2.3 Exploratory Data Analysis

- In T.S, the dependence between the values of the series is important to measure
- If the dependence structure is not regular or is changing at every time point, it is difficult to measure the dependence
- It is crucial for the T.S to meet the weakly stationary conditions
- Need to play down the effects of nonstationary so that the stationary properties of the series may be studied
- Methods: lowess, filtering, differencing, kernel smoothing and smoothing splines etc,.
- General setup

$$x_t = f_t + y_t$$

  where f_t is some smooth function of time, and y_t is a stationary process.

Exmaple 2.3 Detrending Global Temperature

Suppose the model is of the form of

$$x_t = u_t + y_t,$$

where u_t denotes the trend, and y_t is a stationary process.

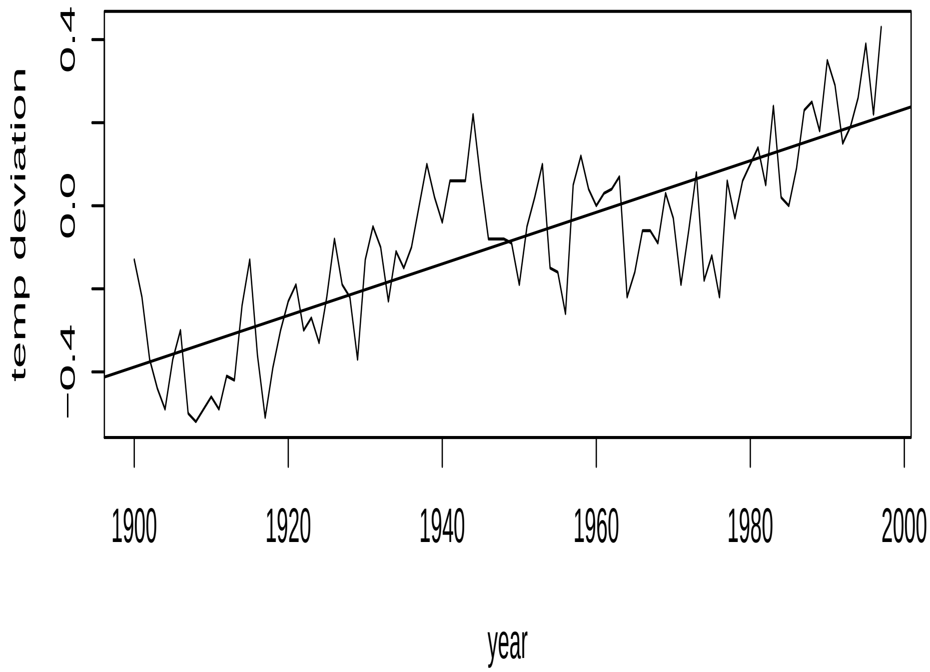
A straight line might be a reasonable model for the trend as we studied in Example 2.1, i.e,

