Lecture 22: Time Series Regression

Ailin Zhang

2023-07-11

Agenda

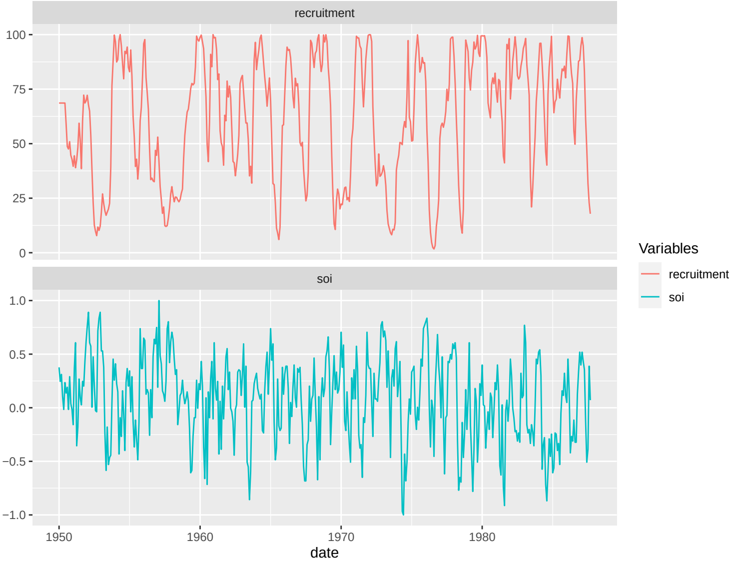
- AR Models
- MA and ARMA Models
- ARIMA Models
- Seasonal ARIMA
- Others (Gaussian Process models, spatio-temporal models)

Southern Oscillation Index & Recruitment

The Southern Oscillation Index (SOI) is an indicator of the development and intensity of El Niño (negative SOI) or La Niña (positive SOI) events in the Pacific Ocean. These data also included the estimate of "recruitment", which indicate fish population sizes in the southern hemisphere.

```
## # A tibble: 453 x 3
##     date    soi recruitment
##    <dbl>  <dbl>       <dbl>
##  1 1950   0.377        68.6
##  2 1950.  0.246        68.6
##  3 1950.  0.311        68.6
##  4 1950.  0.104        68.6
##  5 1950. -0.016        68.6
##  6 1950.  0.235        68.6
##  7 1950.  0.137        59.2
##  8 1951.  0.191        48.7
```

Time series


recruitment
soi
100
75
50
25
0
1.0
0.5
0.0
−0.5
−1.0
1950
1960
1970
1980
date
Variables
recruitment
soi

AR models: Auto Regressive Process

$y_t = y_t - \mu_t$ is the detrended signal

$$AR(1): \quad y_t = \phi\, y_{t-1} + w_t$$

where $w_t \sim N(0, \sigma^2)$

Sample path:

```
{r out.width='80%'} par(mfrow=c(1,2))
plot(arima.sim(list(order=c(1,0,0), ar=.9), n=100),
ylab="x",     main=(expression(AR(1)~~~phi==+.9)))
plot(arima.sim(list(order=c(1,0,0), ar=-.9), n=100),
ylab="x",     main=(expression(AR(1)~~~phi==-.9)))
```

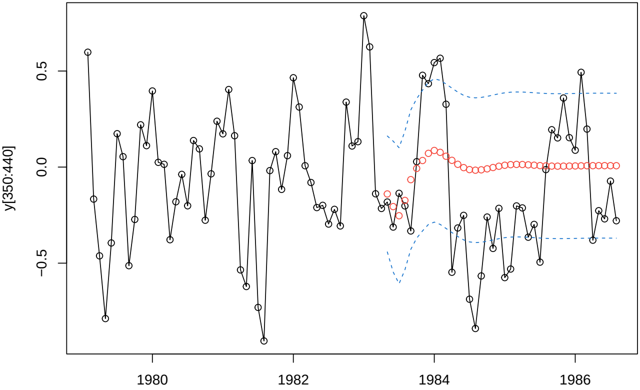
AR models

We can easily generalize from an AR(1) to an AR(p) model by simply adding additional autoregressive terms to the model.

