Intro to time series analysis

FISH 550 – Applied Time Series Analysis

Mark Scheuerell

28 March 2023

Topics for today

Characteristics of time series (ts)

- What is a ts?
- Classifying ts
- Trends
- Seasonality (periodicity)

Classical decomposition

What is a time series?

A set of observations taken sequentially in time

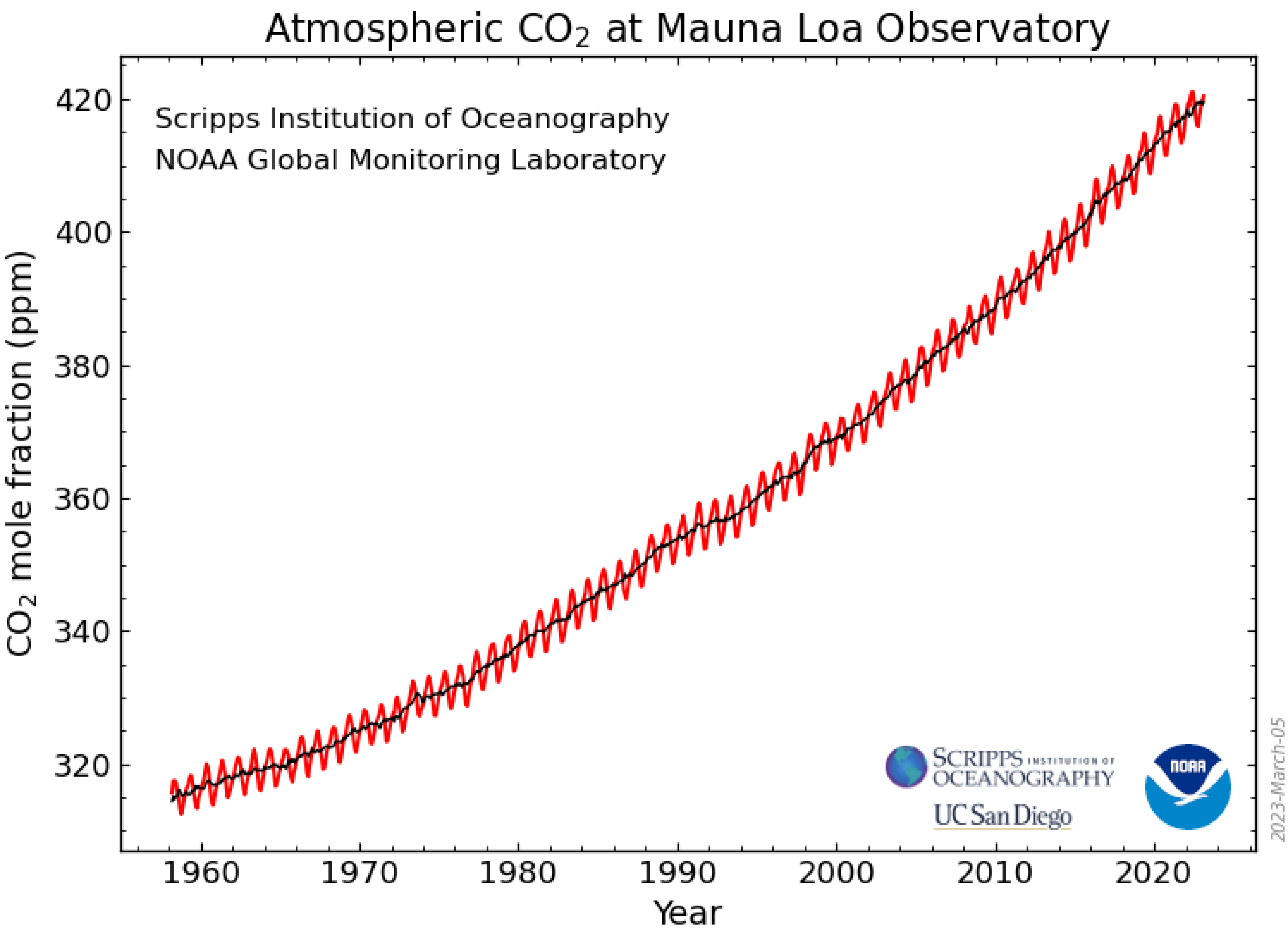
What is a time series?

A ts can be represented as a set

$$\{x_1, x_2, x_3, \ldots, x_n\}$$

For example,

$$\{10, 31, 27, 42, 53, 15\}$$


Atmospheric CO2 at Mauna Loa Observatory
Scripps Institution of Oceanography
NOAA Global Monitoring Laboratory
420
400
380
360
340
320
CO2 mole fraction (ppm)
1960
1970
1980
1990
2000
2010
2020
Year
2023-March-05

Classification of time series

By some *index set*

Interval across real time; $x(t)$

- begin/end: $t \in [1.1, 2.5]$

Classification of time series

By some *index set*

Discrete time; x_t

- Equally spaced: $t = \{1, 2, 3, 4, 5\}$
- Equally spaced w/ missing value: $t = \{1, 2, 4, 5, 6\}$
- Unequally spaced: $t = \{2, 3, 4, 6, 9\}$

Classification of time series

By the *underlying process*

Discrete (eg, total # of fish caught per trawl)

Continuous (eg, salinity, temperature)

Classification of time series

By the *number of values recorded*

Univariate/scalar (eg, total # of fish caught)

Multivariate/vector (eg, # of each spp of fish caught)

Classification of time series

By the *type of values recorded*

Integer (eg, # of fish in 5 min trawl = 2413)

Rational (eg, fraction of unclipped fish = 47/951)

Real (eg, fish mass = 10.2 g)

Complex (eg, $\cos(2\pi 2.43) + i \sin(2\pi 2.43)$)

Classification of time series

We will focus on integers & real-values in discrete time

Univariate (x_t)

Multivariate $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_t$

Time series objects in R

Time series objects have a special designation in R: `ts`

```
ts(data,
   start, end,
   frequency
   )
```

Time series objects in R

Time series objects have a special designation in R: `ts`

```
ts(data,
   start, end,
   frequency
   )
```

`data` should be a vector (univariate)

or a data frame or matrix (multivariate)

Time series objects in R

Time series objects have a special designation in R: `ts`

```
ts(data,
   start, end,
   frequency
   )
```

`start` and `end` give the first and last time indices

For monthly series, specify them as `c(year, month)`

Time series objects in R

Time series objects have a special designation in R: `ts`

```
ts(data,
   start, end,
   frequency
   )
```

`frequency` is the number of observations per unit time

For annual series, `frequency = 1`

For monthly series, `frequency = 12`

Time series objects in R

Time series objects have a special designation in R: `ts`

```
ts(data,
   start, end,
   deltat
   )
```

`deltat` is the fraction of the sampling period

For annual series, `deltat = 1`

For monthly series, `deltat = 1/12`

Time series objects in R

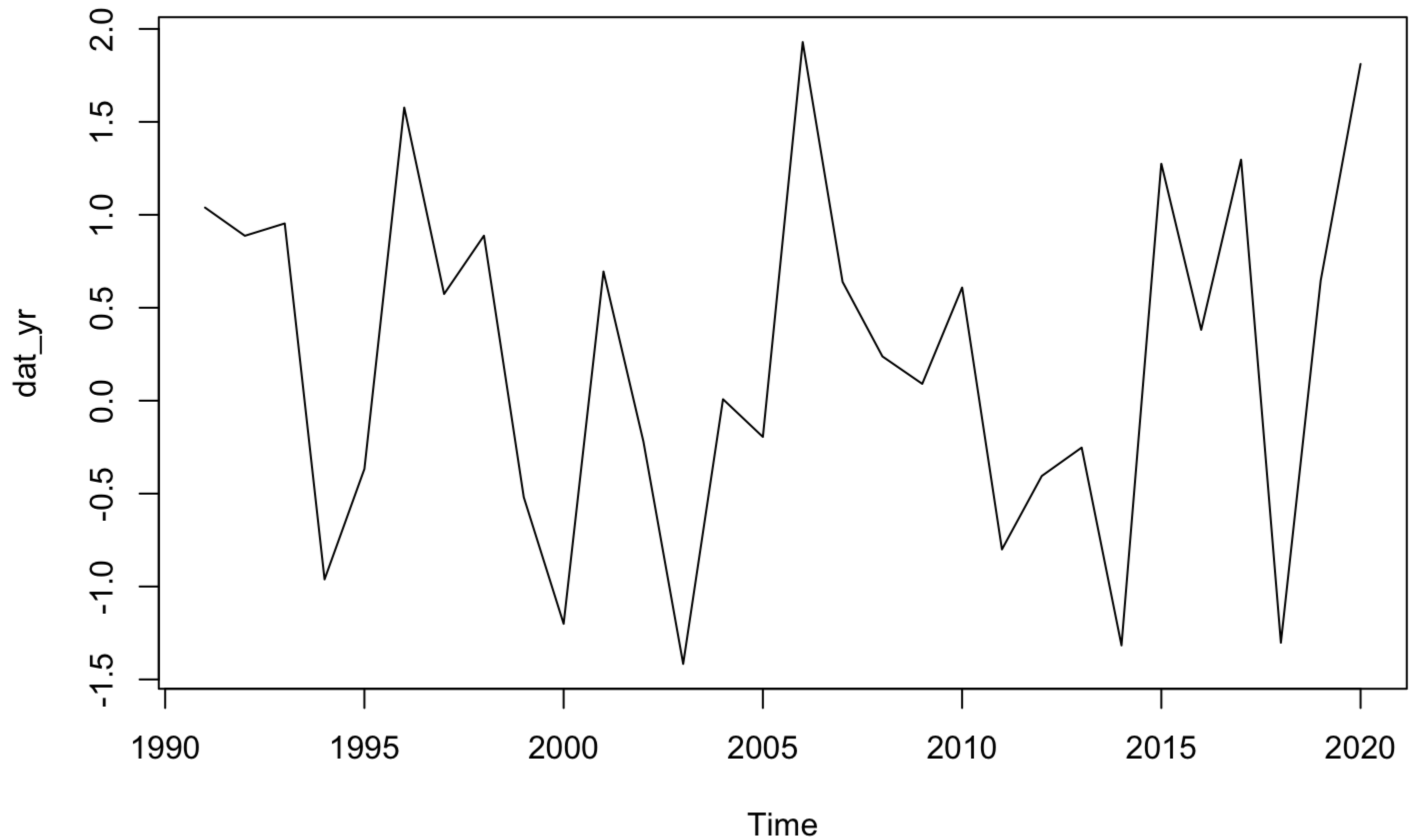
```
set.seed(507)

## annual data
dat_1 <- rnorm(30)
dat_yr <- ts(dat_1,
             start = 1991, end = 2020,
             frequency = 1)

## monthly data
dat_2 <- rnorm(30*12)
dat_mo <- ts(dat_2,
             start = c(1991, 1), end = c(2020, 12),
             frequency = 12)
```

