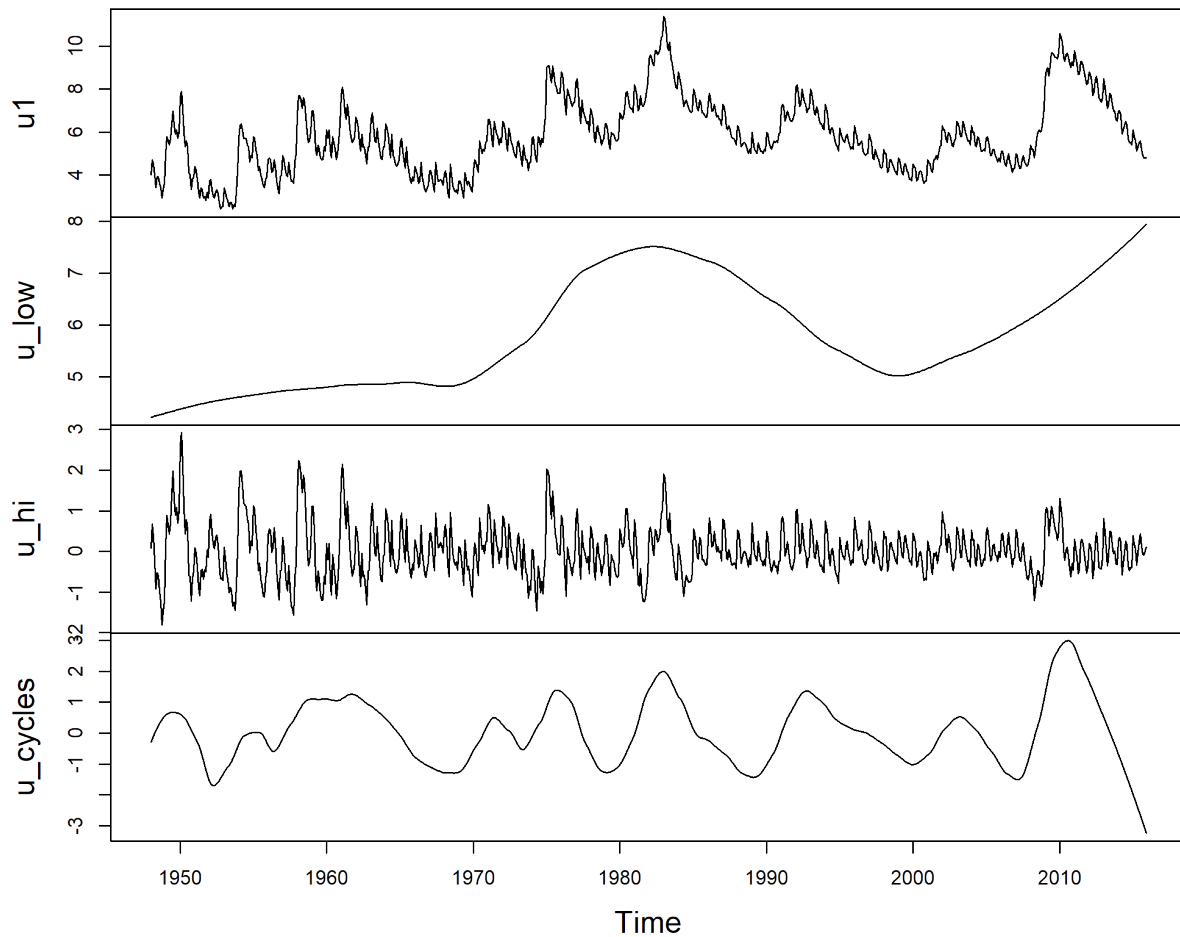

```r
spec_cycle <- spectrum(ts.union(u1_ts,u_cycles),
  spans=c(3,3),
  plot=FALSE)
freq_response_cycle <- spec_cycle$spec[,2]/spec_cycle$spec[,1]
plot(spec_cycle$freq,freq_response_cycle,
  type="l",log="y",
  ylab="frequency ratio", xlab="frequency", xlim=c(0,1.2), ylim=c(5e-6,1.1),
  main="frequency response (dashed line at 1.0)")
abline(h=1,lty="dashed",col="red")
```