$$
\begin{aligned}
AR(p): \quad y_t &= \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + \cdots + \phi_p\, y_{t-p} + w_t \\
&= w_t + \sum_{i=1}^{p} \phi_i\, y_{t-i}
\end{aligned}
$$

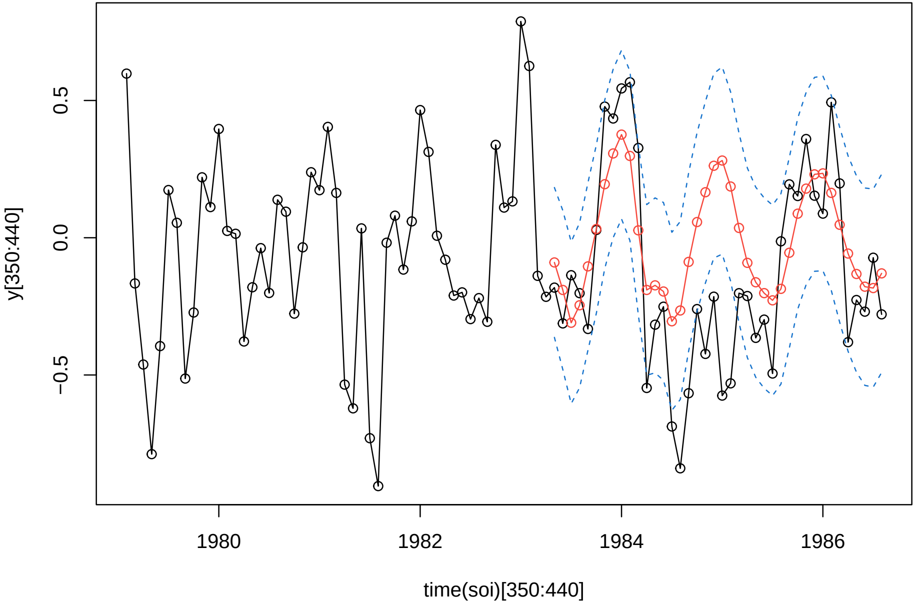
Fit AR(p) models (p given)

```
fit = lm(soi~time(soi), na.action=NULL) # regress soi on time
y = fit$res
ar8<-ar(y[1:400], aic = TRUE, order=8, demean = TRUE, intercep
fore = predict(ar8, n.ahead=40)
```



Fit AR(p) models with p unknown

```
ar_model <-ar(y[1:400], aic = TRUE, order.max=100, demean = TR
ar_model
##
## Call:
## ar(x = y[1:400], aic = TRUE, order.max = 100, demean = TRUE
##
## Coefficients:
##       1        2        3        4        5        6
##  0.3551   0.0733   0.1154   0.0699  -0.0278  -0.0586  -0.03
##       9       10       11       12       13       14
## -0.0267   0.0810   0.1649   0.1681   0.0244  -0.1747  -0.12
##      17       18
## -0.0464  -0.0798
##
## Order selected 18  sigma^2 estimated as  0.0745
fore = predict(ar_model, n.ahead=40)
```

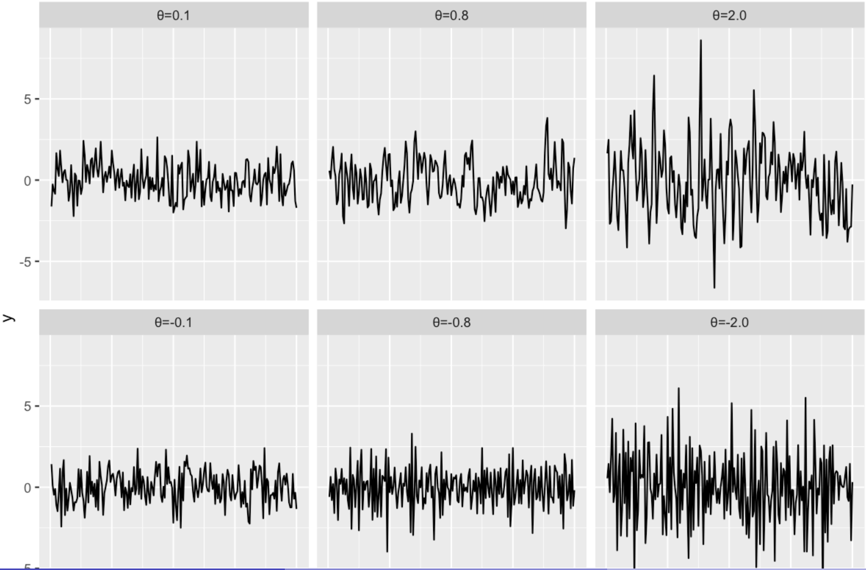

0.5
0.0
−0.5
y[350:440]
1980
1982
1984
1986
time(soi)[350:440]

