Modeling and Analysis of Time Series Data

Chapter 8: Smoothing in the time and frequency domains

Edward L. Ionides

Outline

Introduction to smoothing in time series analysis

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

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

- Smoothers have convenient interpretation in the frequency domain. A smoother typically shrinks high frequency components and preserves low frequency components.

  时域数据并平滑后,高频信号的贡献就减少了

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 US unemployment figures are published by the Bureau of Labor Statistics (BLS).
- Measuring unemployment has subtleties, but these are not our immediate focus.

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

```
 [1] "# Data extracted on: February 3, 2022 (12:29:30 PM)"
 [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] "# Series Id:           LNU04000000"
 [7] "# Labor force status:  Unemployment rate"
 [8] "# Type of data:        Percent or rate"
 [9] "# Age:                 16 years and over"
[10] "Year,Jan,Feb,Mar,Apr,May,Jun,Jul,Aug,Sep,Oct,Nov,Dec"
```

```
U1 <- read.table(file="unadjusted_unemployment.csv",
  sep=",",header=TRUE)
head(U1,3)
```

```
 Year Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
 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
 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
 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
```

Question 8.1. A coding exercise: Explain how the tabulated data in U1 are converted to a time series, below.

R stores matrices column-wisely as a long vector, with a dim attribute to turn it into a matrix
R的保存方法和numpy一样,直接修改dim(shape)就可以展开

```
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="Unemployment rate (unadjusted)")
```

coivd untypical
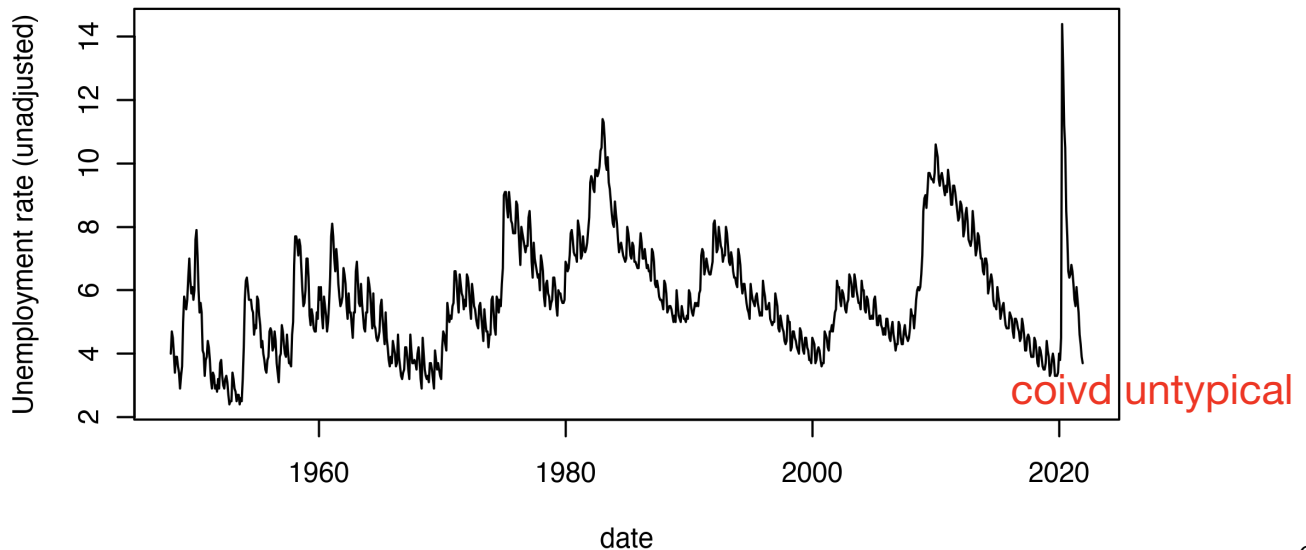
- We see seasonal variation and economic cycles on top of a trend.
- The seasonal variation looks like an additive effect, say an annual fluctation with amplitude around 1 percentage point.
- Sometimes, we may prefer to look at monthly seasonally adjusted unemployment, also provided by BLS.

```
U2 <- read.table(file="adjusted_unemployment.csv",sep=",",
  header=TRUE)
u2 <- t(as.matrix(U2[2:13]))
```