frequency response (dashed line at 1.0)
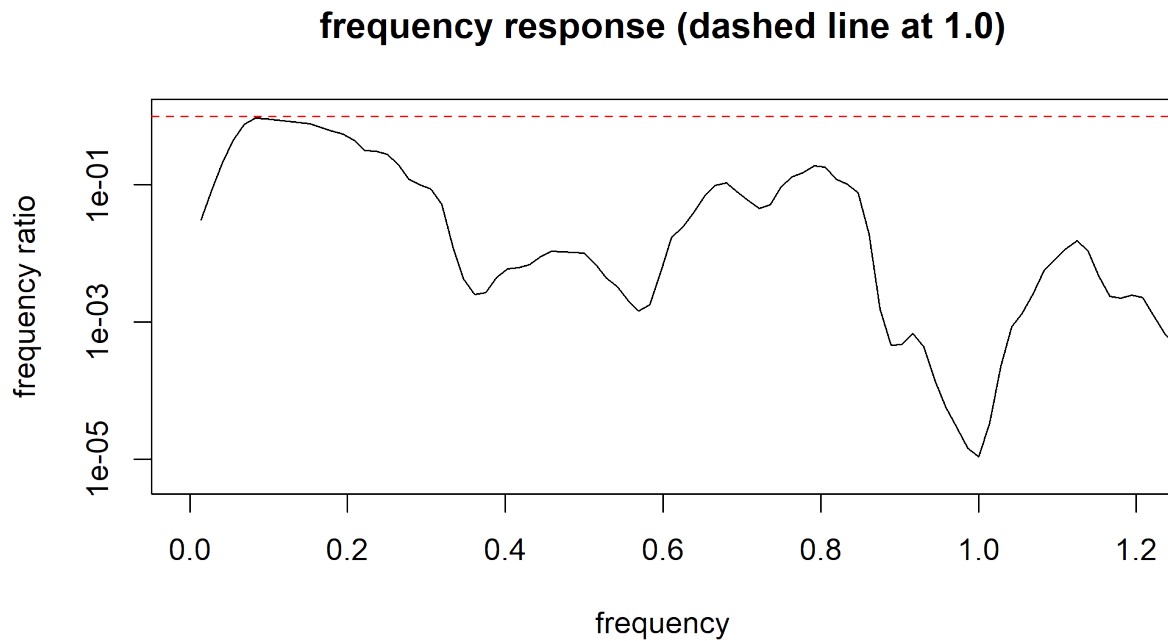
8.3.1 Question: Describe the frequencies (and corresponding periods) that this decomposition identifies as business cycles

- Note: Usually, we should specify units for frequency and period. Here, the units are omitted to give you an exercise!

- To help answer this question, let's add some lines to the previous plot

```
cut_fraction <- 0.5
plot(spec_cycle$freq,freq_response_cycle,
  type="l",log="y",
  ylab="frequency ratio", xlab="frequency", xlim=c(0,0.9), ylim=c(1e-4,1.1),
  main=paste("frequency response, showing region for ratio >", cut_fraction))
abline(h=1,lty="dashed",col="blue")
freq_cycles <- range(spec_cycle$freq[freq_response_cycle>cut_fraction])
abline(v=freq_cycles,lty="dashed",col="blue")
abline(h=cut_fraction,lty="dashed",col="blue")
```

frequency response, showing region for ratio > 0.5
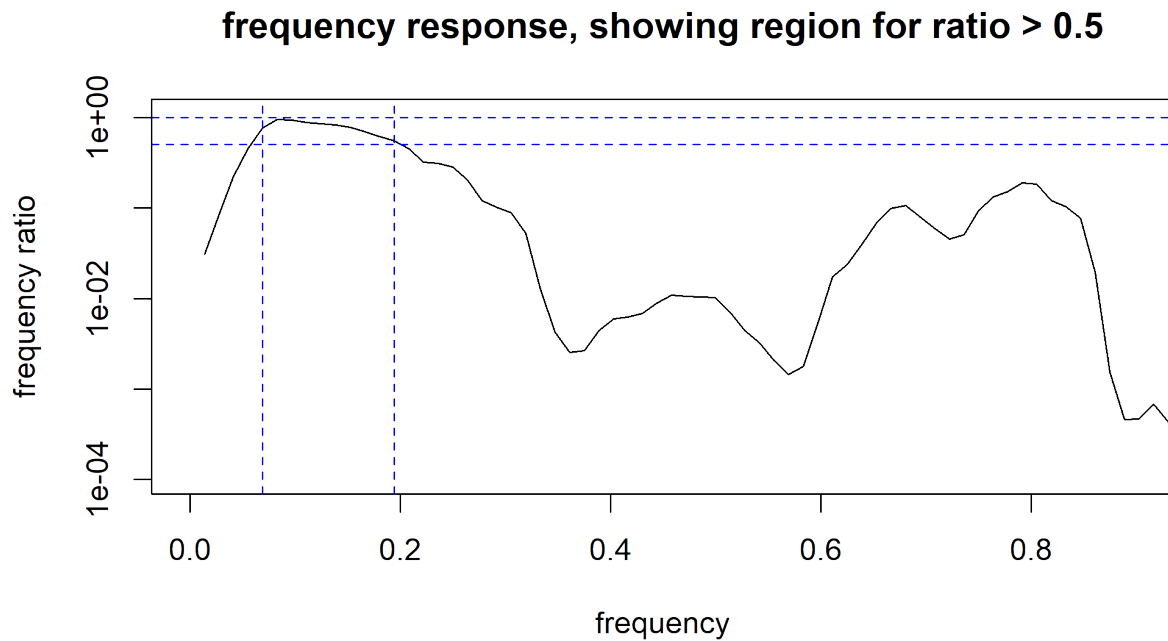

```
kable(matrix(freq_cycles,nrow=1,dimnames=list("frequency",c("low","hi"))),digits=3)
```

	low	hi
frequency	0.069	0.194

8.3.2 Question: So far as we have opinions on business cycles, use them to criticize this decomposition.

### 8.3.3 Question: Criticizing the construction of the blue dashed lines

- Why do the blue dashed lines in the above figure not meet exactly on the frequency response curve?
- What could or should be done to improve this?