MA models: Moving Average (MA) Processes

A moving average process is similar to an AR process, except that the autoregression is on the error term.

$$MA(1): \qquad y_t = w_t + \theta\, w_{t-1}$$

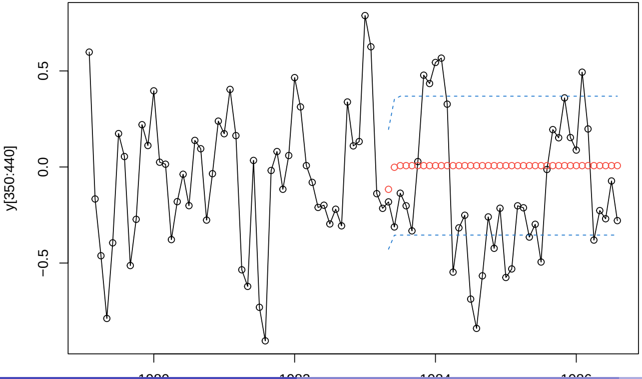
Time series



MA(q) models

$$MA(q): \qquad y_t = \delta + w_t + \theta_1\, w_{t-1} + \theta_2\, w_{t-2} + \cdots + \theta_q\, w_{t-q}$$

```r
ma1<-arima(y[1:400], order = c(0, 0, 2))
fore = predict(ma1, n.ahead=40)
```



ARMA: Autoregressive moving average model

An ARMA model is a composite of AR and MA processes,

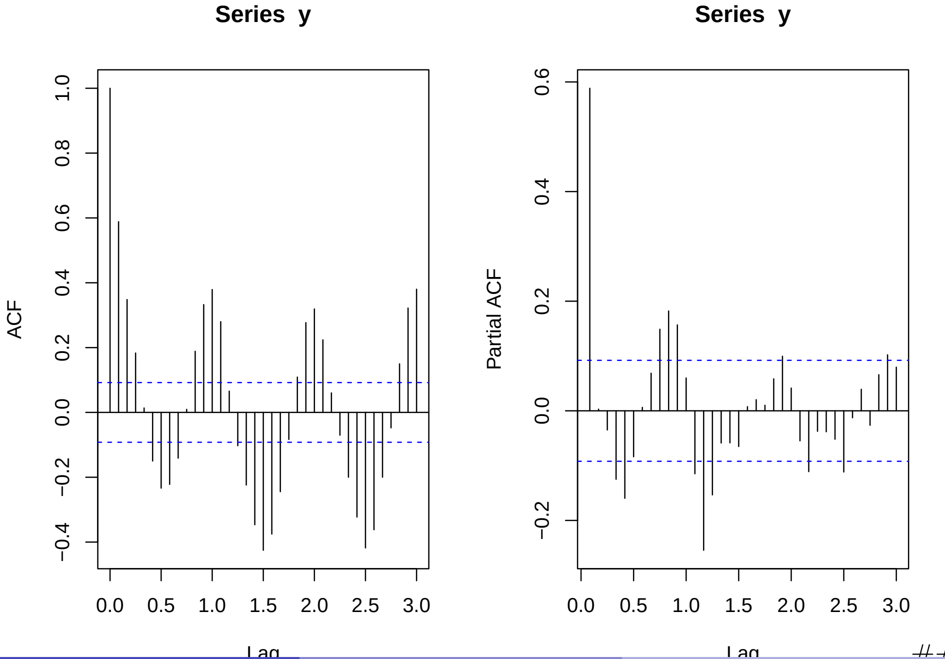
$$ARMA(p,q):$$

$$y_t = \delta + \phi_1\, y_{t-1} + \cdots \phi_p\, y_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t_q}$$