Plotting time series objects in R

There is a designated function for plotting `ts` objects: `plot.ts()`

```
plot.ts(ts_object)
```

Plotting time series objects in R

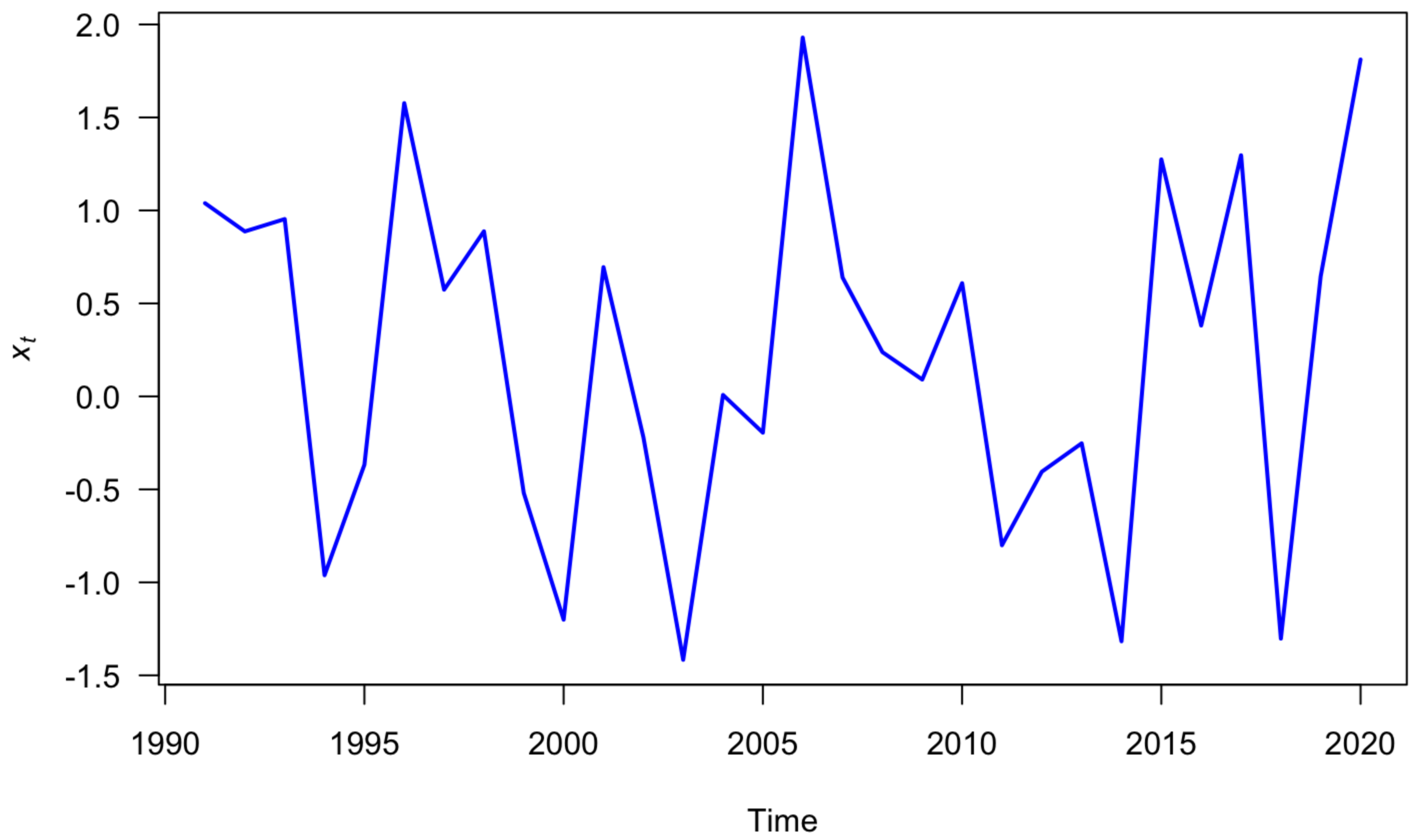


Plotting time series objects in R

We can specify some additional arguments to `plot.ts`

```
plot.ts(dat_yr,
        ylab = expression(italic(x[t])),
        las = 1, col = "blue", lwd = 2)
```

Plotting time series objects in R



Analysis of time series

Statistical analyses of time series

Most statistical analyses are concerned with estimating properties of a population from a sample

For example, we use fish caught in a seine to infer the mean size of fish in a lake

Statistical analyses of time series

Time series analysis, however, presents a different situation:

- Although we could vary the *length* of an observed time series, it is often impossible to make multiple observations at a *given* point in time

Statistical analyses of time series

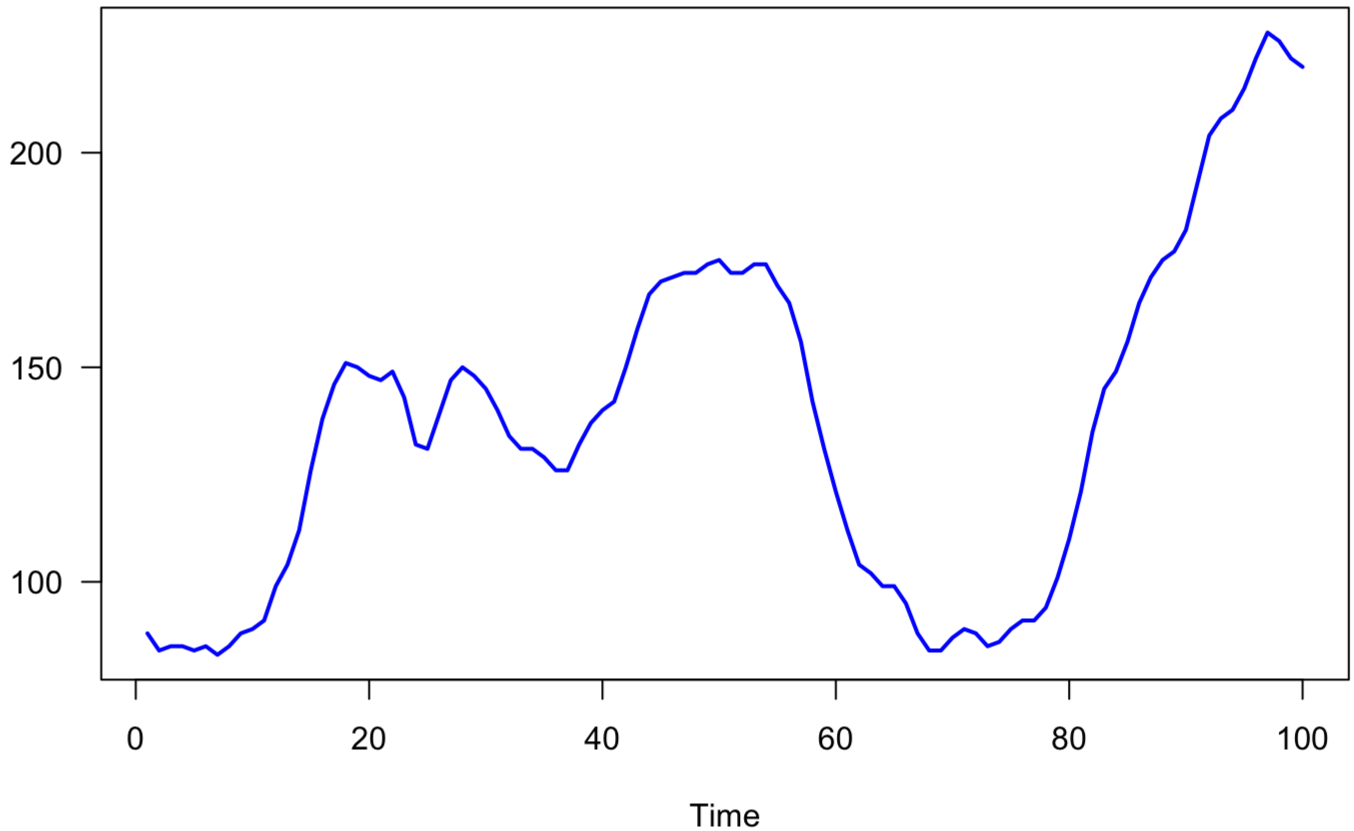
Time series analysis, however, presents a different situation:

- Although we could vary the *length* of an observed time series, it is often impossible to make multiple observations at a *given* point in time

For example, one can't observe today's closing price of Microsoft stock more than once

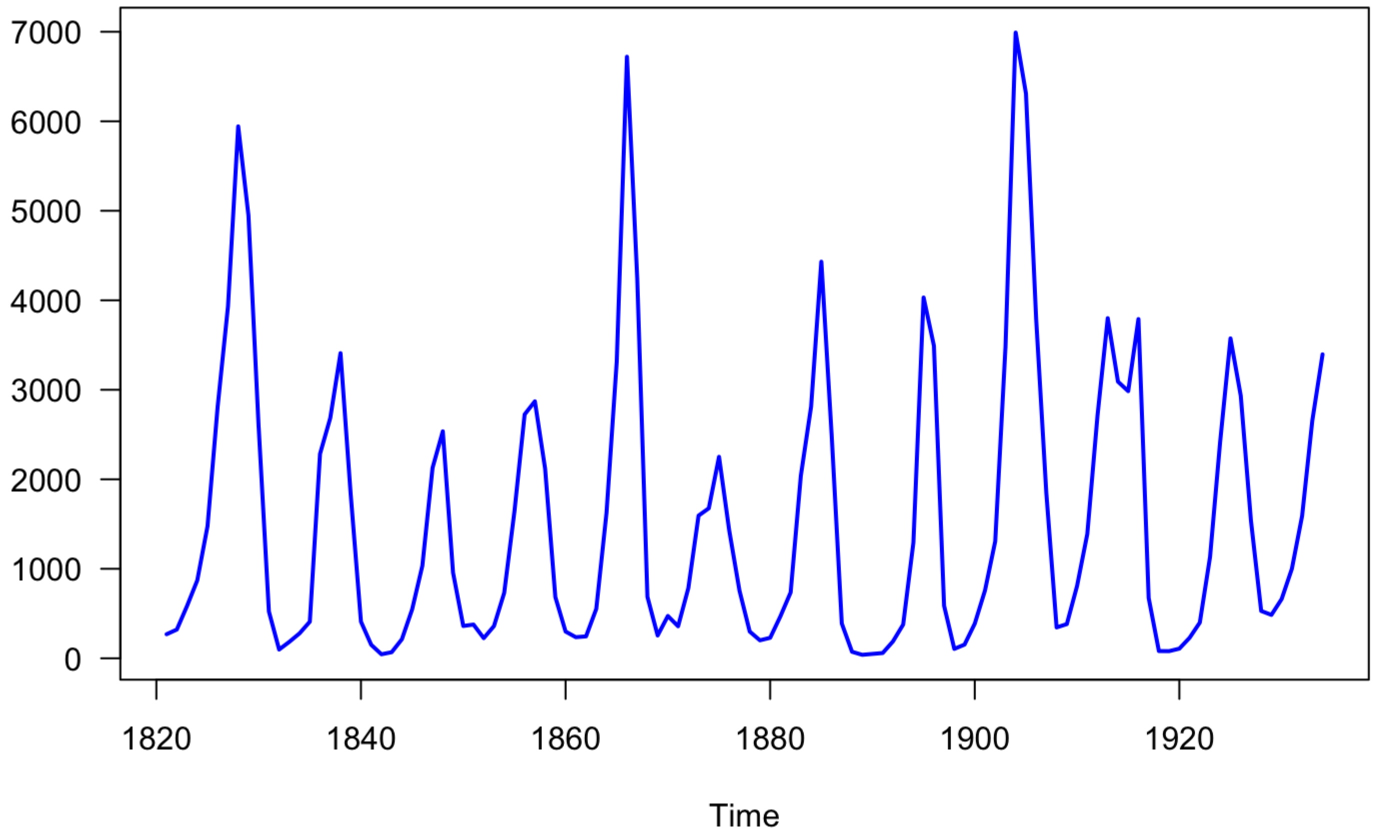
Thus, conventional statistical procedures, based on large sample estimates, are inappropriate