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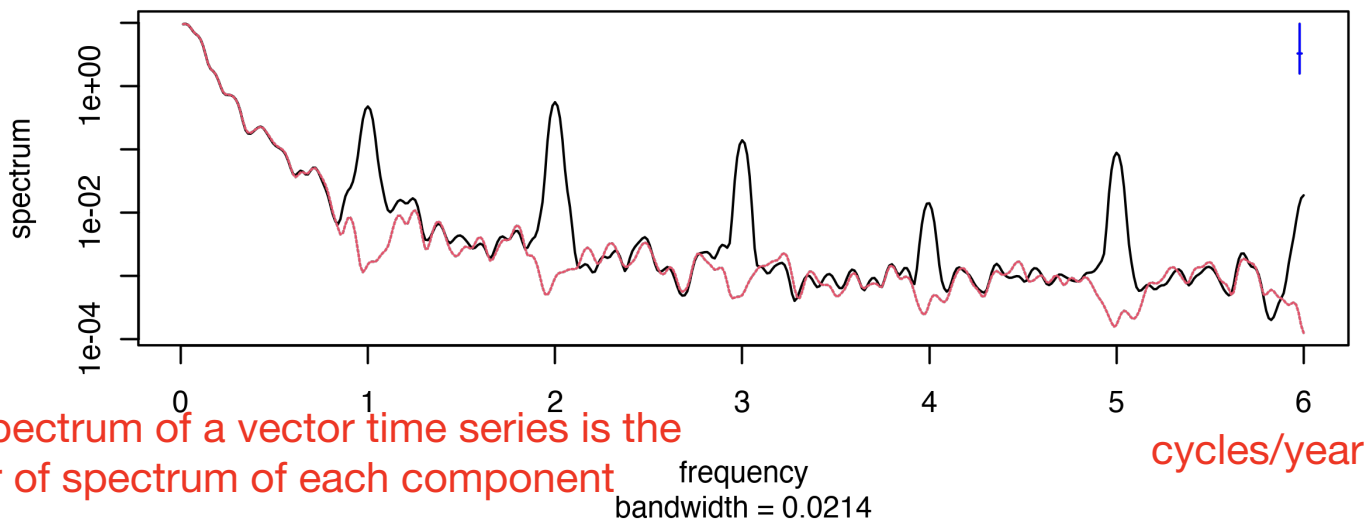

Adjusted的数据去除了一些毛刺,但可能不能被认为是trend
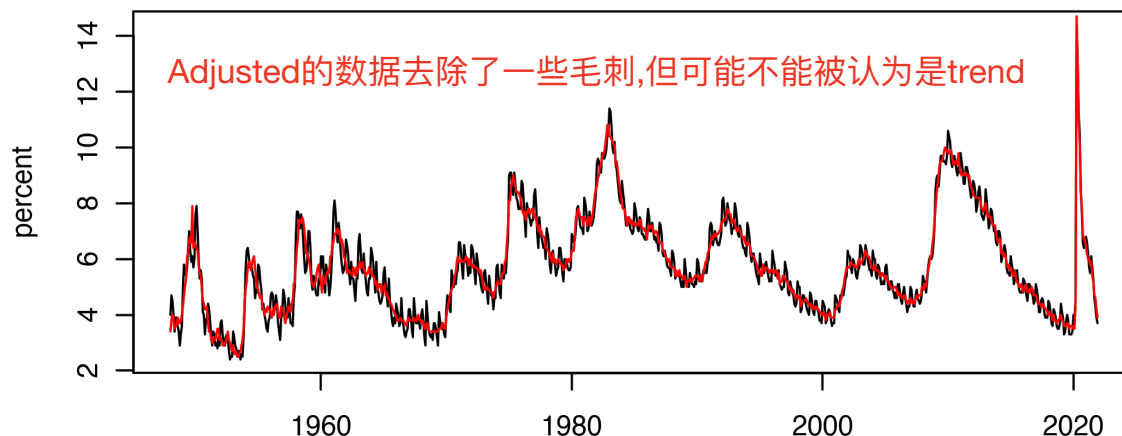
- We can wonder how the BLS adjusts the data, and if this introduces any artifacts that a careful statistician should be aware of.

