

STAT 720 sp 2019 Lec 02

Karl Gregory

1/8/2019

Time series with a trend

Let

$$Y_t = m_t + \varepsilon_t, \quad t = 1, \dots, n,$$

where m_1, \dots, m_n are constants and $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a stationary time series with mean zero. We wish to estimate m_1, \dots, m_n .

Detrending with a parametric model

We could assume, for example, that the trend is a polynomial in t such that

$$m_t = \sum_{j=0}^k \alpha_j t^j,$$

for some $\alpha_0, \alpha_1, \dots, \alpha_k$. Then we could estimate the parameters with

$$(\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_k) = \underset{\alpha_0, \alpha_1, \dots, \alpha_k \in \mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^n \left(Y_t - \sum_{j=0}^k \alpha_j t^j \right)^2,$$

after which we would define the residuals as

$$\hat{\varepsilon}_t = Y_t - \sum_{j=0}^k \hat{\alpha}_j t^j, \quad t = 1, \dots, n.$$

Then one can check whether the residuals are stationary. Example:

```
data(uspop)
Y <- as.numeric(uspop)
t <- seq(1790,1970,by=10)
X <- cbind(1,t-1790,(t-1790)^2)

alpha.hat <- solve( t(X) %*% X) %*% t(X) %*% Y
m.hat <- X %*% alpha.hat

# define autocorrelation function
my.acf <- function(x,max.lag=12)
{

  n <- length(x)
  x.bar <- mean(x)

  gamma.hat <- numeric(max.lag+1)

  for(h in 0:min(max.lag,n-1))
  {
```

```

gamma.hat[h+1] <- 0

for(t in 1:(n-h))
{
    gamma.hat[h+1] <- gamma.hat[h+1] + (x[t] - x.bar)*(x[t+h] - x.bar)
}

}

gamma.hat <- gamma.hat / n
rho.hat <- gamma.hat / gamma.hat[1]

output <- list( gamma.hat = gamma.hat,
                rho.hat = rho.hat,
                lags = 0:max.lag)

return(output)
}

par(mfrow=c(2,2),mar=c( 5.1, 4.1,1.1, 2.1))

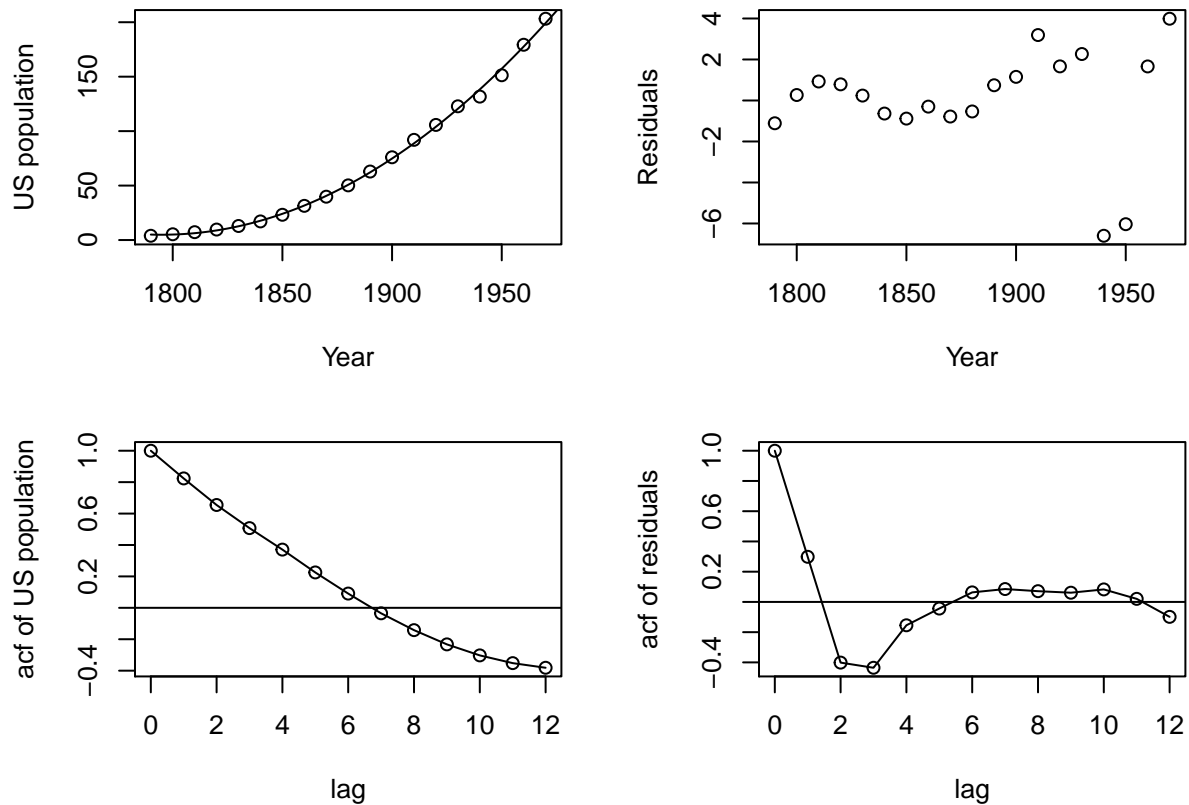
plot(Y~t,ylab="US population",xlab="Year")
t.seq <- seq(1790,1980,length=100)
lines(alpha.hat[1] + (t.seq - 1790)*alpha.hat[2]+(t.seq - 1790)^2*alpha.hat[3] ~ t.seq)

plot(Y-m.hat ~ t,ylab="Residuals",xlab="Year")

plot(my.acf(Y,max.lag=12)$rho.hat~c(0:12),type="o",
     ylab="acf of US population",xlab="lag")
abline(h=0)

plot(my.acf(Y-m.hat,max.lag=12)$rho.hat~c(0:12),type="o",
     ylab="acf of residuals",xlab="lag")
abline(h=0)