Descriptions of time series



Number of users connected to the internet

Descriptions of time series

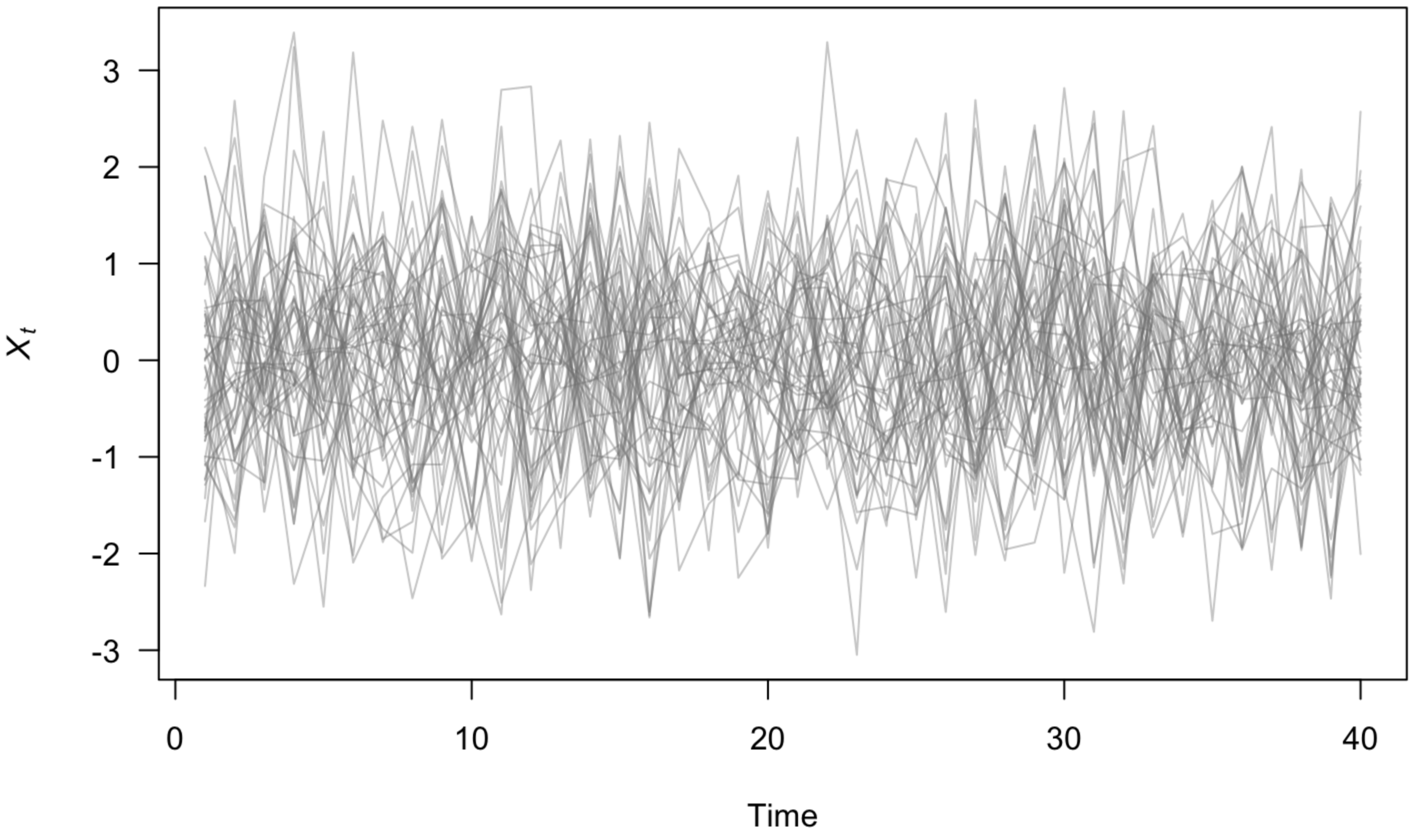


Number of lynx trapped in Canada from 1821-1934

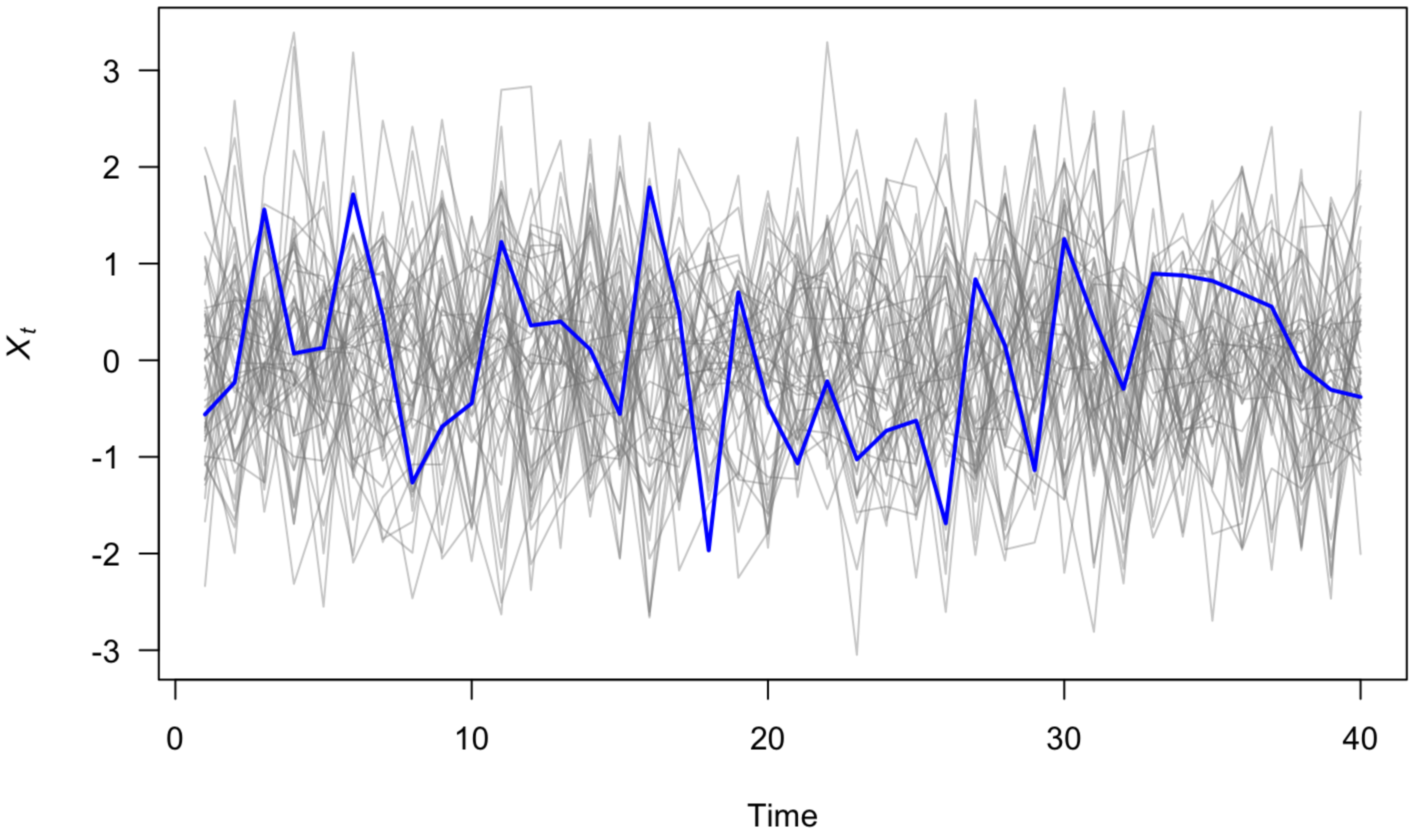
What is a time series model?