- To help understand the seasonal adjustment, we look at what it does to the smoothed periodogram.
- Using the `ts` class we can tell R the units of time.

```
u1_ts <- ts(u1,start=1948,frequency=12)  sample the time series data
u2_ts <- ts(u2,start=1948,frequency=12)
spectrum(ts.union(u1_ts,u2_ts),spans=c(3,5,3), 多次使用滑动窗口
   main="Unemployment: raw (black), seasonally adjusted (red)")
```

ts union combine the time series
into a vector-valued time series

Unemployment: raw (black), seasonally adjusted (red)

The spectrum of a vector time series is the
vector of spectrum of each component

frequency
bandwidth = 0.0214

cycles/year

Comments on the smoothed periodogram

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

Question 8.2. Why does the unadjusted spectrum have peaks at 2,3,4,5,6 cycles per year as well as 1 cycle per year? harmonics of the 1-year cycle

A purely sinusoidal cycle has no harmonics, just a frequency component at the base frequency. Non-sinusoidal periodic functions need higher frequencies to be represented

Question 8.3. Comment on what you learn from comparing these smoothed periodograms.
From the raw data, we learn that there is considerable power added at two cycles per year as well as one cycle per year.
The de-seaonlized data have successfully removed the seasonal peaks and have a very smooth periodogram, with a small dip at seasonal frequencies but some power remaining there

同时有高频和低频长周期信号,且频率是相互的整数倍

The frequency response function

- The ratio of the periodograms of the smoothed and unsmoothed time series is the **frequency response** of the smoother.

- The frequency response function tells us how much the smoother contracts (or inflates) the sine and cosine components at each frequency ω.

- A frequency response may involve change in phase as well as magnitude, but here we consider only magnitude.

- Linear, time invariant transformations do not move power between frequencies, so they are characterized by their frequency response function.

- Smoothers are linear and time invariant, at least approximately. If we scale or shift the data, we expect the smoothed estimate to have the same scale or shift. We expect a smooth approximation to the sum of two time series to be approximately the sum of the two smoothed series.

Calculating a frequency response function

- We investigate 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)
```
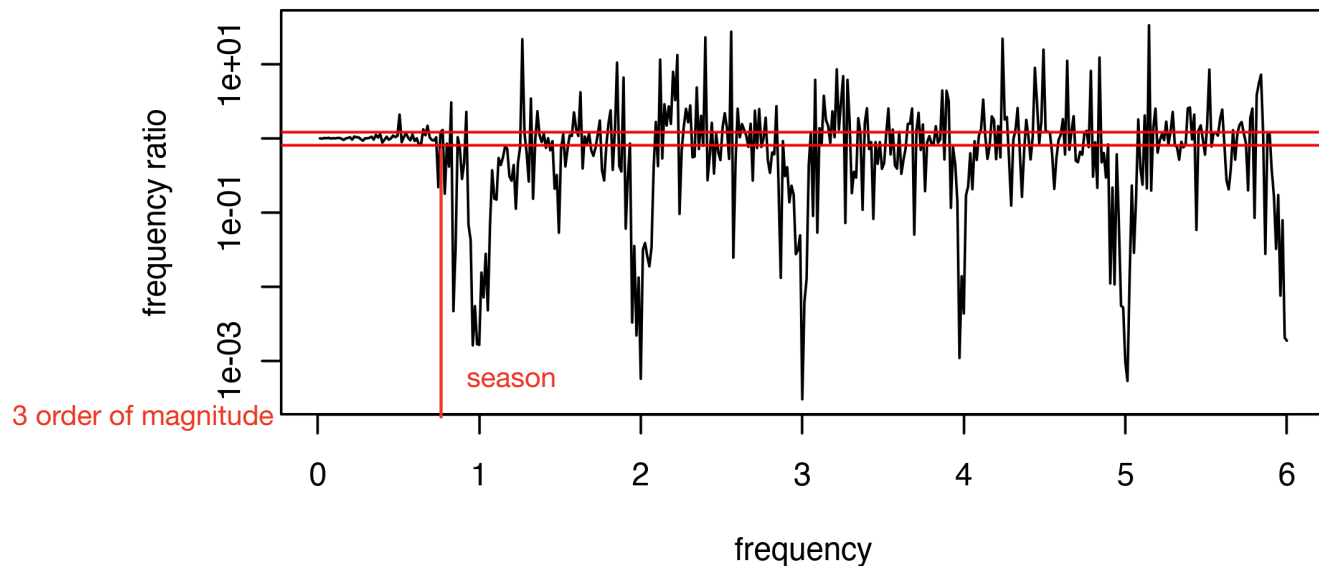
- We find the parts of s that we need to plot the frequency response.