$$u_t = \beta_1 + \beta_2 t,$$

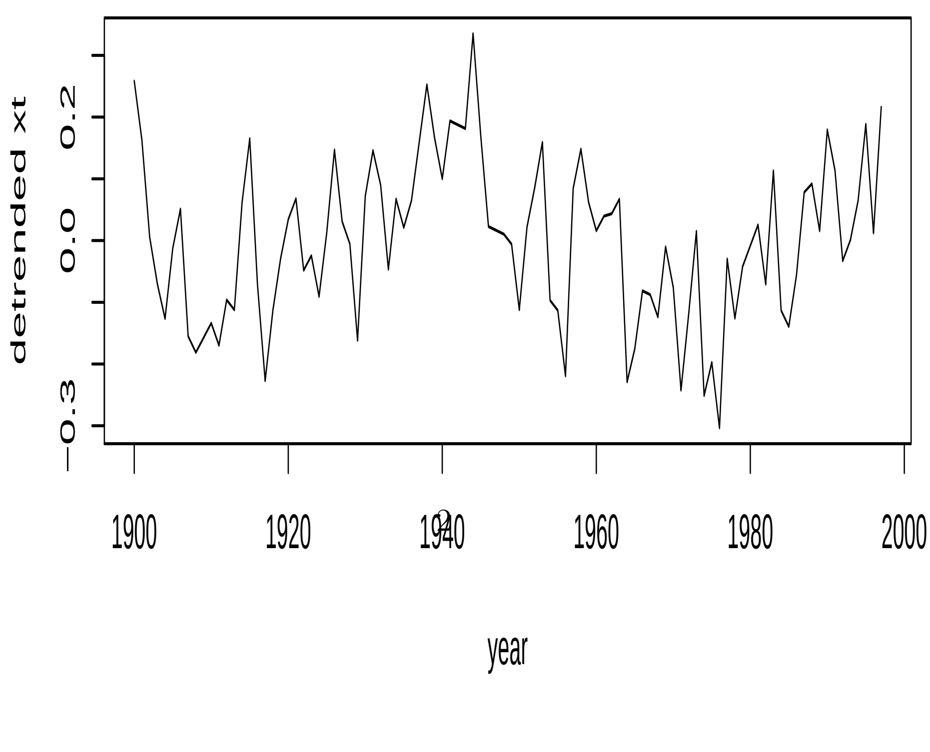
$$\hat{u}_t = -12.186 + .006t.$$

To obtain the detrended series, we simply subtract $\hat{u}_t$ from the observations x_t to obtain the detrended series

$$\hat{y}_t = x_t + 12.186 - .006t$$



Detrended series
Plot



```
gtemp=scan"http://www.stat.pitt.edu/stoffer/tsa2/data/globtemp.dat")
par(mfrow=c(2,1))
x=gtemp[45:142]
t=1900:1997
fit=lm(x ~ t) #regress x on t
summary(fit) #regression output
plot(t,x, type="l", xlab="year", ylab="temp deviation")
abline(fit$coef[1], fit$coef[2])
plot(t, fit$residuals, type="l", xlab="year", ylab="detrended xt", main=("Detrended series Plot"))
```

2.4 Smoothing in the Time Series Context

Example 2.12 Kernel Smoothing

Kernel smoothing is a moving average smoother that uses a weight function, or kernel, to average the observations.