A *time series model* for $\{x_t\}$ is a specification of the *joint distributions* of a sequence of *random variables* $\{X_t\}$, of which $\{x_t\}$ is thought to be a realization

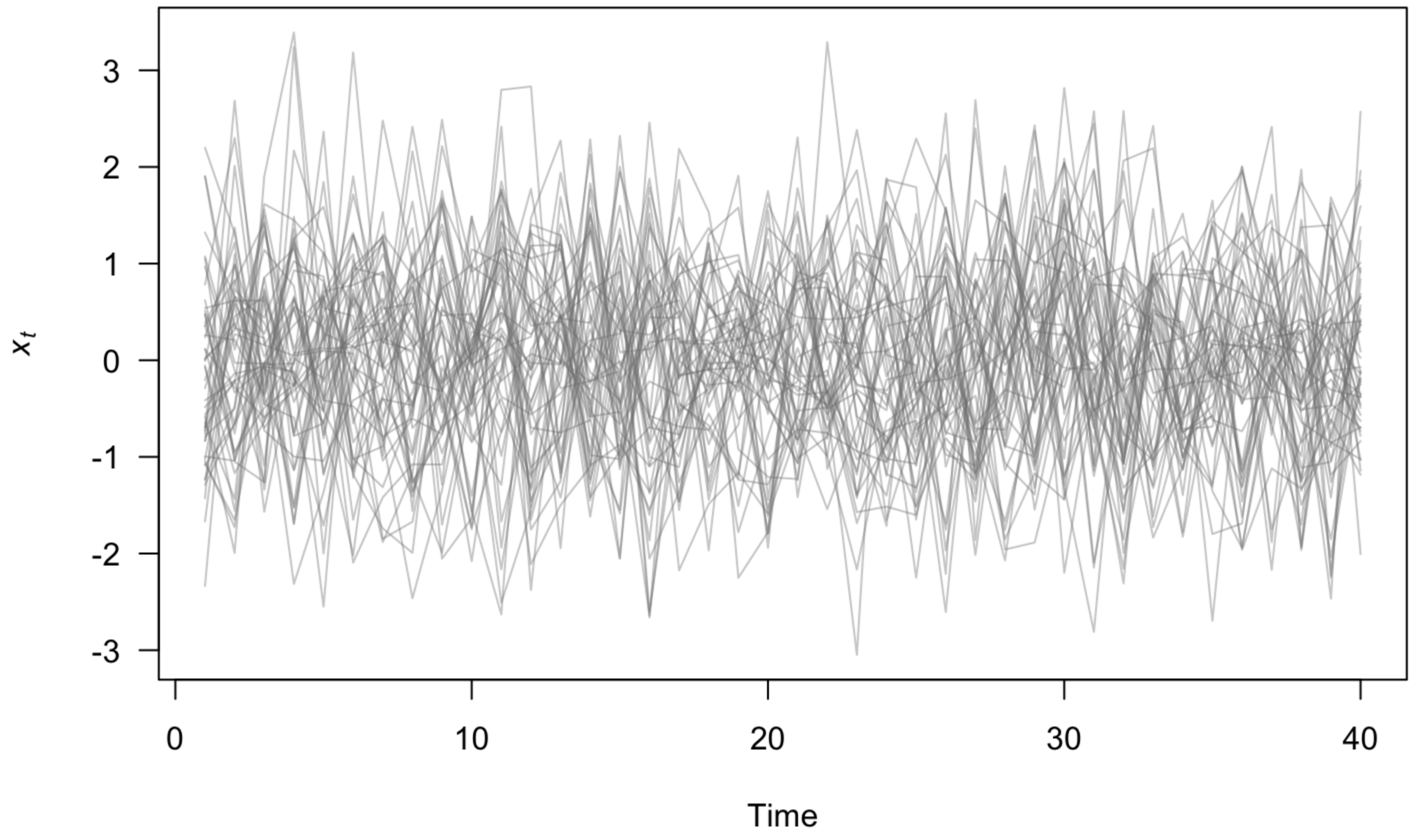
Joint distributions of random variables



We have one realization

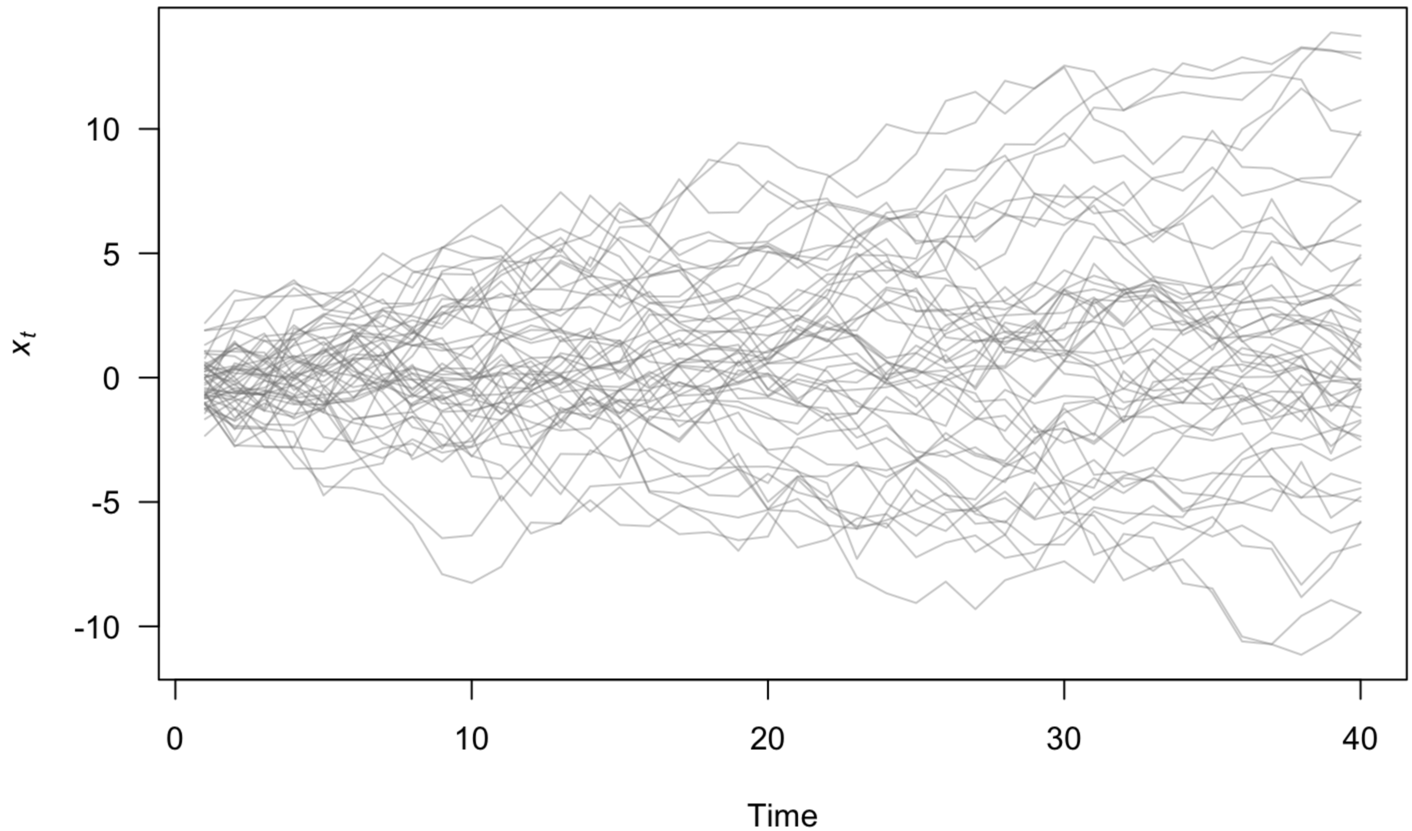


Some simple time series models



White noise: $x_t \sim N(0, 1)$

Some simple time series models



Random walk: $x_t = x_{t-1} + w_t, \text{ with } w_t \sim N(0, 1)$

Classical decomposition

Model time series $\{x_t\}$ as a combination of

1. trend (m_t)
2. seasonal component (s_t)
3. remainder (e_t)

$$x_t = m_t + s_t + e_t$$

Classical decomposition

1. The trend (m_t)

We need a way to extract the so-called *signal* from the *noise*

One common method is via “linear filters”

Linear filters can be thought of as “smoothing” the data

Classical decomposition

1. The trend (m_t)

Linear filters typically take the form

$$\hat{m}_t = \sum_{i=-\infty}^{\infty} \lambda_i x_{t+1}$$

Classical decomposition

1. The trend (m_t)

For example, a moving average

$$\hat{m}_t = \sum_{i=-a}^{a} \frac{1}{2a+1} x_{t+i}$$

Classical decomposition

1. The trend (m_t)

For example, a moving average

$$\hat{m}_t = \sum_{i=-a}^{a} \frac{1}{2a+1} x_{t+i}$$

If $a = 1$, then

$$\hat{m}_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

Classical decomposition

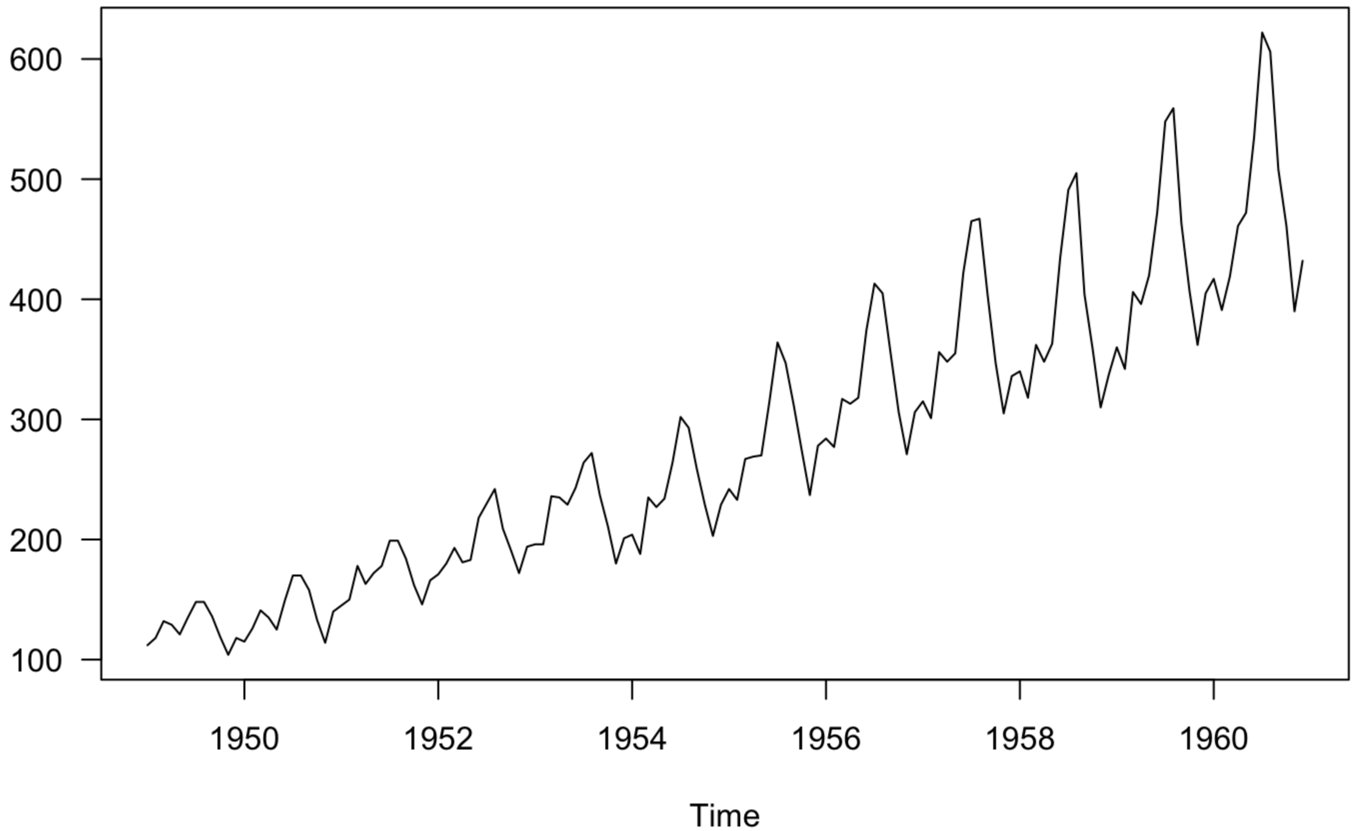
1. The trend (m_t)