```
names(s)
```

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

```
plot(s$freq,s$spec[,2]/s$spec[,1],type="l",log="y",
  ylab="frequency ratio", xlab="frequency",
  main="frequency response (red lines at 0.8 and 1.2)")
abline(h=c(0.8,1.2),col="red")
```

frequency response (red lines at 0.8 and 1.2)

**Question 8.4**. What do you learn from this frequency response plot?

频率小于0.75的信号是trend
Seasonal Frequencies (1 cycle per year and its harmonics) are shrunk 3
order of magnitude
High frequencies have some extra noise

Estimating trend by Loess smoothing

- Loess is a **Local linear regression** approach (perhaps an acronym for LOcal Estimation by Smoothing) also known as **Lowess** (perhaps LOcallyWEighted Sum of Squares). 在每个点做线性回归来平滑

- At each point in time, Loess computes a linear regression (a constant, linear or quadratic trend estimate) using only neighboring times.

- 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.

- We can choose a value of the span that visually separates long term trend from business cycle.
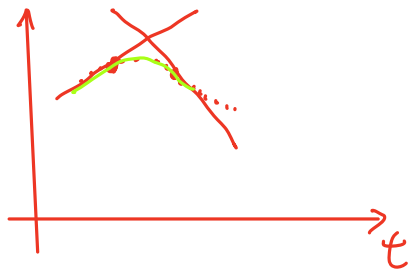
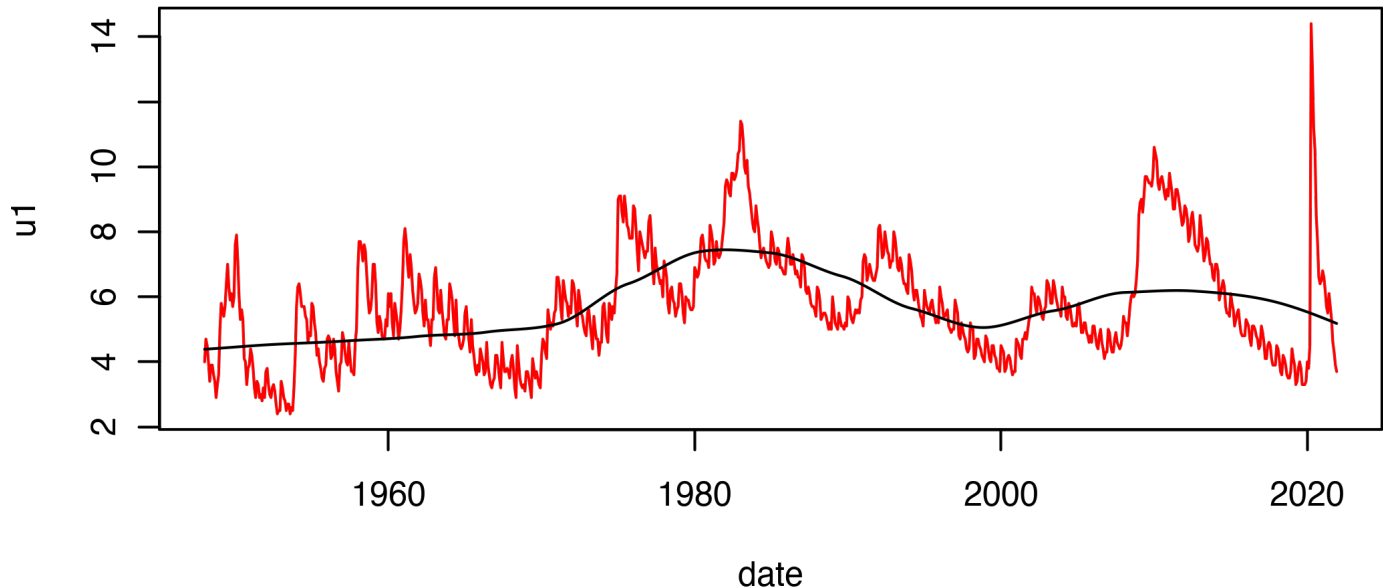window size is determined by the loess 函数的span参数
span is approximately the proportional of all points used in the window

A Loess smooth of unemployment

This is usually chosen by trial an error. It determine
what is estimated as trend or as random variation

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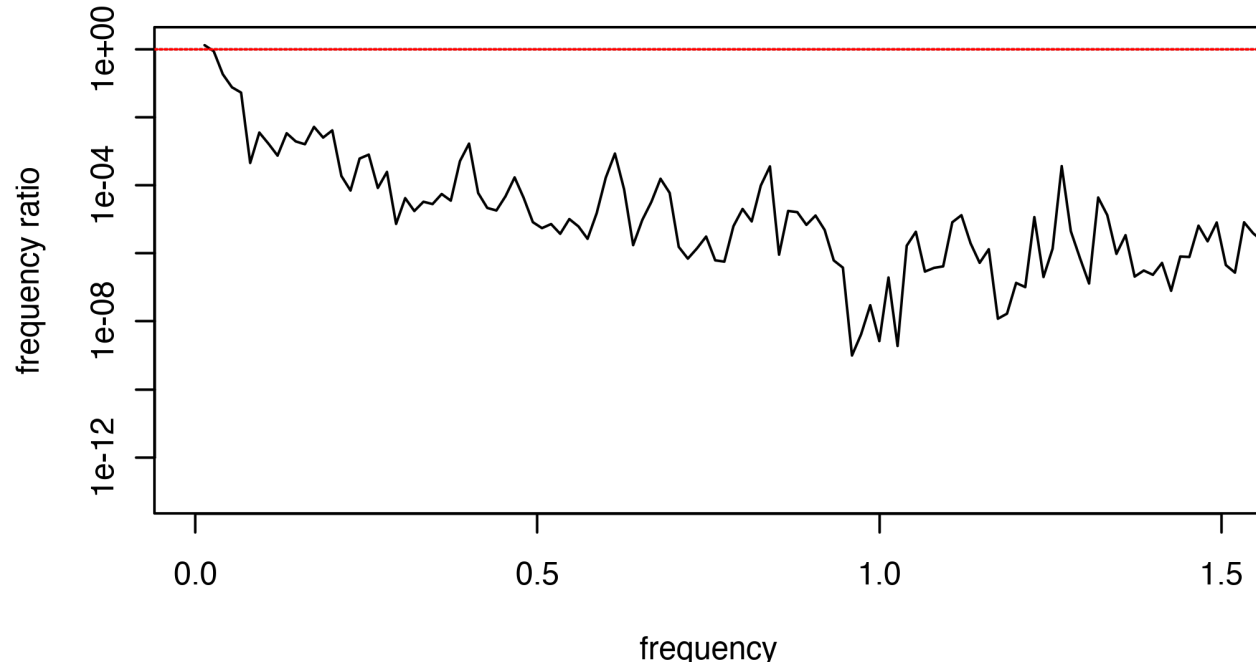
Now, we compute the frequency response function for what we have done.

