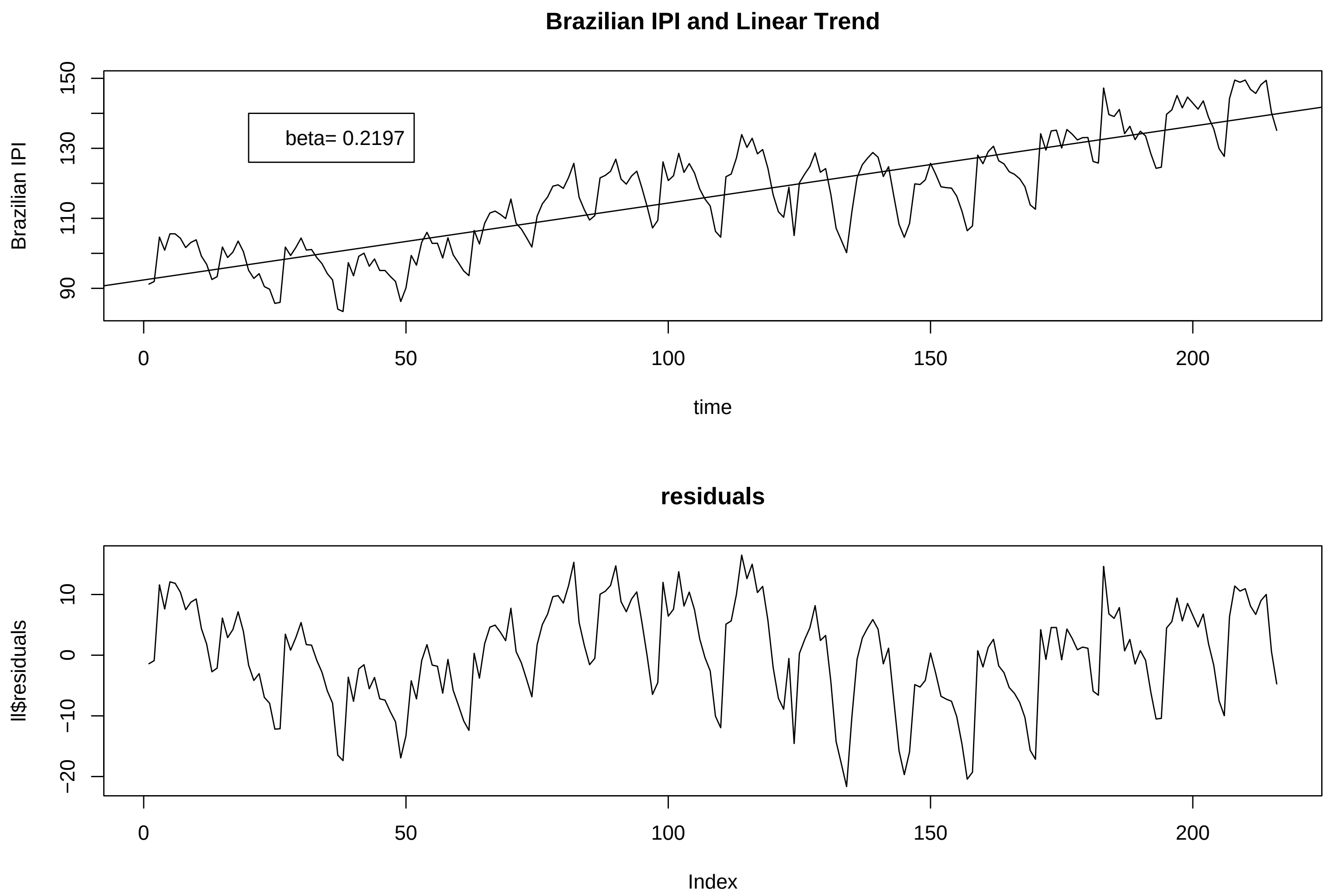
Seasonal variation

- "Variation in the data that is annual in period. For example, unemployment is typically 'high' in the winter but lower in the summer."
- Seasonality also refers to variation that is quarterly, monthly, weekly, etc. in period.
- Apart from seasonal effects, we can find a fixed period due to some other physical cause. Example: daily variability in temperature.
- Both concepts are also related to periodicity.
- If seasonality or cyclic variation are not of interest, they could be removed from the process.