For example, a moving average

$$\hat{m}_t = \sum_{i=-a}^{a} \frac{1}{2a+1} x_{t+i}$$

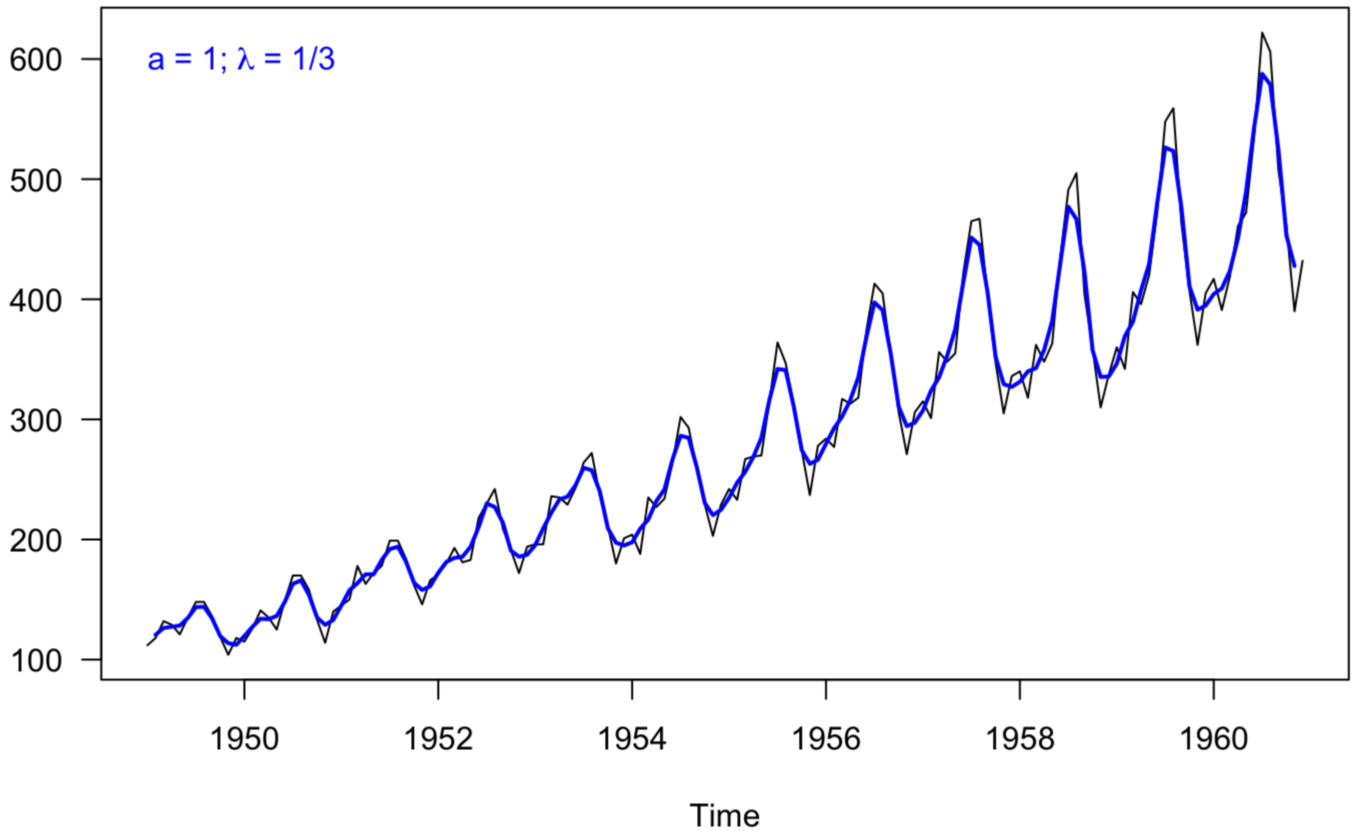
As a increases, the estimated trend becomes more smooth

Example of linear filtering



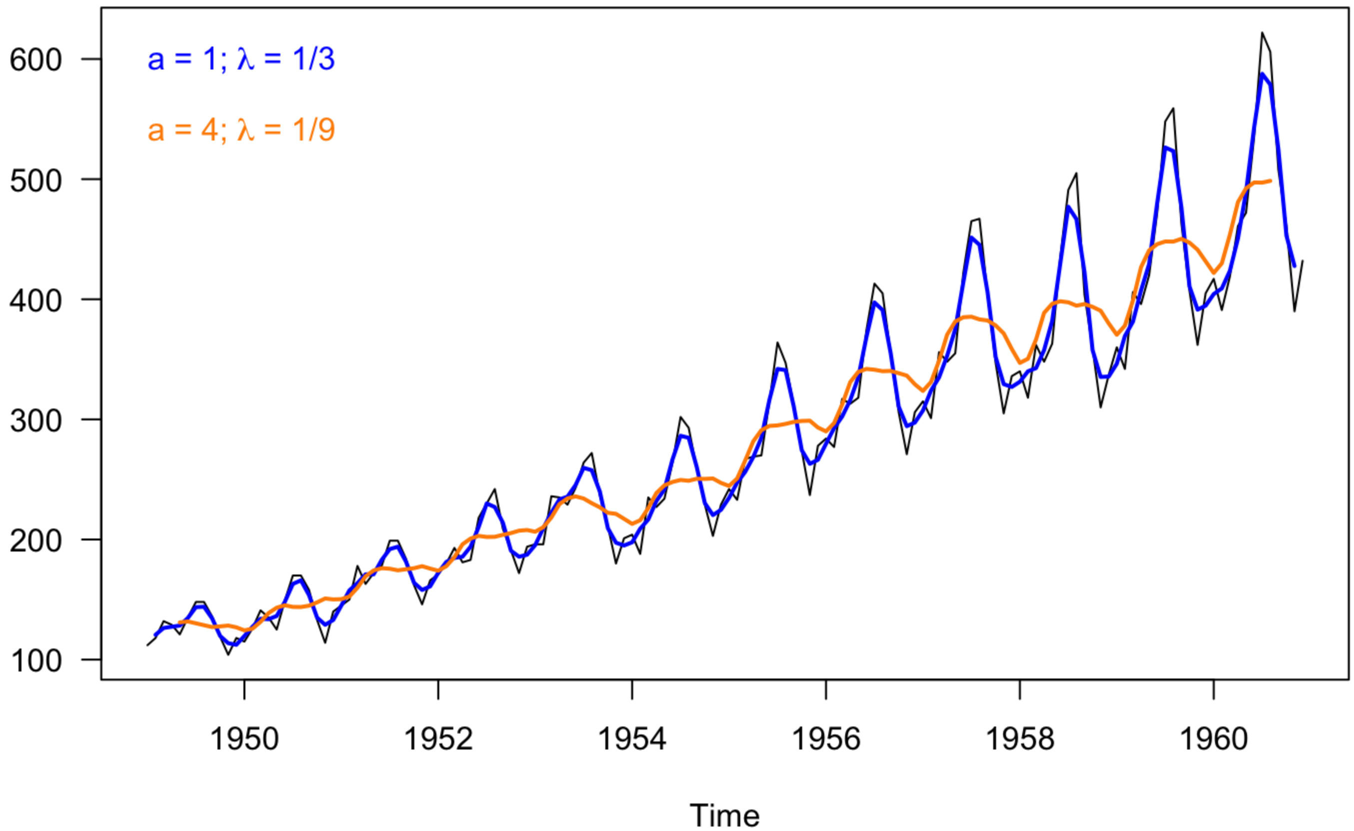
Monthly airline passengers from 1949-1960

Example of linear filtering



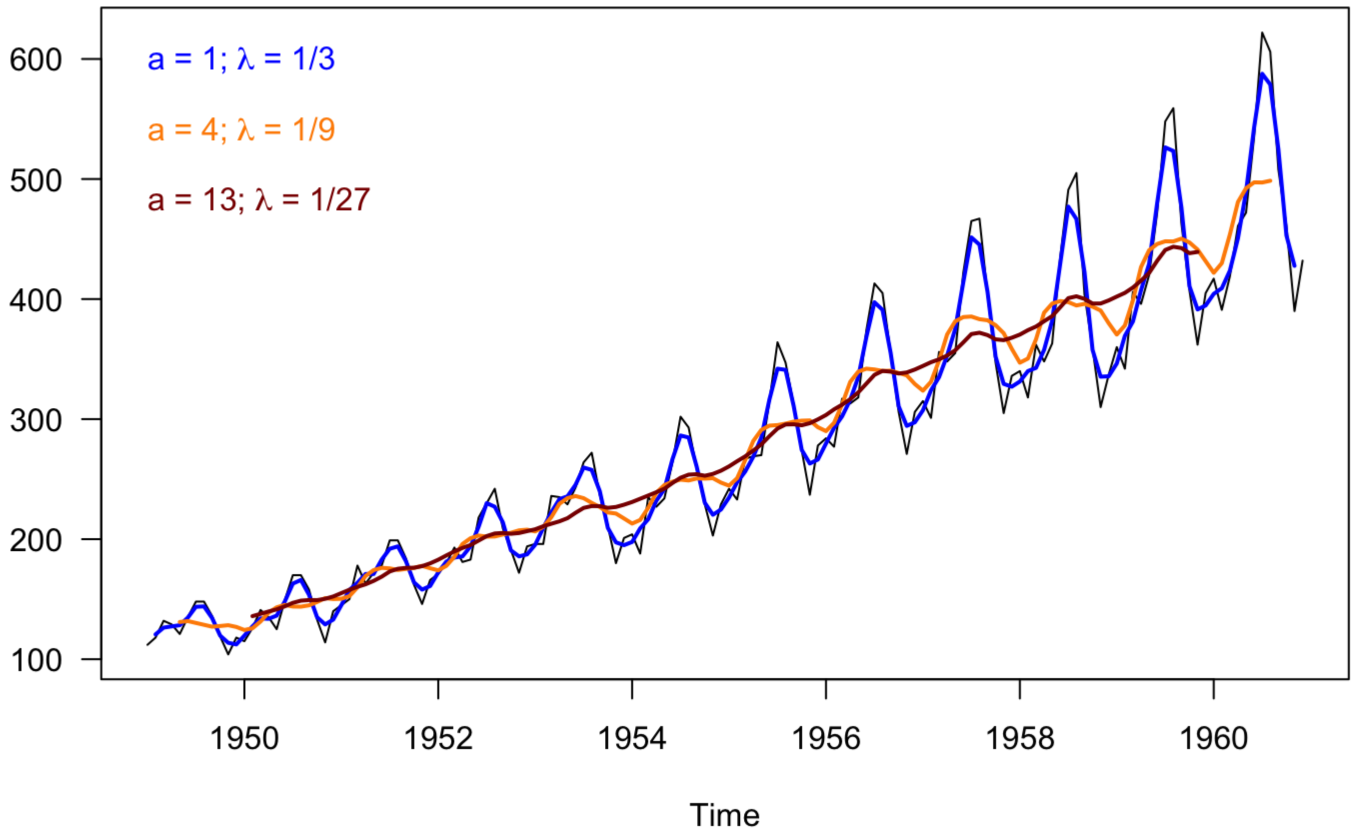
Monthly airline passengers from 1949-1960