```



```
print(alpha.hat)
```

```
##           [,1]
## [1,]  5.041669173
## [2,] -0.063301533
## [3,]  0.006344589
```

Detrending with a moving average

Another way to estimate the trend is by taking a moving average of Y_1, \dots, Y_n . One way (out of many possible ways!) to do this is the following:

For some $q \geq 1$ define

$$Y_{1-q} = \dots = Y_0 = Y_1 \quad \text{and} \quad Y_{n+1} = \dots = Y_{n+q} = Y_n,$$

and then set

$$\hat{m}_t = (2q + 1)^{-1} \sum_{j=-q}^q Y_{t+j}.$$

Then the residuals are

$$\hat{\varepsilon}_t = Y_t - \hat{m}_t, \quad t = 1, \dots, n.$$

Then one can check whether the residuals are stationary. Example:

```
data(nhtemp)
Y <- as.numeric(nhtemp)
t <- 1912:1971
q <- 10
```

```

n <- length(Y)

Ymod <- c(rep(Y[1],q),Y,rep(Y[n],q))
m.hat <- numeric(n)
for(i in 1:n)
{
  m.hat[i] <- mean( Ymod[i:(i+2*q)] )
}

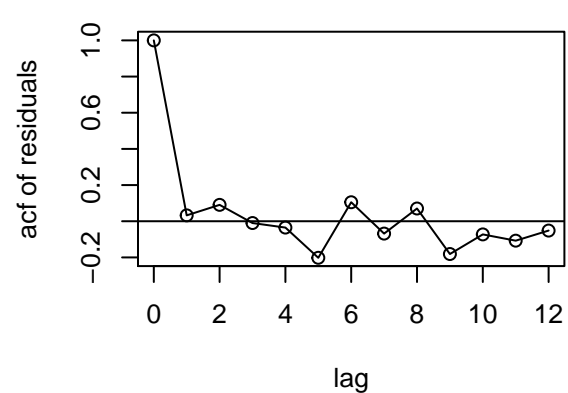
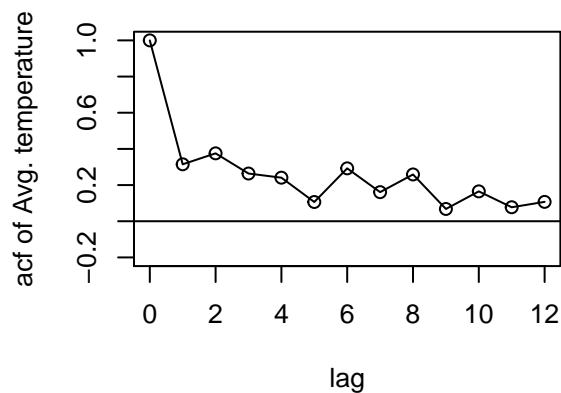
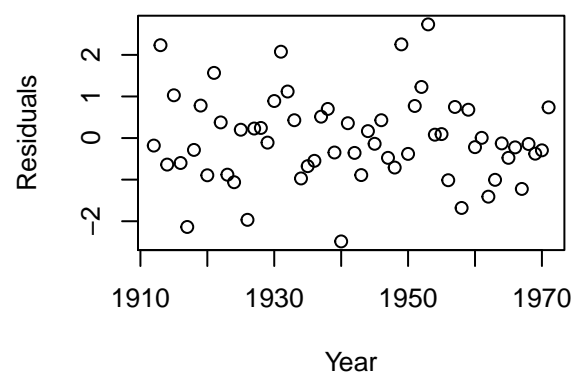
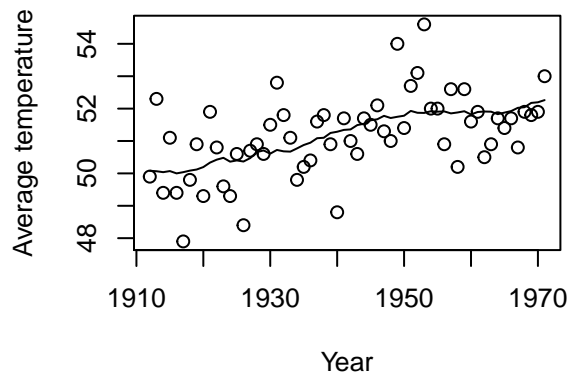
par(mfrow=c(2,2),mar=c( 5.1, 4.1,1.1, 2.1))
plot(Y~t,ylab="Average temperature",xlab="Year")
lines(m.hat~t)

plot(Y-m.hat~t,ylab="Residuals",xlab="Year")

plot(my.acf(Y,max.lag=12)$rho.hat~c(0:12),type="o",ylab="acf of Avg. temperature",
      xlab="lag", ylim=c(-.2,1))
abline(h=0)

plot(my.acf(Y-m.hat,max.lag=12)$rho.hat~c(0:12),type="o",ylab="acf of residuals",
      xlab="lag",ylim=c(-.2,1))
abline(h=0)

```



Detrending with differencing

Another approach is differencing. Define the backward shift operator B by

$$BY_t = Y_{t-1}$$

as well as the first difference operator ∇ by

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - B)Y_t.$$

For $j \geq 1$, define powers of the operators as

$$B^j Y_t = Y_{t-j} \quad \text{and} \quad \nabla^j Y_t = \nabla(\nabla^{j-1} Y_t), \quad \text{with} \quad \nabla^0 Y_t := Y_t.$$

This gives, for example,

$$\nabla^2 Y_t = \nabla(\nabla Y_t) = (1 - B)(1 - B)Y_t = 1 - 2BY_t + B^2 Y_t = 1 - 2Y_{t-1} + Y_{t-2}.$$