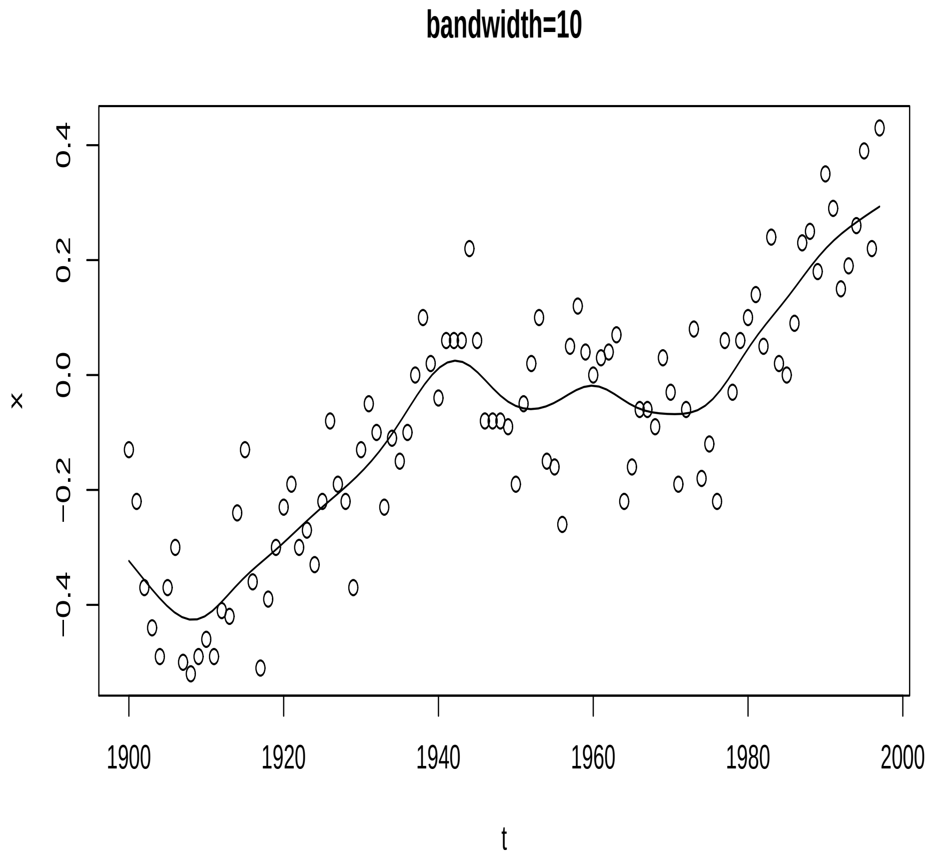
$f(t)$ is estimated by

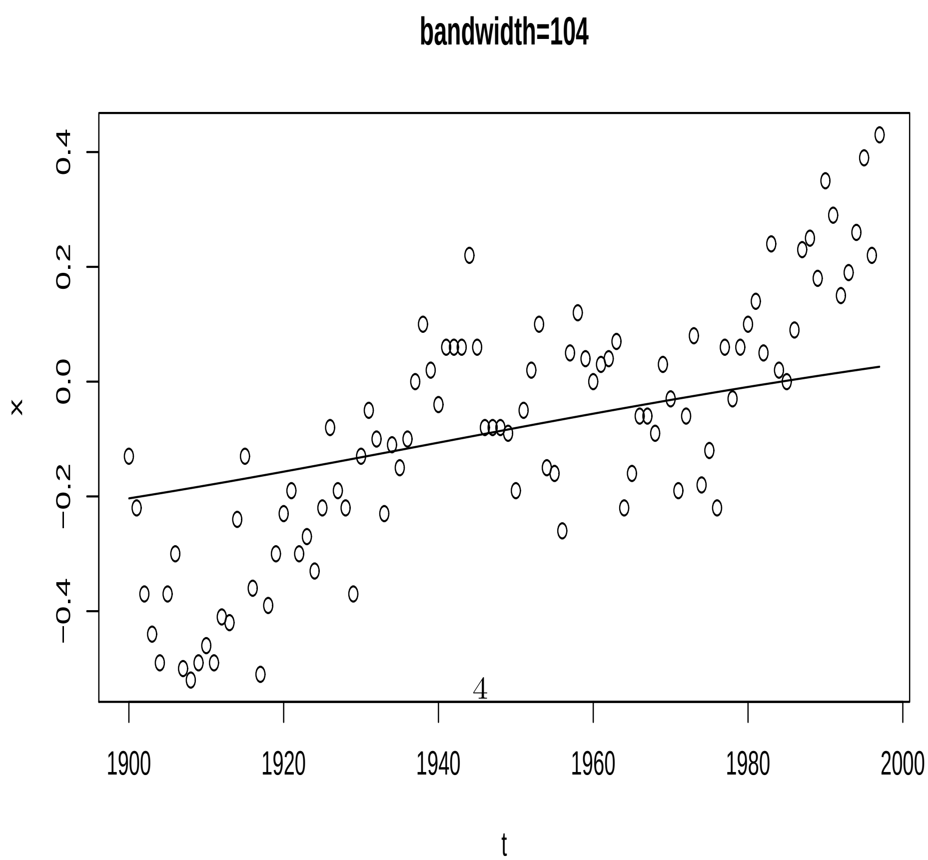
$$\hat{f}(t) = \sum_{i=1}^{n} w_t(i)x_t$$

where

$$w_t(i) = K(\frac{t-i}{b})/\sum_{j=1}^{n} K(\frac{t-j}{b})$$

The estimator is called the Naradaya-Watson estimator. $K(.)$ is a kernel function; typically, the normal kernel, $K(z) = \frac{1}{\sqrt{2\pi}}exp(-z^2/2)$, is used. $b = 10$ is roughly weighted monthly averages; $b = 104$ is roughly weighted yearly averages for the trend component.


bandwidth=10
x
t
1900
1920
1940
1960
1980
2000
−0.4
−0.2
0.0
0.2
0.4


bandwidth=104
x
t
1900
1920
1940
1960
1980
2000
−0.4
−0.2
0.0
0.2
0.4

plot(t,x, main="bandwidth=10")

lines(ksmooth(t, x,"normal", bandwidth=10))

plot(t,x, main="bandwidth=104")

lines(ksmooth(t, x,"normal", bandwidth=104))

Example 2.14 Smoothing Splines

- an extension of polynomial regression
- divide time $t = 1, \cdots, n$ into k intervals, $[t_0 = 1, t_1]$, $[t_1 + 1, t_2] \cdots, [t_{k-1} + 1, t_k = n]$.
- $t_0, t_1 \cdots, t_k$ are called knots.
- in each interval, fits a polynomial regression