Example of linear filtering



Monthly airline passengers from 1949-1960

Example of linear filtering



Monthly airline passengers from 1949-1960

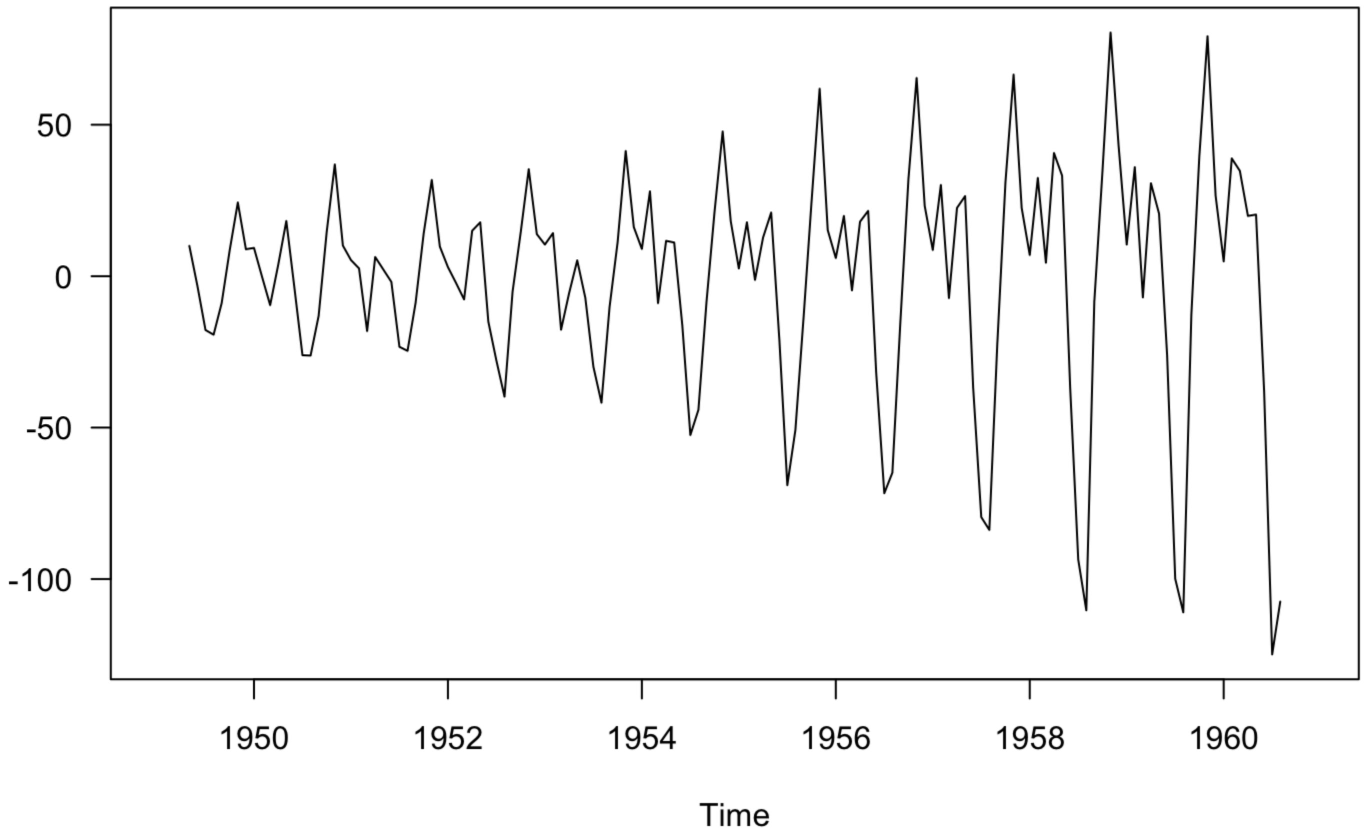
Classical decomposition

2. Seasonal effect (s_t)

Once we have an estimate of the trend $\hat{m}_t$, we can estimate $\hat{s}_t$ simply by subtraction:

$$\hat{s}_t = x_t - \hat{m}_t$$

Classical decomposition



Seasonal effect ($\hat{s}_t$), assuming $\lambda = 1/9$

Classical decomposition

2. Seasonal effect (s_t)

But, $\hat{s}_t$ really includes the remainder e_t as well

$$\begin{aligned}\hat{s}_t &= x_t - \hat{m}_t \\ (s_t + e_t) &= x_t - m_t\end{aligned}$$

Classical decomposition

2. Seasonal effect (s_t)

So we need to estimate the *mean* seasonal effect as

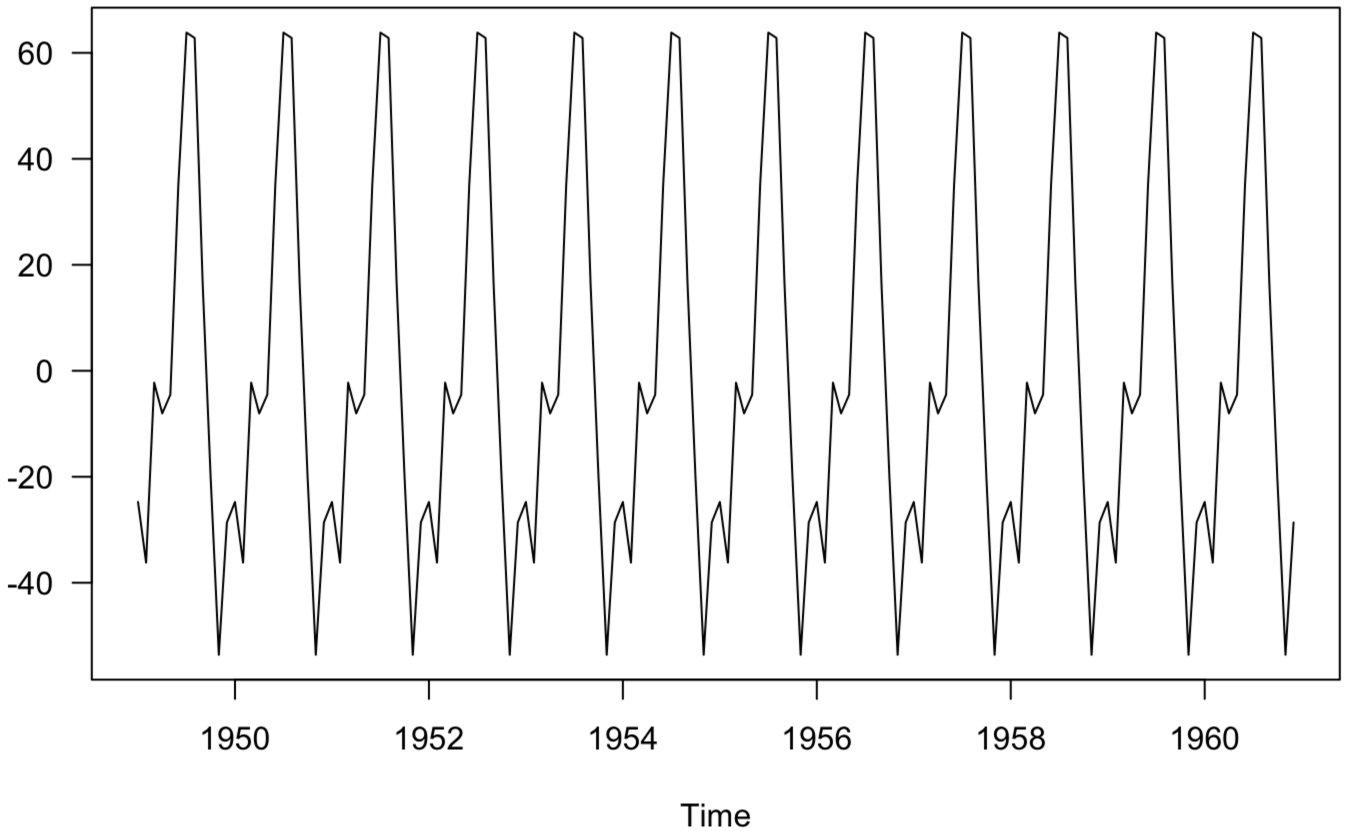
$$\hat{s}_{Jan} = \sum \frac{1}{(N/12)} \{s_1, s_{13}, s_{25}, \ldots\}$$

$$\hat{s}_{Feb} = \sum \frac{1}{(N/12)} \{s_2, s_{14}, s_{26}, \ldots\}$$

$$\vdots$$

$$\hat{s}_{Dec} = \sum \frac{1}{(N/12)} \{s_{12}, s_{24}, s_{36}, \ldots\}$$

Mean seasonal effect (s_t)