Trend

- Loosely speaking 'long term change in mean' (20 years or 50 years?)
- A trend can be confounded with cyclic variation depending on the number of observations in our data.
- The simplest type of trend is *linear* for which $X_t = \alpha + \beta t + \omega_t$.
- α and β are constants. ω_t is a zero mean error.
- The mean level $\mu_t = \alpha + \beta t$ is also known as the "trend term".
- Others prefer to think of β as the trend; change of the mean level per unit of time.

- This model is too rigid and usually inadeaquate to fit real data.
- Consider again the Brazilian Industrial Production Index.


Brazilian IPI and Linear Trend
beta= 0.2197
Brazilian IPI
90
110
130
150
0
50
100
150
200
time
residuals
ll$residuals
−20
−10
0
10
0
50
100
150
200
Index

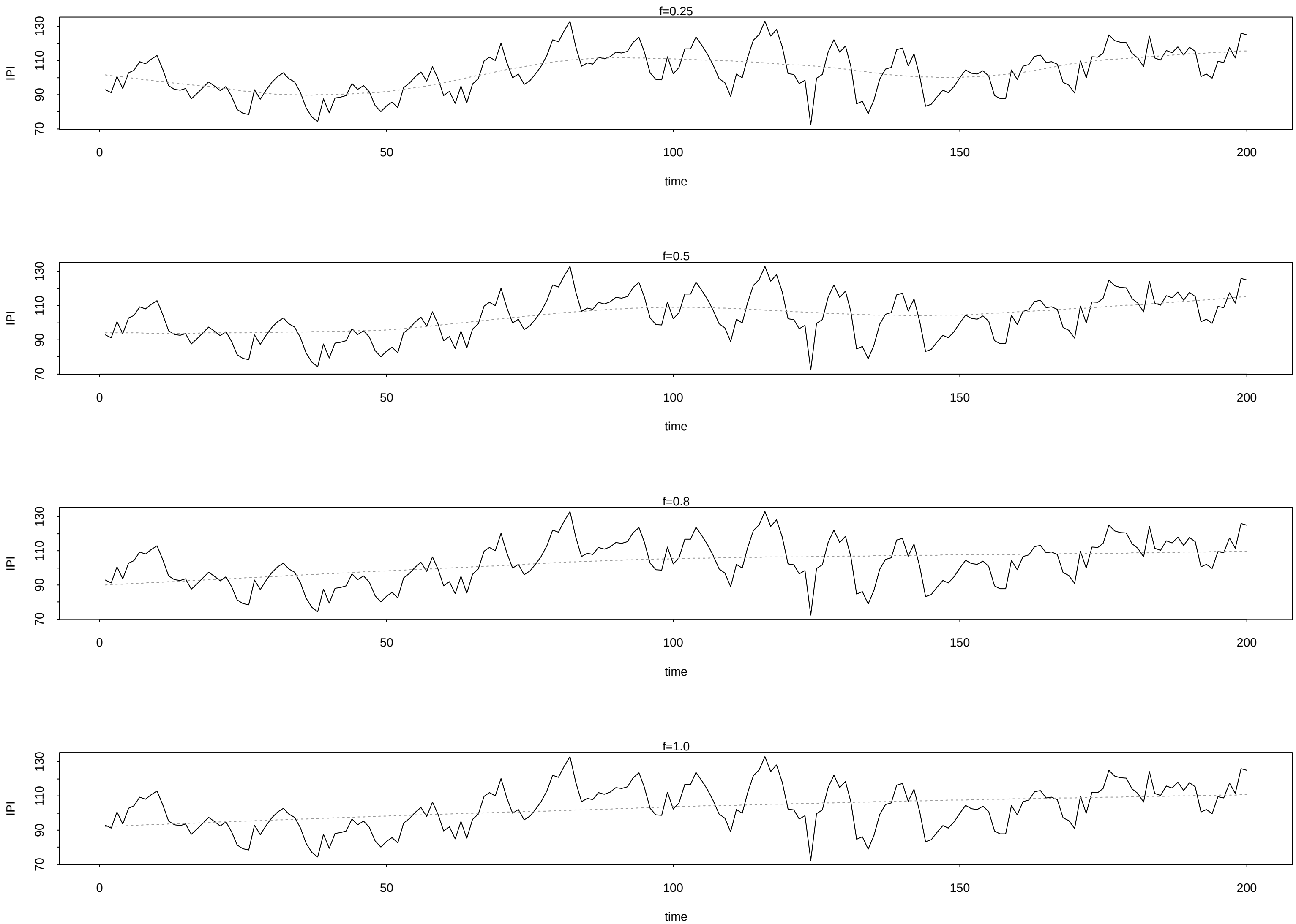
- Alternatively, we could use other *polynomials* to represent the mean level. For example: $X_t = \alpha + \beta t + \delta t^2 + \omega_t$ or $X_t = \alpha + \beta t + \delta t^2 + \gamma t^3 + \omega_t$.
- Still, these more general mean functions could be inadequate in practice.
- Another idea is to use locally linear regression. $X_t = \alpha_t + \beta_t t + \omega_t$
- Now the parameters α_t, β_t are time dependent.
- This falls within the state space modeling approach.