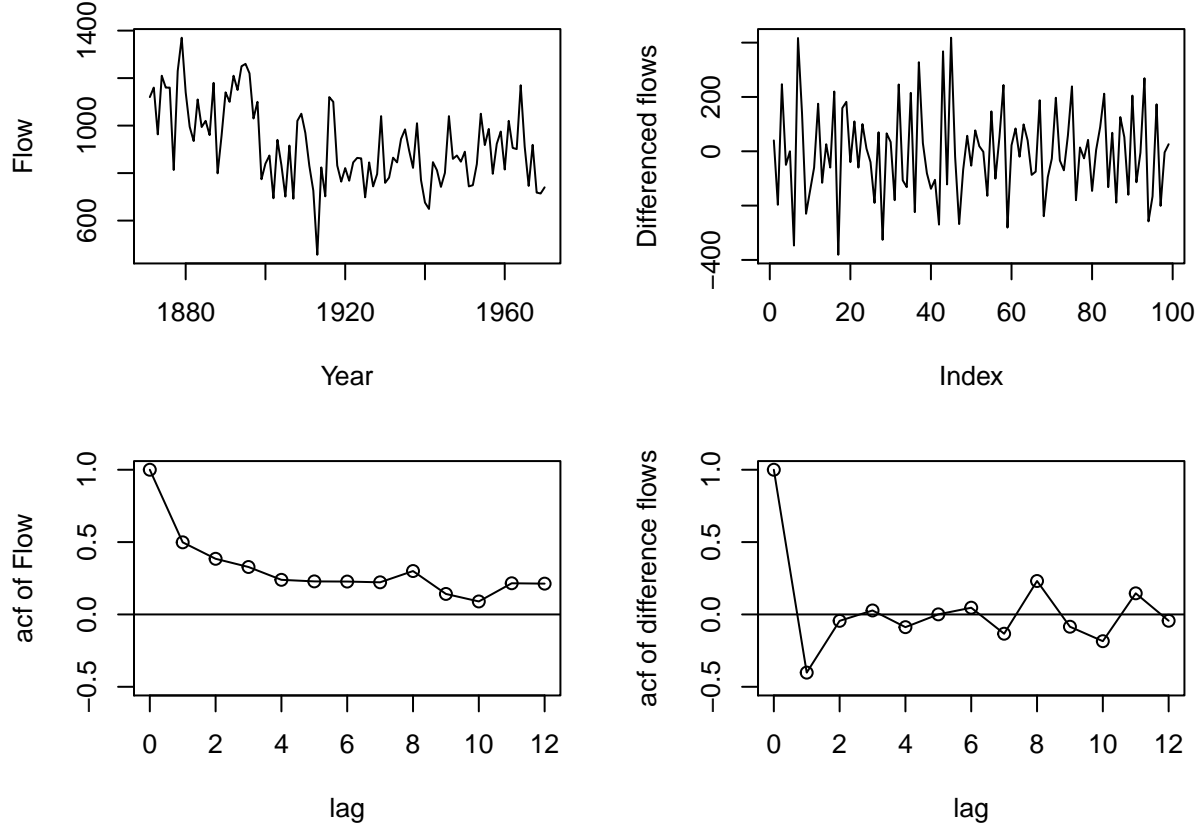
If the trend is a polynomial of degree k , then k th-order differencing eliminates the trend (See prob. 1.4 of B&D Theory). After differencing, we can check whether the differenced series is stationary. Example:

```
data(Nile)
Y <- as.numeric(Nile)
t <- 1871:1970

par(mfrow=c(2,2),mar=c( 5.1, 4.1,1.1, 2.1))
plot(Y~t,xlab="Year",ylab="Flow",type="l")
plot(diff(Y),type="l",ylab="Differenced flows")

plot(my.acf(Y,max.lag=12)$rho.hat~c(0:12),type="o",ylab="acf of Flow",
      xlab="lag",ylim=c(-.5,1))
abline(h=0)

plot(my.acf(diff(Y),max.lag=12)$rho.hat~c(0:12),type="o",ylab="acf of difference flows",
      xlab="lag",ylim=c(-.5,1))
abline(h=0)
```



Time series with trend and seasonal components

Let

$$Y_t = m_t + s_t + \varepsilon_t, \quad t = 1, \dots, n,$$

where m_1, \dots, m_n and s_1, \dots, s_n are constants such that for some $d \geq 1$, $s_t = s_{t+d}$ and $\sum_{j=1}^d s_{t+j} = 0$ for all t and where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a stationary time series with mean zero. We wish to estimate m_1, \dots, m_n and s_1, \dots, s_n . The constant d is called the period of the seasonal effect.

Detrending and deseasonalizing with moving average estimation

The following is one way to estimate m_1, \dots, m_n and s_1, \dots, s_n :

First estimate the trend with a moving average, setting the width of the moving average window equal to the period of the seasonality; more precisely, choose q such that $d = 2q + 1$ if d is odd and such that $d = 2q$ if d is even. Then define, as before,

$$Y_{1-q} = \dots = Y_0 = Y_1 \quad \text{and} \quad Y_{n+1} = \dots = Y_{n+q} = Y_n,$$

and for $t = 1, \dots, n$ set

$$\hat{m}_t = \begin{cases} (2q+1)^{-1} \sum_{j=-q}^q Y_{t+j} & \text{if } d \text{ is odd} \\ (2q)^{-1} (0.5Y_{t-q} + \sum_{j=-(q-1)}^{q-1} Y_{t+j} + 0.5Y_{t+q}) & \text{if } d \text{ is even.} \end{cases}$$

Assume n/d is an integer and denote

$$(s_1, \dots, s_n) = (s_{1,1}, \dots, s_{1,d}, \dots, s_{n/d,1}, \dots, s_{n/d,d}).$$