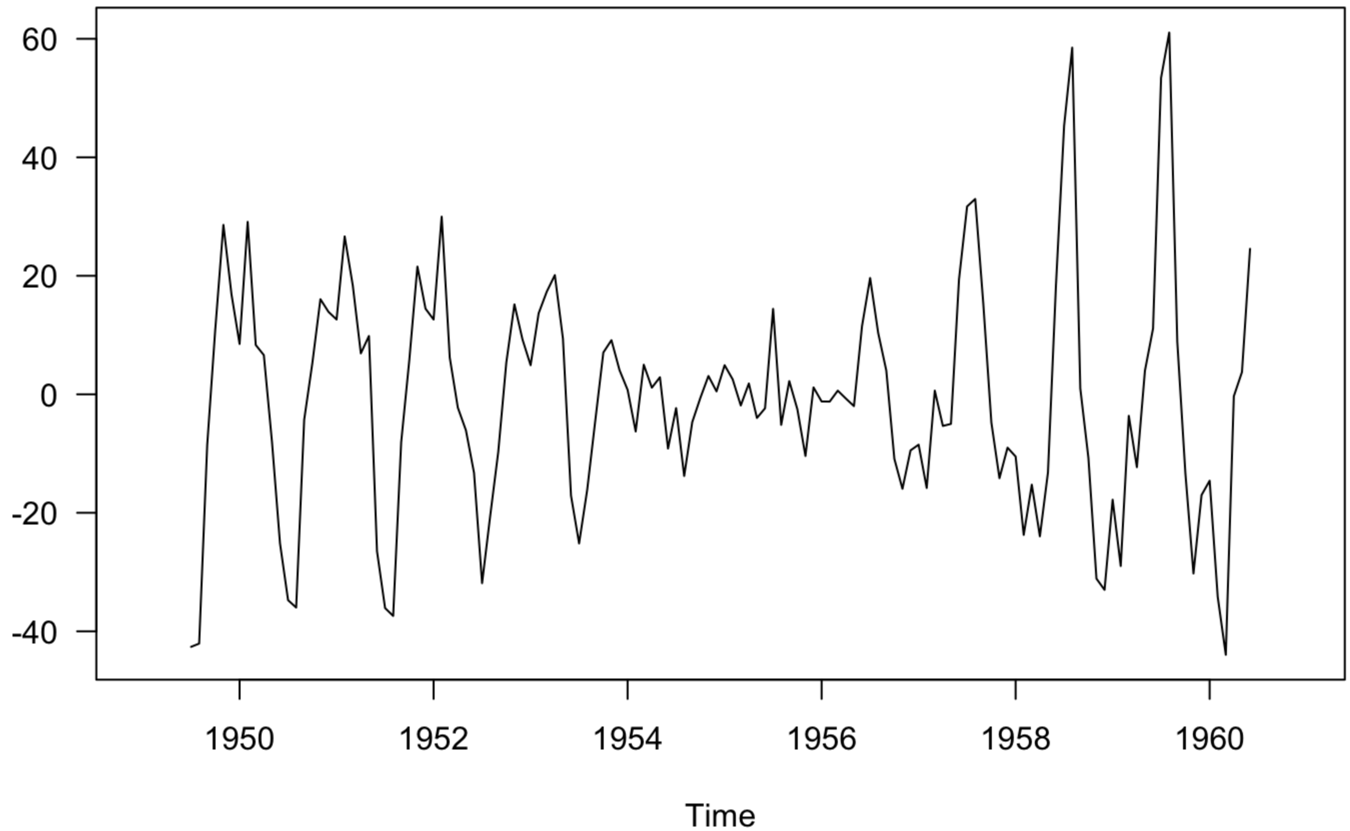
Classical decomposition

3. Remainder (e_t)

Now we can estimate e_t via subtraction:

$$\hat{e}_t = x_t - \hat{m}_t - \hat{s}_t$$

Remainder (e_t)