Lowess

- The trend can also be represented by a *lowess* estimator.
- This is based on locally least squares regression of the type $\alpha_t + \beta_t t$
- Lowess depends on a *smoothing parameter* denoted by f which basically determines the shape of the mean estimator.
- Different values of f must be tested before reporting a lowess curve.
- *lowess* is available as an R function so further information is available through the R help option.

  ```
  > help(lowess)
  ```

- Reference: Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74, 829-836.
- The following Figure shows different lowess estimators for the Brazilian IPI with $f = 0.25$, $f = 0.5$, $f = 0.8$ and $f = 1.0$.
- Notice that when f is small the estimated mean level is flexible.
- When f is close to one, the mean level resembles our regression line $\alpha + \beta t$.


f=0.25
f=0.5
f=0.8
f=1.0
IPI
time
70
90
110
130
0
50
100
150
200

Filtering

- Another exploratory procedure for dealing with a trend.
- The data x_t is converted into a "new" series y_t by the linear operation.

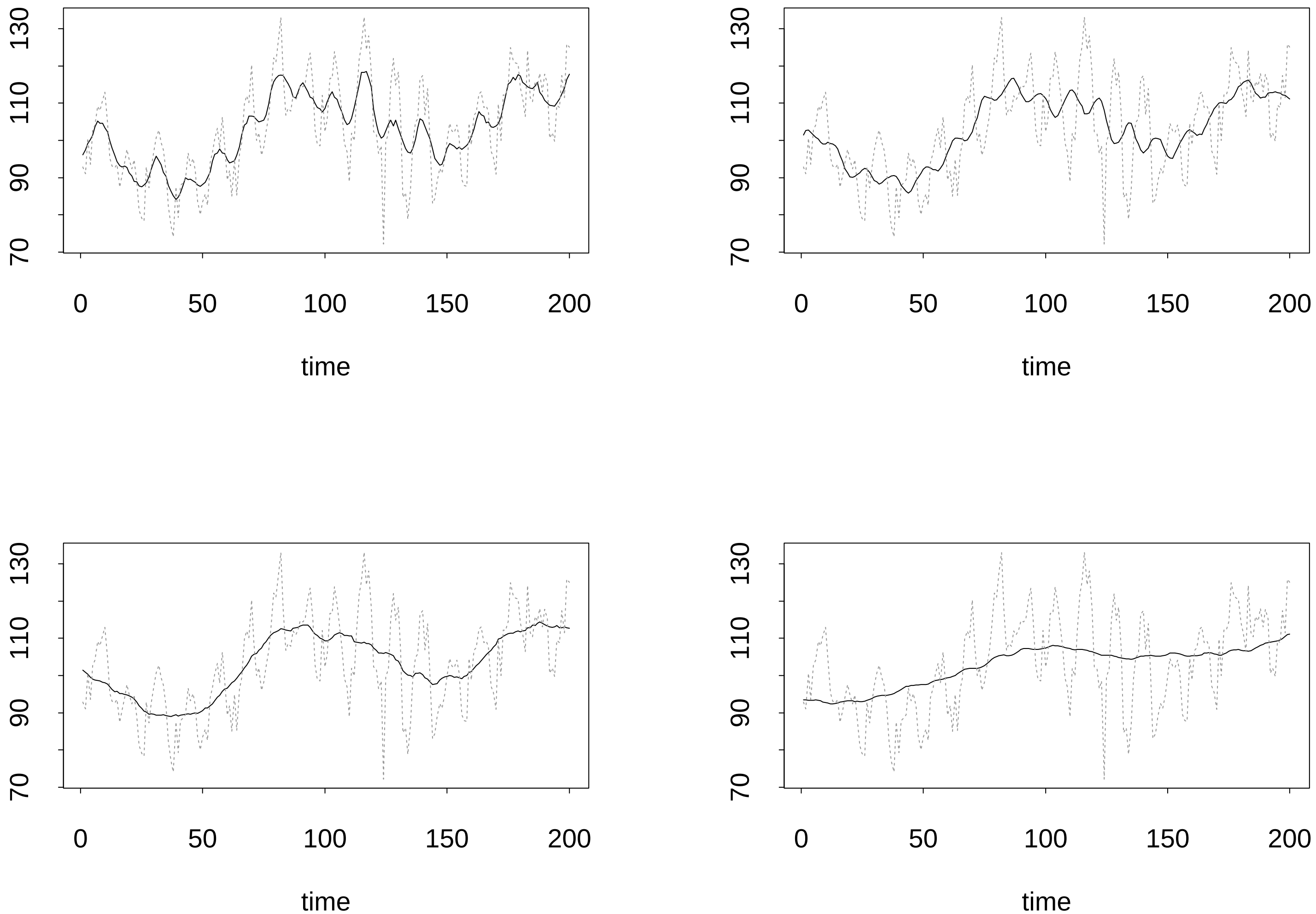
$$y_t = \sum_{j=-q}^{+s} w_j x_{t+j}$$

- The $w_j's$ are weights such that $\sum_{j=-q}^{+s} w_j = 1$
- y_t is considered a "smooth" version of x_t.
- This operation is also known as a *moving average.*
- Moving averages are often symmetric; $s = q$ and $a_j = a_{-j}$
- This is a weighted average of the q-previous and the q-next observation corresponding to each x_t.

- Simplest moving average; $w_j = 1/(2q+1); j = -q, \ldots, +q$
- The smooth value of x_t is given by

$$Sm(x_t) = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t+j}$$

- The next figure shows the Brazilian IPI with $Sm(x_t)$ for $q = 4, 8, 12, 40$.


time
time
time
time

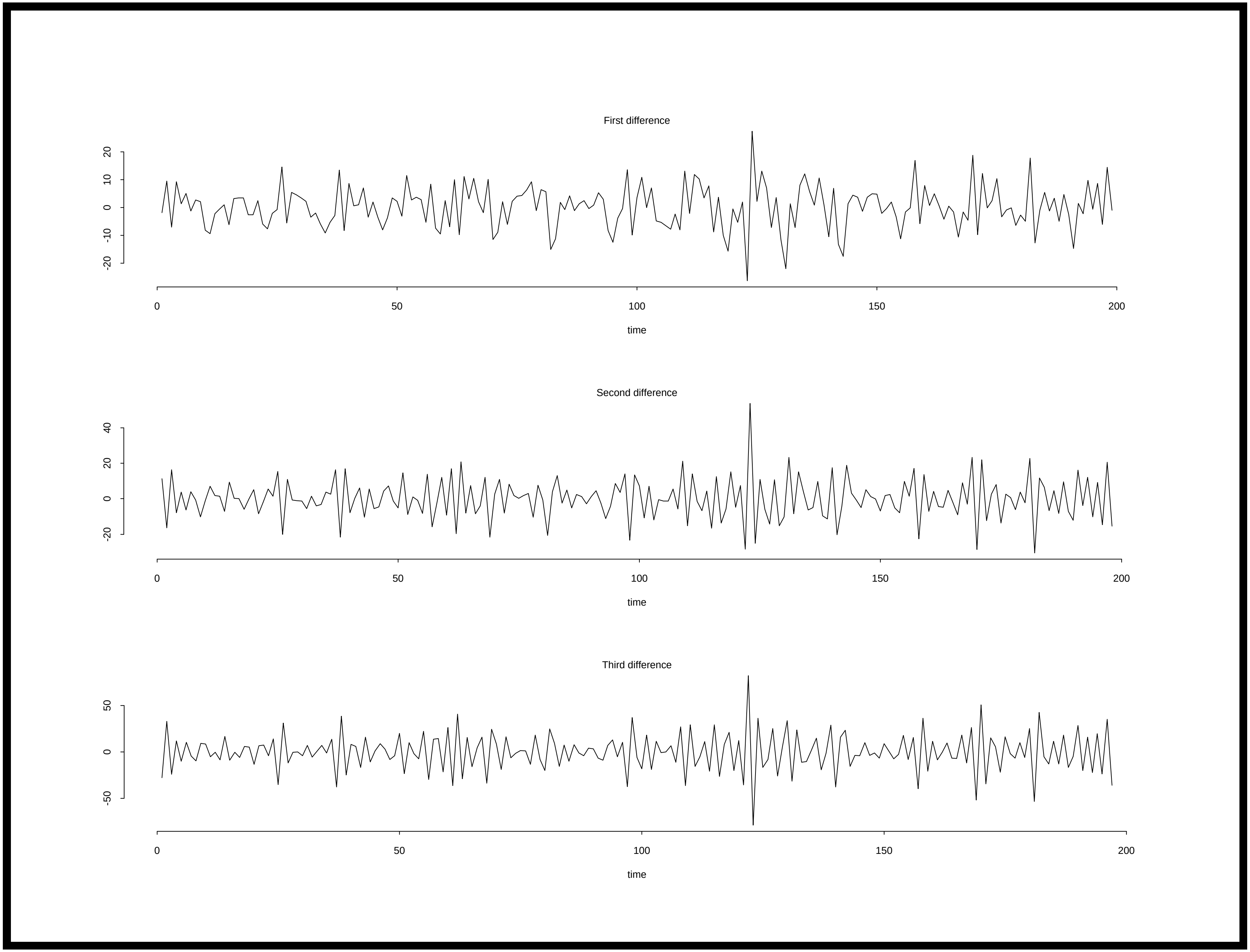
- This moving average is not always recommended to remove trend.
- Notice that $Sm(x_t)$ is only valid from $t = q + 1$ to $t = N - q$; N equal to the sample size.
- Alternatively, we could use *exponential smoothing*