```
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 (red line at 1.0)")
abline(h=1,lty="dashed",col="red")
```

低频可以获得trend

frequency response (red line at 1.0)

Question 8.5. Describe the frequency domain behavior of this filter.

Smoothers are called low-pass filter. They preserve low frequency and drop out high ones.

Frequencies higher than ~0.1 cycles per-year (10 yr period) are reduced by over 1 order of magnitude

# 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 <u>mid-range</u> frequencies might be considered to correspond to the business cycle.
- We 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
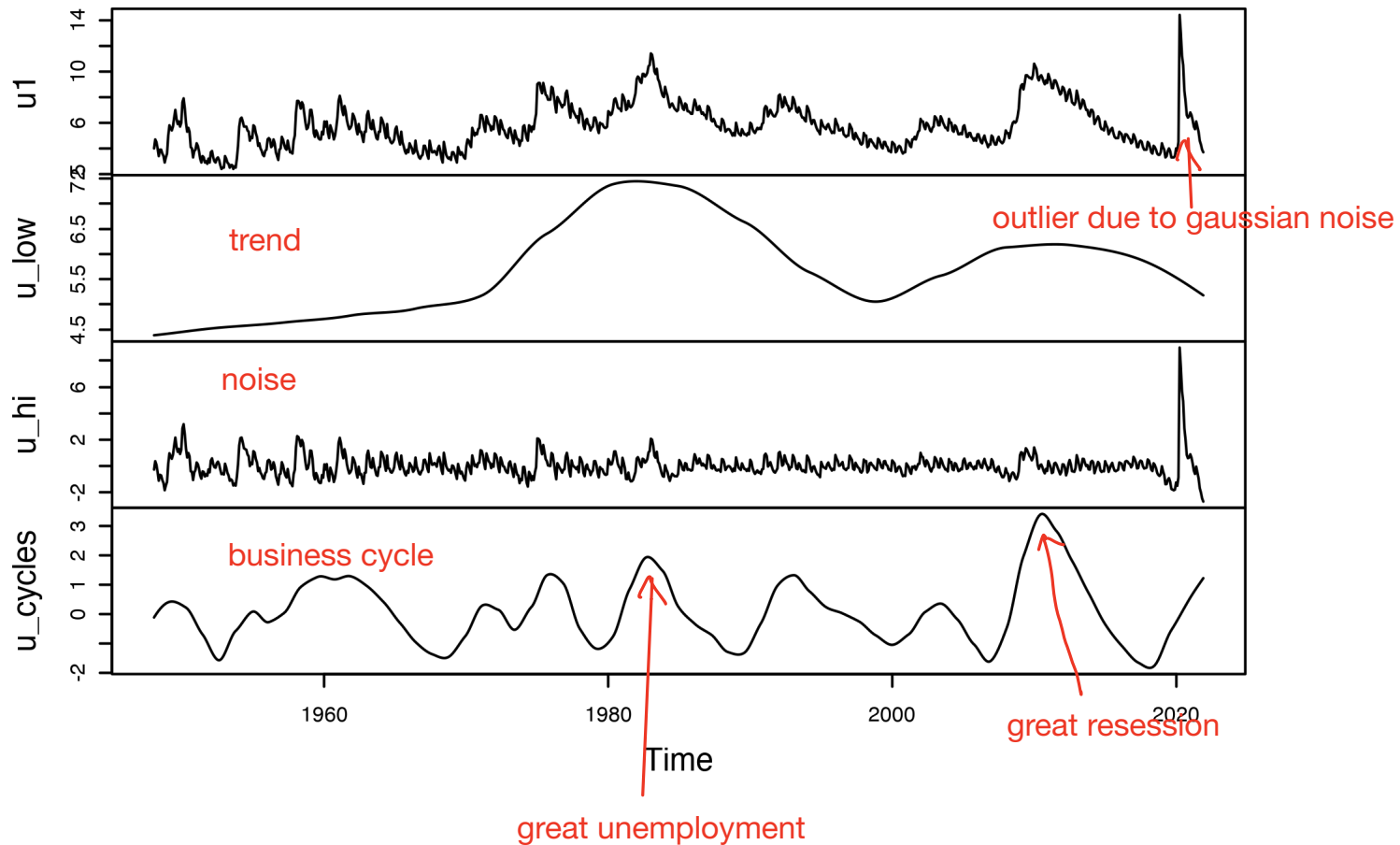
	low	hi
frequency range, region for ratio greater than 0.5	0.048	0.211

spectrum ratio

5-year period

frequency unit is cycle per year
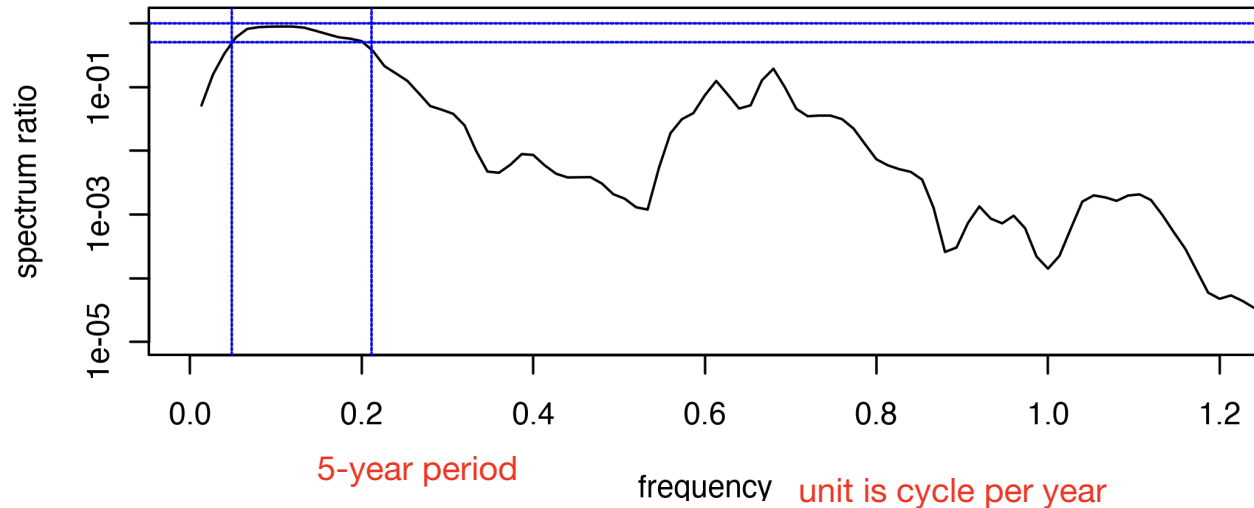
Question 8.6. Describe the frequencies (and corresponding periods) that this decomposition identifies as business cycles. Note: units of frequency are omitted to give you an exercise!

This shows we have estimated 5-20 year cycles as "business cycle".
This makes reasonable sense. Economists sometimes concern shorter cycles: to include there

Below is a smoothed periodogram for the raw unemployment data, plotted up to 0.7 cycles per year to focus on relevant frequencies.
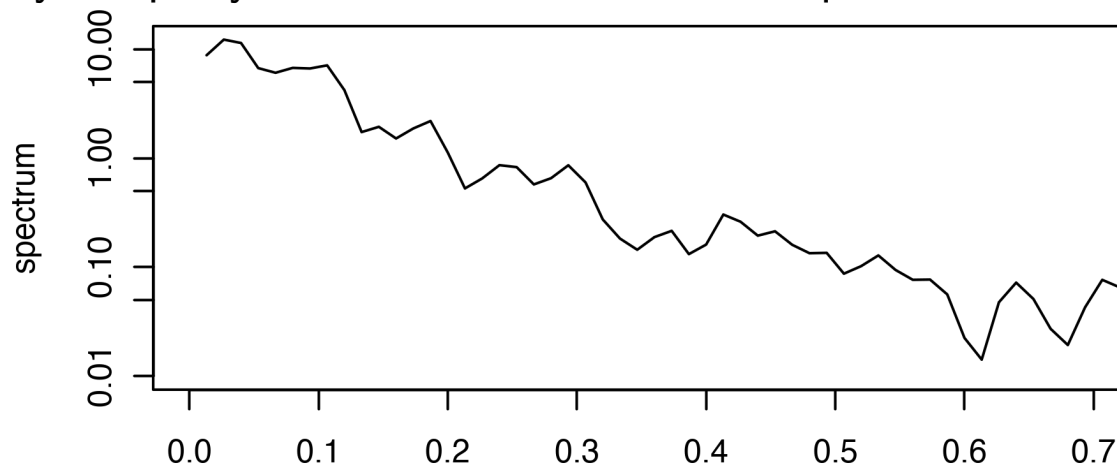
Note the error bar got cut off by truncating the x-axis range

frequency
bandwidth = 0.0102

Question 8.7. Comment on the evidence for and against the concept of a business cycle in the above figure.