$$f_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p$$

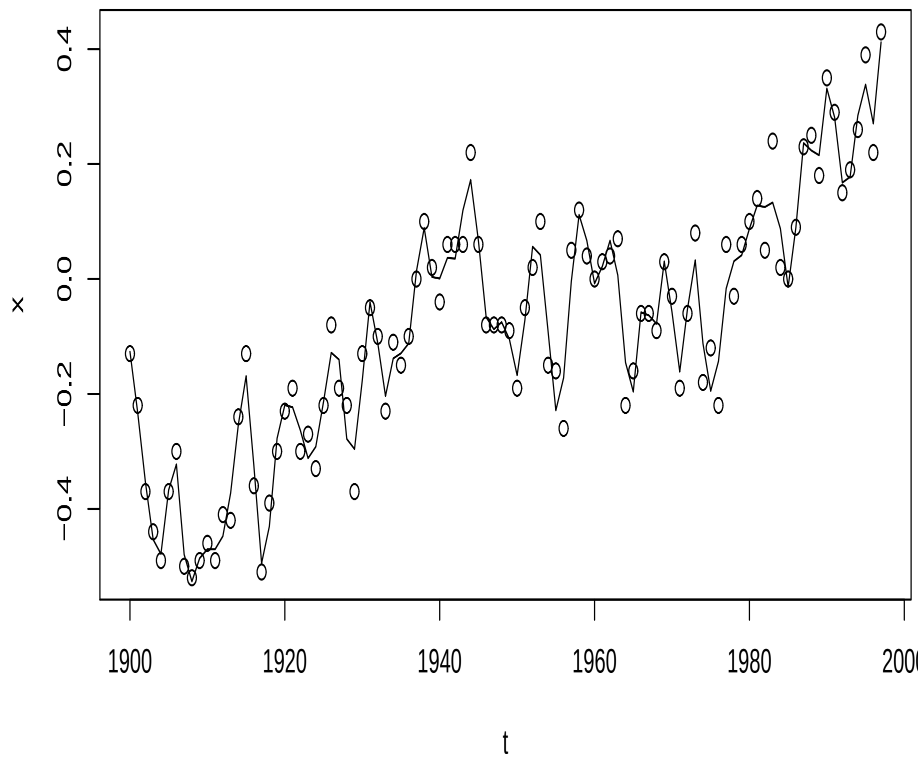
- a related method is smoothing splines, which minimize a compromise between the fit and the degree of smoothness given by

$$\sum_{i=1}^{n} [x_t - f_t]^2 + \lambda \int (f_t'')^2 dt,$$

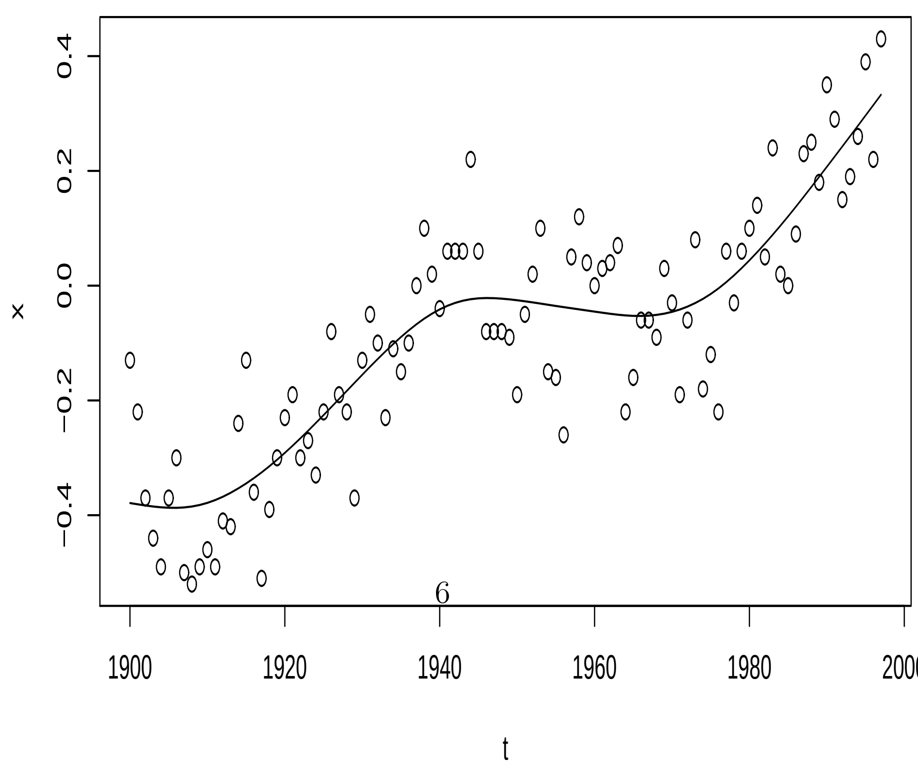
  where $f(t)$ is a cubic spline with a know at each t. The degree of smoothness is controlled by $\lambda > 0$.

- when $p = 3$, this is called cubic splines.

ss spar=.1



ss spar=.8



```
plot(t,x, main="ss spar=.1")
lines(smooth.spline(t, x, spar=.1))
plot(t,x, main="ss spar=.8")
lines(smooth.spline(t, x, spar=.8))
```

spar: smoothing parameter, typically (but not necessarily) in (0,1].