For $k = 1, \dots, d$, define

$$\hat{w}_k = (n/d)^{-1} \sum_{j=1}^{n/d} (Y_{(j-1)d+k} - \hat{m}_{(j-1)d+k}),$$

and set

$$\hat{s}_{1,k} = \dots = \hat{s}_{n/d,k} = \hat{w}_k - d^{-1} \sum_{j=1}^d \hat{w}_j.$$

From these, define

$$(\hat{s}_1, \dots, \hat{s}_n) = (\hat{s}_{1,1}, \dots, \hat{s}_{1,d}, \dots, \hat{s}_{n/d,1}, \dots, \hat{s}_{n/d,d}).$$

The centering of the $\hat{w}_1, \dots, \hat{w}_d$ ensures that $\sum_{j=1}^d \hat{s}_{t+j} = 0$ for all t .

The residuals are then

$$\hat{\varepsilon}_t = Y_t - \hat{m}_t - \hat{s}_t, \quad \text{for } t = 1, \dots, n.$$

One can check whether the residuals appear stationary. Example:

```
data(AirPassengers)

Y <- as.numeric(AirPassengers)
n <- length(Y)
t <- 1:n

d <- 12
q <- d/2

Ymod <- c(rep(Y[1],q),Y,rep(Y[n],q))
m.hat <- numeric(n)
for(i in 1:n)
{
  m.hat[i] <- (.5*Ymod[i] + sum(Ymod[(i+1):(i+2*q-1)]) + .5*Ymod[(i+2*q)])/(2*q)
}

w.hat <- apply(matrix(Y - m.hat,d,n/d),1,mean)
s.hat.d <- w.hat - mean(w.hat)
s.hat <- rep(s.hat.d,n/d)

par(mfrow=c(2,3),mar=c( 5.1, 4.1,1.1, 2.1))
plot(Y~t,type="l")
lines(m.hat~t,lty=2)

plot(Y - m.hat ~ t,type="l")
lines(s.hat~t,lty=2)

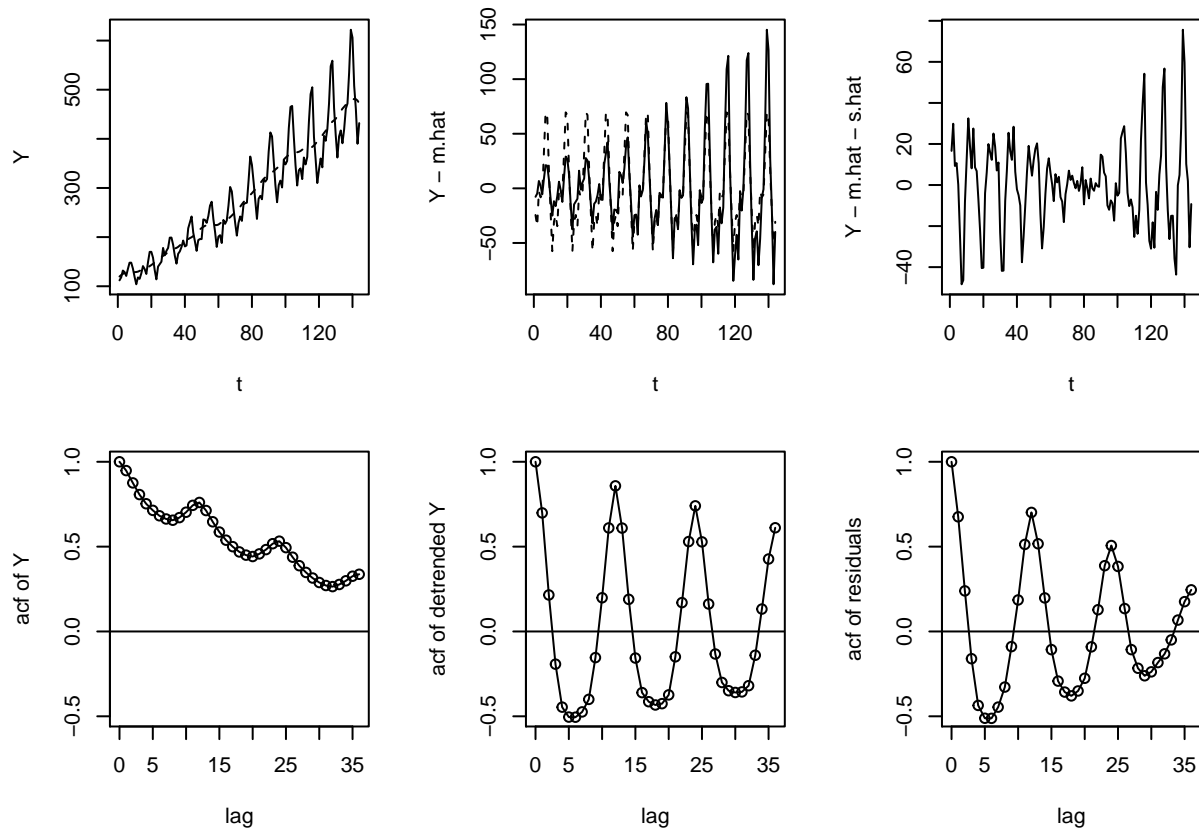
plot(Y - m.hat - s.hat ~ t,type="l")

plot(my.acf(Y,max.lag=36)$rho.hat~c(0:36),type="o",ylab="acf of Y",xlab="lag",
ylim=c(-.5,1))
abline(h=0)

plot(my.acf(Y - m.hat,max.lag=36)$rho.hat~c(0:36),type="o",
ylab="acf of detrended Y",xlab="lag",ylim=c(-.5,1))
abline(h=0)

plot(my.acf(Y - m.hat - s.hat,max.lag=36)$rho.hat~c(0:36),type="o",
```

```
ylab="acf of residuals",xlab="lag",ylim=c(-.5,1))
abline(h=0)
```