There are not clear peaks in the periodogram. The economy doesn't appear to have resonant the frequencies where specific business cycle patterns appear repeatedly. The economy does cycle, but there doesn't appear to be a well-defined length of cycles or the government does a good job at dumping the cycles.

Common smoothers in R

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

- `ksmooth` is a **kernel smoother**. The default periodogram smoother in `spectrum` is also a kernel smoother. See
https://en.wikipedia.org/wiki/Kernel_smoother

- `smooth.spline` is a **spline smoother**.
https://en.wikipedia.org/wiki/Smoothing_spline

- You can learn about alternative smoothers, and try them out if you like, but `loess` is a good practical choice for many smoothing applications.

Bandwidth for a 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 to focus on the tuning parameter for the smoother in question, which is not the bandwidth unless you are going kernel smoothing.

Further reading

- Section 2.3 of Shumway and Stoffer (2017) discusses smoothing of time series, in the time domain.

- Section 4.2 of Shumway and Stoffer (2017) presents a frequency response function for linear filters, related to this chapter but in a different context.

References and Acknowledgements

Shumway RH, Stoffer DS (2017). *Time Series Analysis and its Applications: With R Examples*. Springer.

- Compiled on February 3, 2022 using R version 4.1.2.
- Licensed under the Creative Commons Attribution-NonCommercial license. Please share and remix non-commercially, mentioning its origin.
- We acknowledge previous versions of this course.