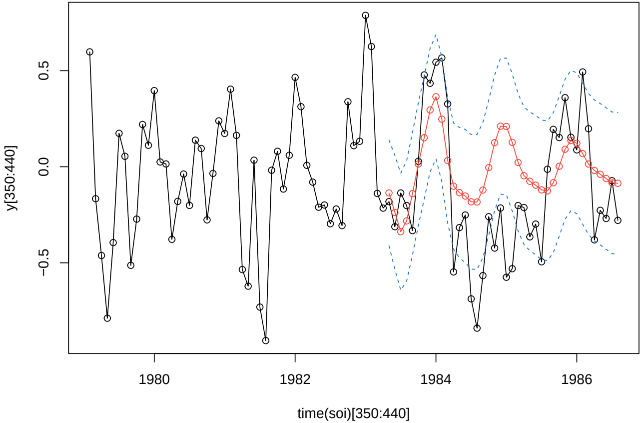
$$\phi_p(L) y_t = \delta + \theta_q(L) w_t$$

ACF and PACF

- ACF, Autocorrelation Function shows the direct as well as indirect relation between a value with a lagged version of itself (we can count the number of significant lags on the ACF graph to find q for MA process)
- PACF, Partial Autocorrelation Function only shows the direct relationship between a value and a lagged version of itself. (We can use the PACF to count the number of significant lags for an AR process)



Series y
ACF
Lag
Series y
Partial ACF
Lag



0.5
0.0
−0.5
y[350:440]
1980
1982
1984
1986
time(soi)[350:440]

ARIMA: Differencing opearator

We will need to define one more notational tool for indicating differencing

$$\Delta y_t = y_t - y_{t-1}$$

just like the lag operator we will indicate repeated applications of this operator using exponents

$$\begin{aligned}
\Delta^2 y_t &= \Delta(\Delta y_t) \\
&= (\Delta y_t) - (\Delta y_{t-1}) \\
&= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\
&= y_t - 2y_{t-1} + y_{t-2}
\end{aligned}$$

Δ can also be expressed in terms of the lag operator L,

$$\Delta^d = (1 - L)^d$$

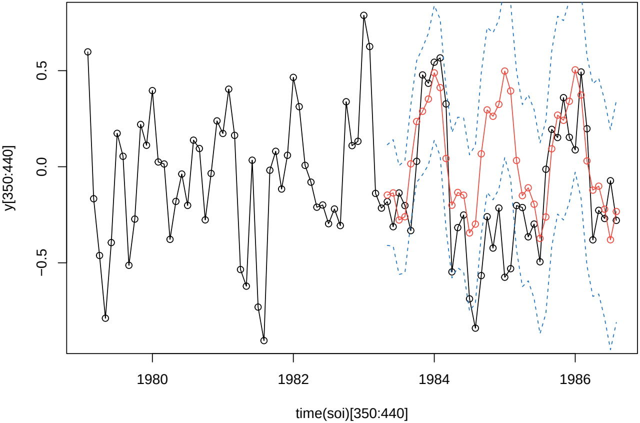
ARIMA Models

Autoregressive integrated moving average are just an extension of an $ARMA$ model to include differencing of degree d to y_t before include the autoregressive and moving average components.

$$ARIMA(p,d,q): \qquad \phi_p(L)\,\Delta^d\,y_t = \delta + \theta_q(L)w_t$$

ARIMA Models

```
arima_mod <- arima(y[1:400], order = c(14,1,4), include.mean = FALSE)
arima_mod
##
## Call:
## arima(x = y[1:400], order = c(14, 1, 4), include.mean = FALSE)
##
## Coefficients:
##          ar1      ar2     ar3      ar4      ar5      ar6      ar7      ar8
##       1.0203  -1.2574  0.9039  -0.4892  -0.1828  -0.1631  -0.1464  -0.1455
## s.e.  0.0759   0.0911  0.0989   0.1063   0.1083   0.1088   0.1085   0.1084
##           ar9     ar10    ar11    ar12     ar13     ar14      ma1     ma2
##       -0.0637  -0.0389  0.0354  0.0233  -0.0115  -0.0941  -1.6803  1.8146
## s.e.   0.1088   0.1084  0.1060  0.0975   0.0773   0.0565   0.0594  0.0958
##           ma3     ma4
##       -1.5648  0.9336
## s.e.   0.0607  0.0310
##
## sigma^2 estimated as 0.06834:  log likelihood = -36.63,  aic = 111.26
```