The residuals do not appear to be stationary. The seasonality appears to change over time; the fluctuations increase in amplitude.

Handling a time-evolving seasonal effect:

Using the notation

$$(s_1, \dots, s_n) = (s_{1,1}, \dots, s_{1,d}, \dots, s_{n/d,1}, \dots, s_{n/d,d}),$$

suppose that $s_{j,1}, \dots, s_{j,d}$ change as j changes, and that $\sum_{k=1}^d s_{j,k} = 0$ for all j . Then we might estimate the seasonal effects with a moving average across neighboring seasons as is done with the following R code:

```
data(AirPassengers)

Y <- as.numeric(AirPassengers)
n <- length(Y)
t <- 1:n

d <- 12
q <- d/2

Ymod <- c(rep(Y[1],q),Y,rep(Y[n],q))
m.hat <- numeric(n)
for(i in 1:n)
{
  m.hat[i] <- (.5*Ymod[i] + sum(Ymod[(i+1):(i+2*q-1)]) + .5*Ymod[(i+2*q)])/(2*q)
```



```

}

# now do a moving average over neighboring seasons:
S <- Y - m.hat
ns <- n/d
qs <- 2
Smod <- matrix(c( rep(S[1:d],qs), S , rep(S[(n-d+1):n],qs)),ns+2*qs,byrow=TRUE)
S.hat <- numeric(n)
for(j in 1:ns)
{
  ind <- ((j-1)*d+1):(j*d)
  w.j <- apply(Smod[j:(j+2*qs),],2,mean)
  S.hat[ind] <- w.j - mean(w.j)
}

par(mfrow=c(2,3),mar=c( 5.1, 4.1,1.1, 2.1))
plot(Y~t,type="l")
lines(m.hat~t,lty=2)

plot(Y - m.hat ~ t,type="l")
lines(S.hat~t,lty=2)

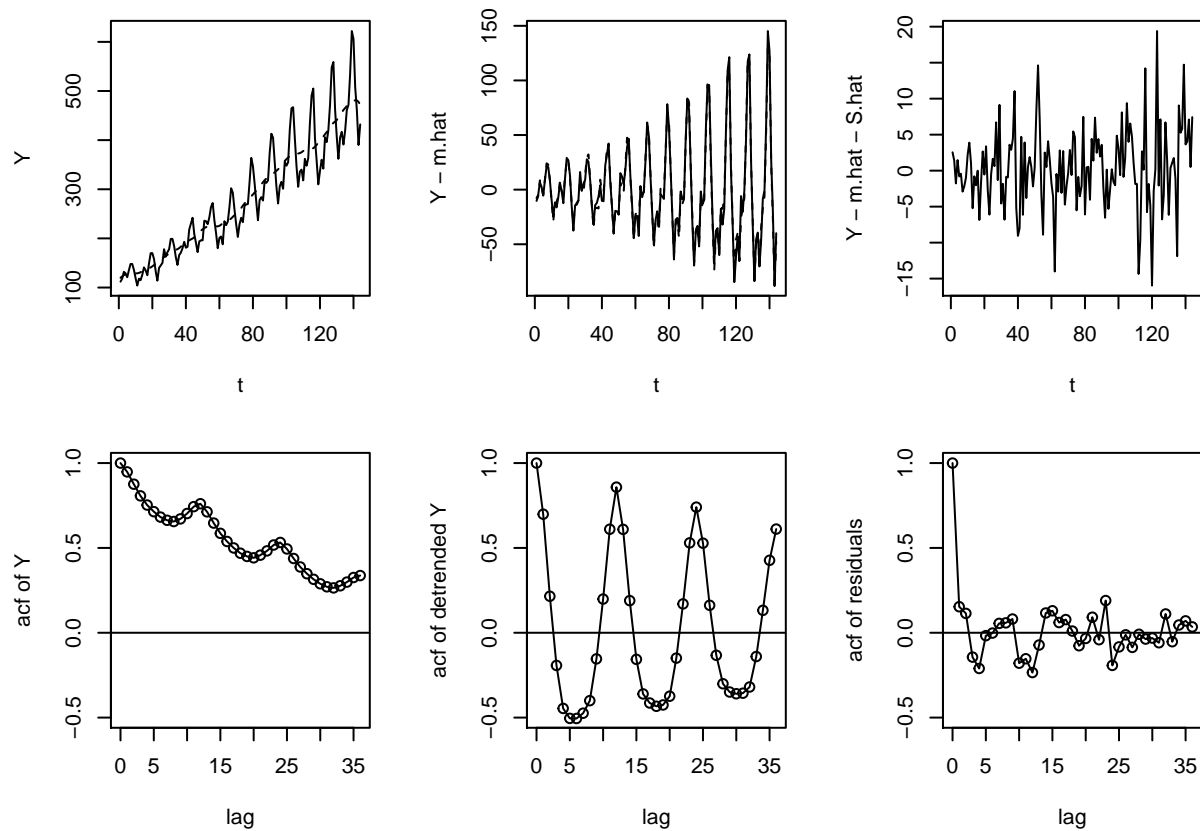
plot(Y - m.hat - S.hat ~ t,type="l")

plot(my.acf(Y,max.lag=36)$rho.hat~c(0:36),type="o",ylab="acf of Y",xlab="lag",
ylim=c(-.5,1))