$$x_t = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j}$$

- α is such that $0 < \alpha < 1$ and the weights $w_j = \alpha(1-\alpha)^j$ decay geometrically in j.
- This filter only involves present and past values of x_t.

Differencing

- The purpose is to remove the trend in the data.
- This technique is integral part of the so-called Box-Jenkins methodology.
- If data is non-seasonal usually a first order difference is sufficient.
- The first order difference of the data is defined as $z_t = x_t - x_{t-1} = \nabla x_t; t = 2, 3, \ldots, N$.
- The second difference is $\nabla^2 x_t = \nabla z_t = x_t - 2x_{t-1} + x_{t-2}; t = 3, \ldots, N$.
- d differences are denoted by $\nabla^d x_t$.
- Next: first, second and third differences for Brazilian IPI.


First difference
20
10
0
-10
-20
0
50
100
150
200
time
Second difference
40
20
0
-20
0
50
100
150
200
time
Third difference
50
0
-50
0
50
100
150
200
time

More R code

```
# Moving Average function
movavg <- function(x, k)
{
        m <- 2 * k + 1
        n <- length(x)
        y <- rep(0, n)
  for(i in 1:n) {
       k1 <- max( - k, (1 - i))
       k2 <- min(k, (n - i))
       y[i] <- mean(x[(i + k1):(i + k2)])
        }
  return(y)
}
```

```
m<- read.table("braipi",skip=1)
y <- as.vector(m[,2])
# Col 2 is the IPI
ym <- matrix(NA,nrow=length(y),ncol=4)
qvals <- c(4,8,12,40)
for(i in 1:4){ym[,i] <- movavg(y,qvals[i])}
par(mfrow=c(2,2))
for(i in 1:4){plot(ym[,i],xlab="time",
ylab=" ",type="l")}
mtext("The Brazilian IPI and different moving
average smoothers",outer=T)
# Lowess Estimators
par(mfrow=c(4,1))
l <- lowess(y,f=0.25)
```

```
plot(y,l$y,xlab="time",ylab="IPI",type="l")
mtext("f=0.25")
l<-lowess(y,f=0.5)
plot(y,l$y,xlab="time",ylab="IPI",type="l")
mtext("f=0.5")
l<-lowess(y,f=0.8)
plot(y,l$y,xlab="time",ylab="IPI",type="l")
mtext("f=0.8")
l<-lowess(y,f=1.0)
plot(y,l$y,xlab="time",ylab="IPI",type="l")
mtext("f=1.0")
mtext("Lowess estimators of trend for
Brazilian IPI data",outer=T)
```