8.3.4 Looking for business cycles

- We can plot just the lower frequencies of a smoothed periodogram for the raw unemployment data, to zoom in on the frequencies around the business cycle frequency.

- Standard periodogram smoothers use the same smoothing bandwidth across all frequencies. This may not always be appropriate. Why?

- Sometimes in practice we want to use less smoothing when we are focusing on low frequency behaviors.

```
s1 <- spectrum(u1_ts,spans=c(3),plot=FALSE)
plot(s1,xlim=c(0,0.7),ylim=c(1e-2,max(s1$spec)))
```

Series: x
Smoothed Periodogram
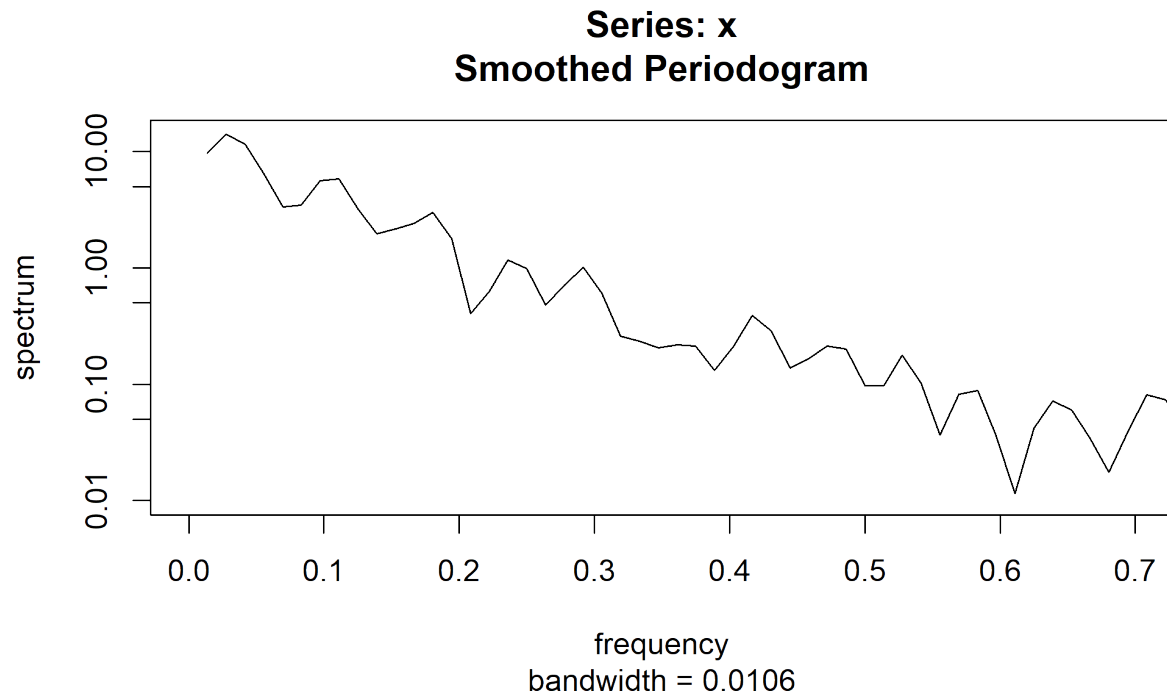
8.3.5   Question: Comment on the evidence for and against the concept of a business cycle in the above figure.

8.4   Common smoothers in R

- Above, we have used the **local regression smoother loess** but there are other options.

- Our immediate goal is to get practical experience using a smoother and then statistically assessing what we have done.

- You can learn about alternative smoothers, and try them out, if you like.

- `ksmooth` is a **kernel smoother**. The default periodogram smoother in `spectrum` is also a kernel smoother.

- `smooth.spline` is a **spline smoother**.

- All these smoothers have some concept of a **bandwidth**, which is a measure of the size of the neighborhood of time points in which data affect the smoothed value at a particular time point.

- The concept of bandwidth is most obvious for kernel smoothers, but exists for other smoothers.

- We usually only interpret bandwidth up to a constant. For a particular smoothing algorithm and software implementation, you learn by experience to interpret the comparative value (smaller bandwidth means less smoothing).

- Typically, when writing reports, it makes sense not to present or discuss smoothing bandwidth since it is not directly interpretable for most readers.