abline(h=0)

plot(my.acf(Y - m.hat,max.lag=36)$rho.hat~c(0:36),type="o",
      ylab="acf of detrended Y",xlab="lag",ylim=c(-.5,1))
abline(h=0)

plot(my.acf(Y - m.hat - S.hat,max.lag=36)$rho.hat~c(0:36),type="o",
      ylab="acf of residuals", xlab="lag",ylim=c(-.5,1))
abline(h=0)

```



This results in residuals which look much more stationary.

Detrending and deseasonalizing with differencing

If the period of the seasonal component of a time series is equal to d , then taking $Y_t - Y_{t-d}$ may eliminate the seasonality. Define the lag- d difference operator ∇_d by

$$\nabla_d Y_t = Y_t - Y_{t-d} = (1 - B^d)Y_t.$$

Note that this is different from $\nabla^d Y_t = (1 - B)^d Y_t$. If a trend remains after lag- d differencing, one might take the first difference of the de-seasonalized series, as in the following example:

```
library(MASS)
data(deaths)

par(mfrow=c(2,3),mar=c( 5.1, 4.1,1.1, 2.1))

plot(deaths)
plot(diff(deaths,12))
plot(diff(diff(deaths,12),1))

plot(my.acf(deaths,max.lag=14)$rho.hat~c(0:14),type="o",ylab="acf of Y",
      xlab="lag", ylim=c(-.8,1))
abline(h=0)

plot(my.acf(diff(deaths,12),max.lag=14)$rho.hat~c(0:14),type="o",
      ylab="acf of diff(deaths,12)",xlab="lag",ylim=c(-.8,1))
abline(h=0)
```

```
plot(my.acf(diff(diff(deaths,12),1),max.lag=14)$rho.hat~c(0:14),type="o",
      ylab="acf of diff(diff(deaths,12),1)",xlab="lag",ylim=c(-.8,1))
abline(h=0)
```