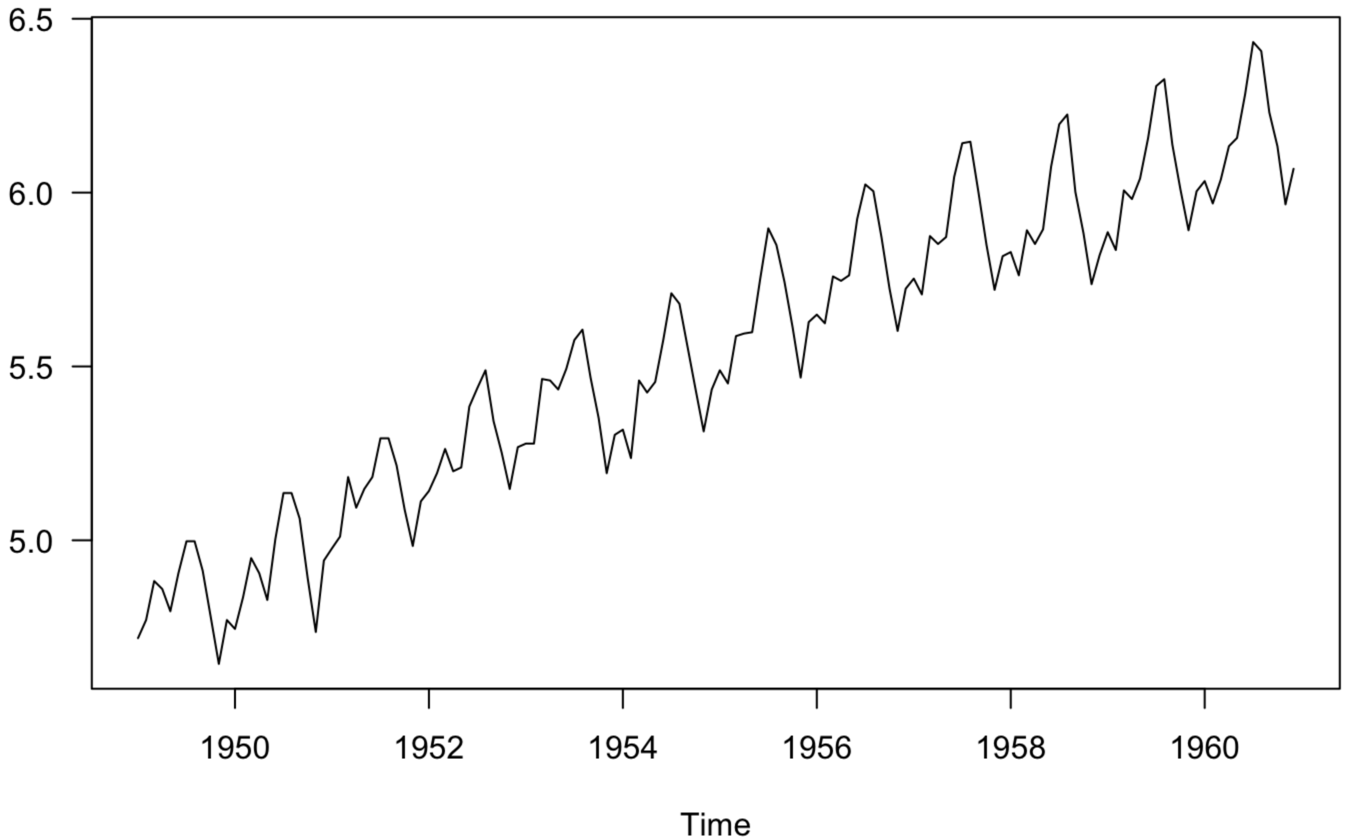


Let's try a different model

With some other assumptions

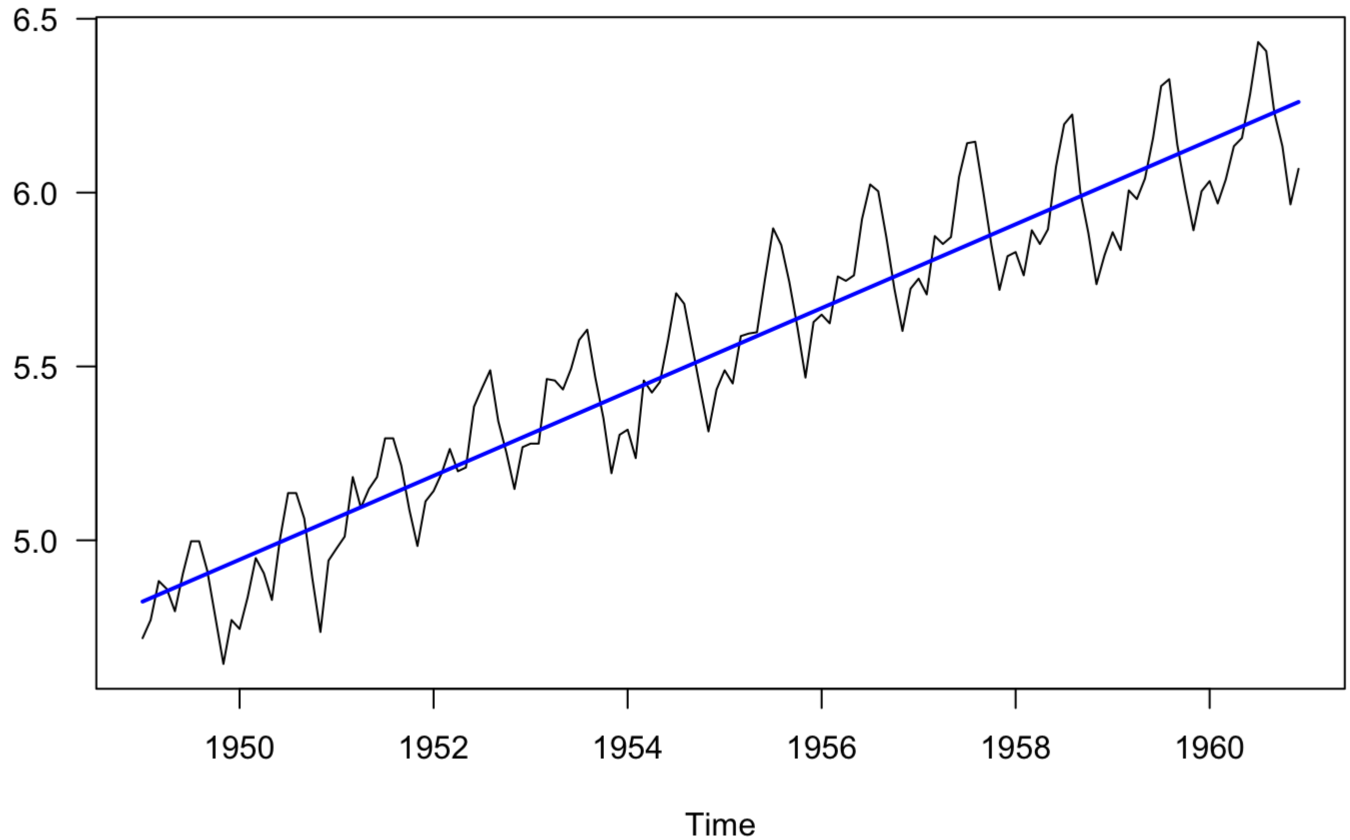
1. Log-transform data
2. Linear trend

Log-transformed data

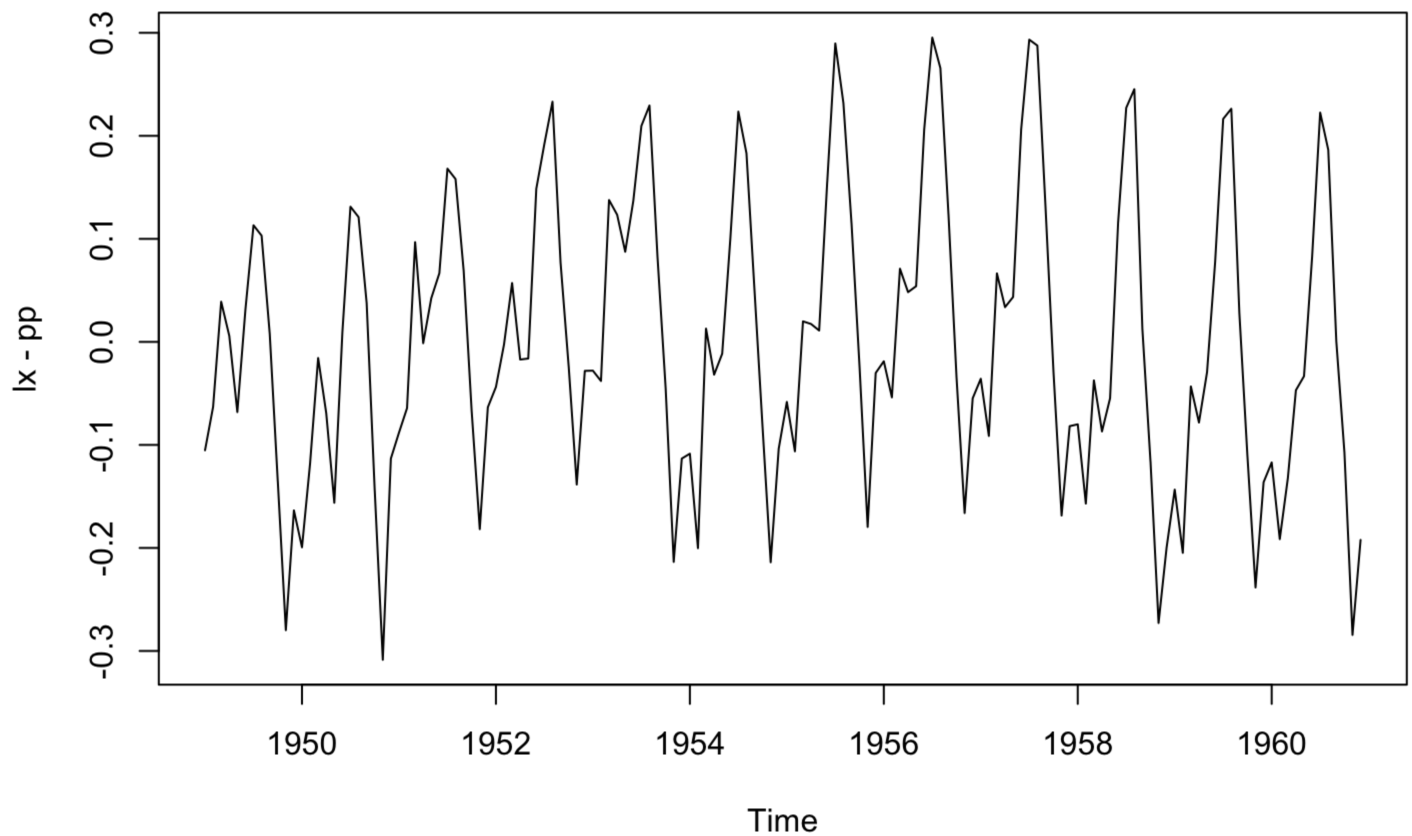


Monthly airline passengers from 1949-1960

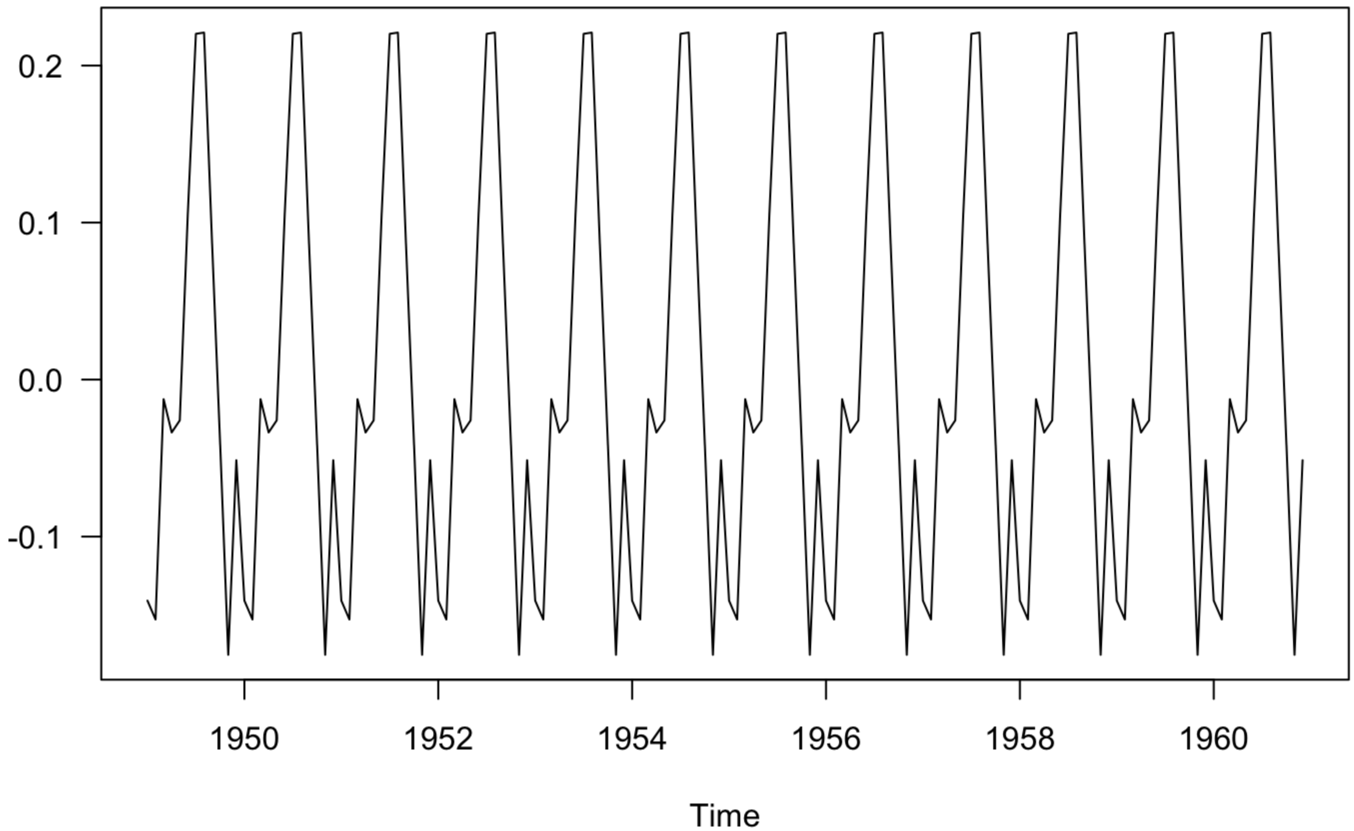
The trend (m_t)



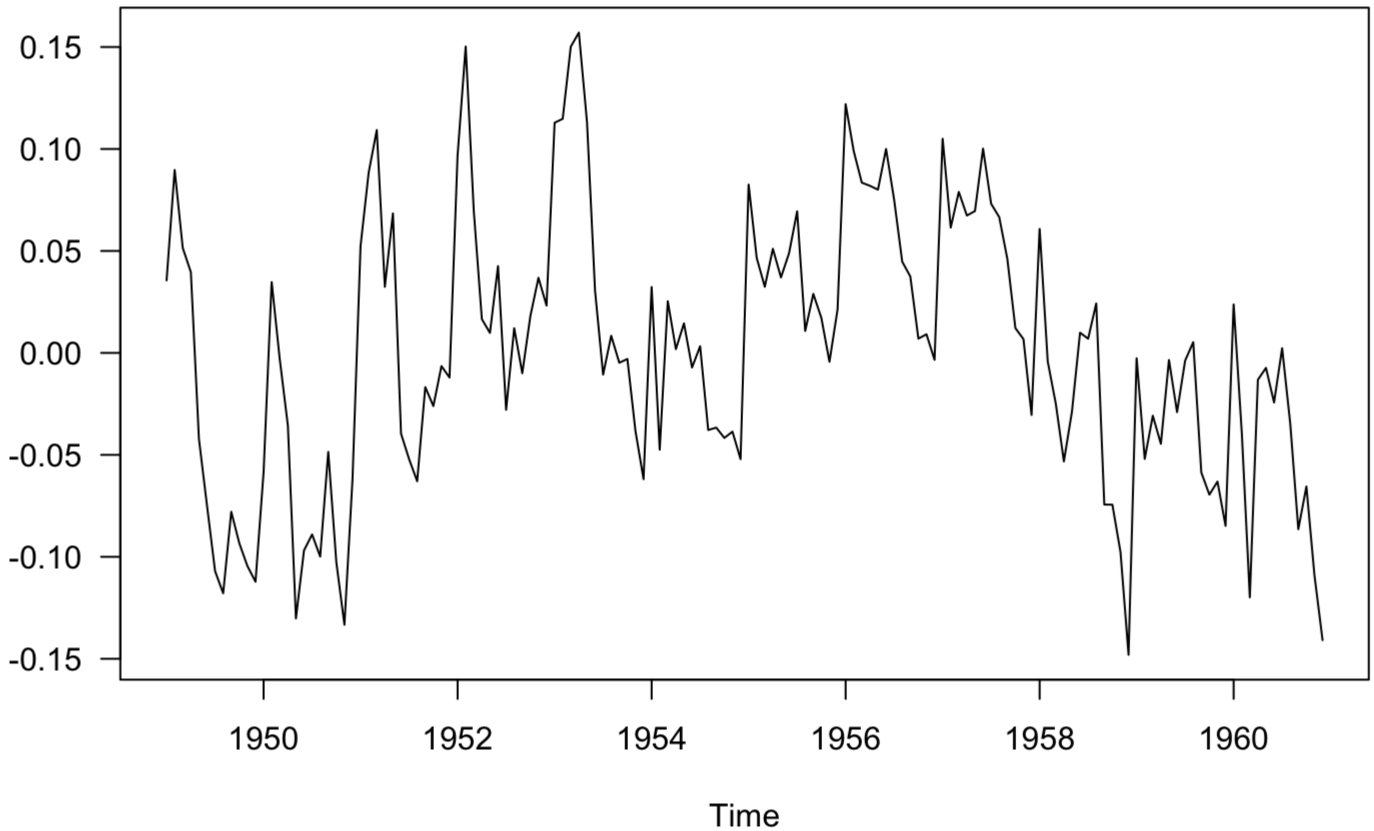
Seasonal effect (s_t) with error (e_t)



Mean seasonal effect (s_t)



Remainder (e_t)



Summary

Today's topics

Characteristics of time series (ts)

- What is a ts?
- Classifying ts
- Trends
- Seasonality (periodicity)

Classical decomposition