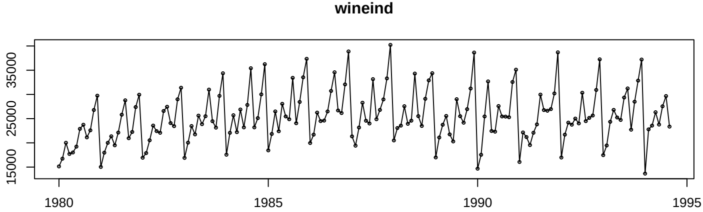

0.5
0.0
−0.5
y[350:440]
1980
1982
1984
1986
time(soi)[350:440]

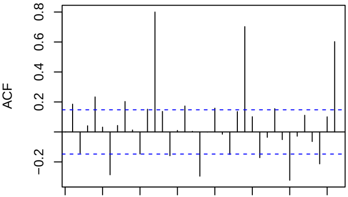
ARIMA Model

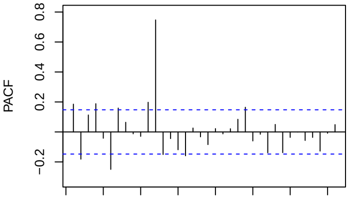
```r
arima(y[1:400], order = c(14,1,4), include.mean = FALSE)$aic
## [1] 111.2576
arima(y[1:400], order = c(13,1,4), include.mean = FALSE)$aic
## [1] 113.2476
arima(y[1:400], order = c(13,1,5), include.mean = FALSE)$aic
## [1] 116.4798
arima(y[1:400], order = c(13,0,3), include.mean = FALSE)$aic
## [1] 93.27956
```

Seasonal ARIMA for `wineind`

Australian total wine sales by wine makers in bottles <= 1 litre. Jan 1980 – Aug 1994.







Seasonal ARIMA

We can extend the existing ARIMA model to handle these higher order lags (without having to include all of the intervening lags).

Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$:

$$\Phi_P(L^s)\,\phi_p(L)\,\Delta_s^D\,\Delta^d\,y_t = \delta + \Theta_Q(L^s)\,\theta_q(L)\,w_t$$

Seasonal ARIMA

We can extend the existing ARIMA model to handle these higher order lags (without having to include all of the intervening lags).

Seasonal ARIMA $(p, d, q) \times (P, D, Q)_s$:

$$\Phi_P(L^s)\,\phi_p(L)\,\Delta_s^D\,\Delta^d\,y_t = \delta + \Theta_Q(L^s)\,\theta_q(L)\,w_t$$

where

$$\begin{aligned} \phi_p(L) &= 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p \\ \theta_q(L) &= 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^q \\ \Delta^d &= (1 - L)^d \end{aligned}$$

$$\begin{aligned} \Phi_P(L^s) &= 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \ldots - \Phi_P L^{Ps} \\ \Theta_Q(L^s) &= 1 + \Theta_1 L + \Theta_2 L^{2s} + \ldots + \theta_p L^{Qs} \\ \Delta_s^D &= (1 - L^s)^D \end{aligned}$$

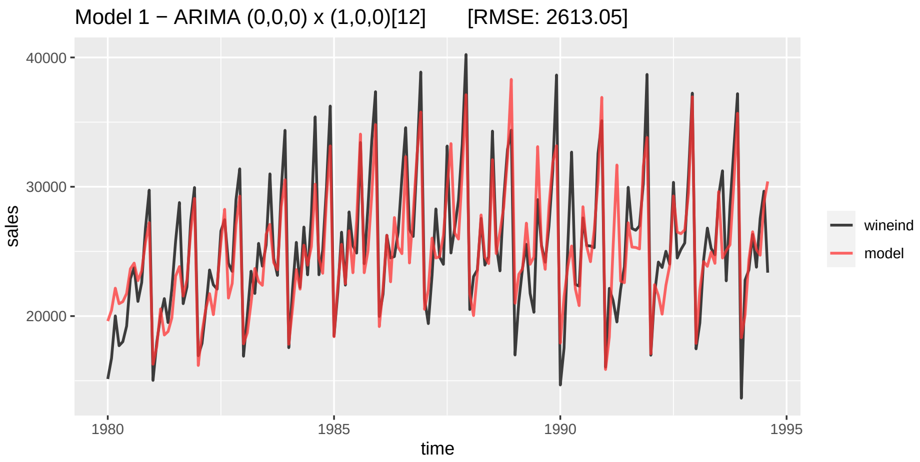
Seasonal ARIMA for `wineind`

Lets consider an ARIMA$(0,0,0) \times (1,0,0)_{12}$:

$$(1 - \Phi_1 L^{12})\, y_t = \delta + w_t$$
$$y_t = \Phi_1 y_{t-12} + \delta + w_t$$

```
m1 = Arima(wineind, seasonal=list(order=c(1,0,0), period=12))
m1
## Series: wineind
## ARIMA(0,0,0)(1,0,0)[12] with non-zero mean
##
## Coefficients:
##          sar1      mean
##        0.8780  24489.24
## s.e.   0.0314   1154.48
##
## sigma^2 = 6906536:  log likelihood = -1643.39
## AIC=3292.78   AICc=3292.92   BIC=3302.29
```

Fitted model



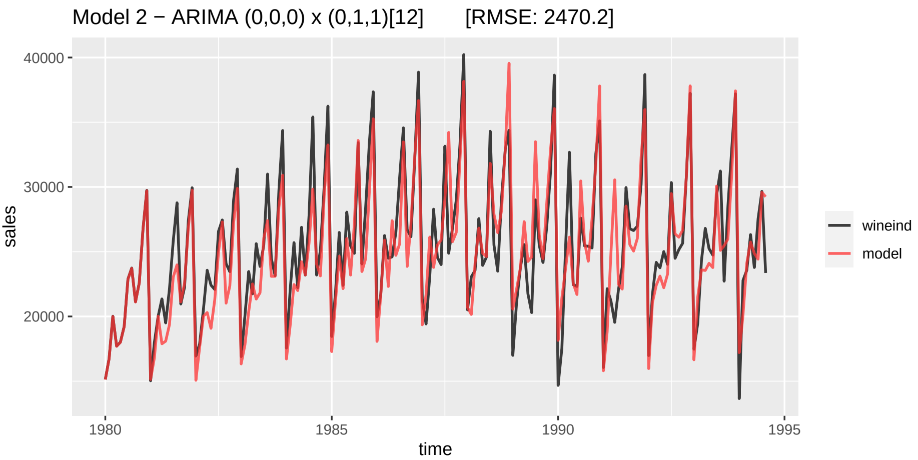
Model 2

ARIMA$(0,0,0) \times (0,1,1)_{12}$:

$$
\begin{aligned}
(1 - L^{12}) y_t &= \delta + (1 + \Theta_1 L^{12}) w_t \\
y_t - y_{t-12} &= \delta + w_t + \Theta_1 w_{t-12} \\
y_t &= \delta + y_{t-12} + w_t + \Theta_1 w_{t-12}
\end{aligned}
$$

```
(m2 = Arima(wineind, order=c(0,0,0), seasonal=list(order=c(0,1
## Series: wineind
## ARIMA(0,0,0)(0,1,1)[12]
##
## Coefficients:
##          sma1
##       -0.3246
## s.e.   0.0807
##
## sigma^2 = 6588531:  log likelihood = -1520.34
```

Fitted model



Model 3

ARIMA$(3,0,0) \times (0,1,1)_{12}$

$$(1-\phi_1 L-\phi_2 L^2-\phi_3 L^3)\,(1-L^{12})y_t = \delta+(1+\Theta_1 L)w_t$$
$$(1-\phi_1 L-\phi_2 L^2-\phi_3 L^3)\,(y_t-y_{t-12}) = \delta+w_t+w_{t-12}$$
$$y_t = \delta+\sum_{i=1}^{3}\phi_i y_{t-1}+y_{t-12}-\sum_{i=1}^{3}\phi_i y_{t-12-i}+w_t+w_{t-12}$$

```
(m3 = Arima(wineind, order=c(3,0,0),
            seasonal=list(order=c(0,1,1), period=12)))
## Series: wineind
## ARIMA(3,0,0)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2     ar3     sma1
##       0.1402  0.0806  0.3040  -0.5790
## s.e.  0.0755  0.0813  0.0823   0.1023
```

Fitted model