Seasonal ARIMA models

FISH 550 – Applied Time Series Analysis   Download Rmd pdf

Eli Holmes

6 Apr 2023

Seasonality

Load the chinook salmon data set

```
load("chinook.RData")
head(chinook)
```

```
##   Year Month Species log.metric.tons metric.tons
## 1 1990   Jan Chinook        3.397858        29.9
## 2 1990   Feb Chinook        3.808882        45.1
## 3 1990   Mar Chinook        3.511545        33.5
## 4 1990   Apr Chinook        4.248495        70.0
## 5 1990   May Chinook        5.200705       181.4
## 6 1990   Jun Chinook        4.371976        79.2
```

The data are monthly and start in January 1990. To make this into a ts object do

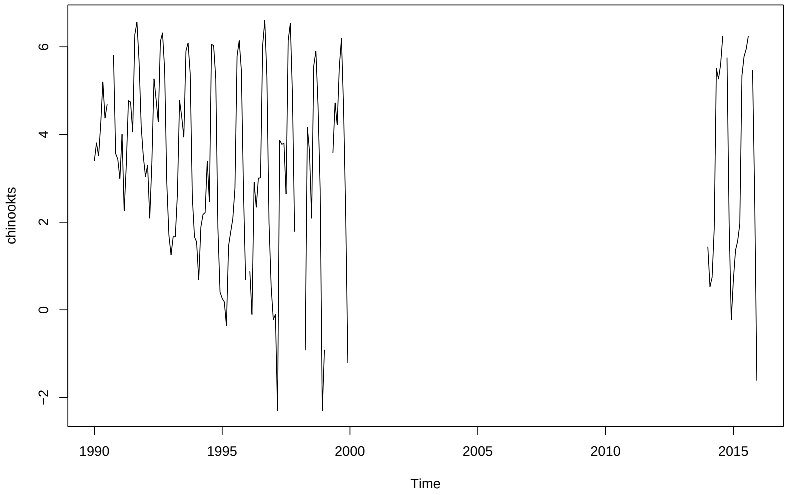
```
chinookts <- ts(chinook$log.metric.tons, start=c(1990,1),
                 frequency=12)
```

`start` is the year and month and frequency is the number of months in the year.

Use `?ts` to see more examples of how to set up ts objects.

Plot seasonal data

```
plot(chinookts)
```



Seasonal ARIMA model SARIMA

Seasonally differenced data, e.g. chinook data January 1990 - chinook data January 1989.

$$z_t = x_t - x_{t+s} - m$$

Basic structure of a seasonal AR model

seasonal differenced data = AR(seasonal) + AR(non-seasonal) + AR(p+season)

For example, SARIMA model could capture this process

1. AR(seasonal) January in t is correlated with January in $t-1$
2. AR(non-seasonal) January differences are correlated with February differences
3. AR(p+season) appears because of 1 and 2

$z_t = \text{AR}(1) + \text{AR}(12) + \text{AR}(1{+}12)$

$$z_t = \phi_1 z_{t-1} + \Phi_1 z_{t-12} - \phi_1 \Phi_1 z_{t-13}$$

Notation

ARIMA (p,d,q)(ps,ds,qs)S

ARIMA (non-seasonal part)(seasonal part)Frequency

ARIMA (non-seasonal) means t correlated with $t-1$

ARIMA (seasonal) means t correlated with $t-s$

Examples

ARIMA (1,0,0)(1,0,0)[12]

What data are we modeling? Get the differences

$$z_t = x_t - m$$

Write out the AR parts with z_t

$$z_t = \phi_1 z_{t-1} + \Phi_1 z_{t-12} - \phi_1 \Phi_1 z_{t-13} + w_t$$

Write out the MA parts, the w_t. No MA in this model.

$$w_t = e_t$$

ARIMA (1,0,0)(1,1,0)[12]

Figure out z_t. Just a seasonal difference.

$$z_t = x_t - x_{t-12} - m$$

Write out the AR parts with z_t

$$z_t = \phi_1 z_{t-1} + \Phi_1 z_{t-12} - \phi_1 \Phi_1 z_{t-13} + w_t$$

Write out the MA parts, the w_t. No MA in this model. w_t is white noise.

$$w_t = e_t$$

Seasonal random walk model

ARIMA(0,0,0)(0,1,0)[12]

expected January 1990 = January 1989 + constant mean

Figure out z_t. m is the mean seasonal difference.

$$z_t = x_t - x_{t-12} - m$$

Write out the AR parts with z_t. No AR part.

$$z_t = w_t$$

Write out the MA parts, the w_t.

$$w_t = e_t$$

Seasonal random walk model with random trend

ARIMA(0,1,0)(0,1,0)[12]

expected Feb 1990 = Feb 1989 + (Jan 1990 - Jan 1989)

Figure out z_t. m is the mean seasonal difference.

$$z_t = (x_t - x_{t-12}) - (x_{t-1} - x_{t-13}) - m$$

Write out the AR parts with z_t. No AR part.

$$z_t = w_t$$

Write out the MA parts, the w_t.

$$w_t = e_t$$

airline model

ARIMA(0, 1, 1)(0, 1, 1)[12]

Figure out z_t.

$$z_t = (x_t - x_{t-12}) - (x_{t-1} - x_{t-13}) - m$$

Write out the AR parts with z_t. No AR part.

$$z_t = w_t$$

Write out the MA parts, the w_t.

$$w_t = e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13}$$

Example with longer lags

ARIMA (2,0,1)(1,0,2)[12]

What data are we modeling? Get the differences

$$z_t = x_t - m$$

Write out the AR parts with z_t

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \Phi_1 z_{t-12} - (\text{cross products}) + w_t$$

Write out the MA parts, the w_t.

$$w_t = e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-12} - \Theta_2 e_{t-24} + (\text{cross products})$$

auto.arima() for seasonal ts

auto.arima() will recognize that our data has season and fit a seasonal ARIMA model to our data by default. We will define the training data up to 1998 and use 1999 as the test data.

```
traindat <- window(chinookts, c(1990,10), c(1998,12))
testdat <- window(chinookts, c(1999,1), c(1999,12))
fit <- forecast::auto.arima(traindat)
fit
```

```
## Series: traindat
## ARIMA(1,0,0)(0,1,0)[12] with drift
##
## Coefficients:
##          ar1    drift
##       0.3676  -0.0320
## s.e.  0.1335   0.0127
##
## sigma^2 = 0.8053:  log likelihood = -107.37
## AIC=220.73   AICc=221.02   BIC=228.13
```

Summary for seasonal models

Basic steps for identifying a seasonal model. **forecast** automates most of this.

- Check that you have specified your season correctly in your ts object.
- Plot your data. Look for trend, seasonality and random walks.

Summary

- Use differencing to remove season and trend.
  - Season and no trend. Take a difference of lag = season
  - No seasonality but a trend. Try a first difference
  - Both. Do both types of differences
  - Neither. No differencing
  - Random walk. First difference
  - Parametric looking curve. Transform

Summary

- Examine the ACF and PACF of the differenced data.
  - Look for patterns (spikes) at seasonal lags
- Estimate likely models and compare with model selection criteria (or cross-validation). Use `TRACE=TRUE`
- Do residual checks

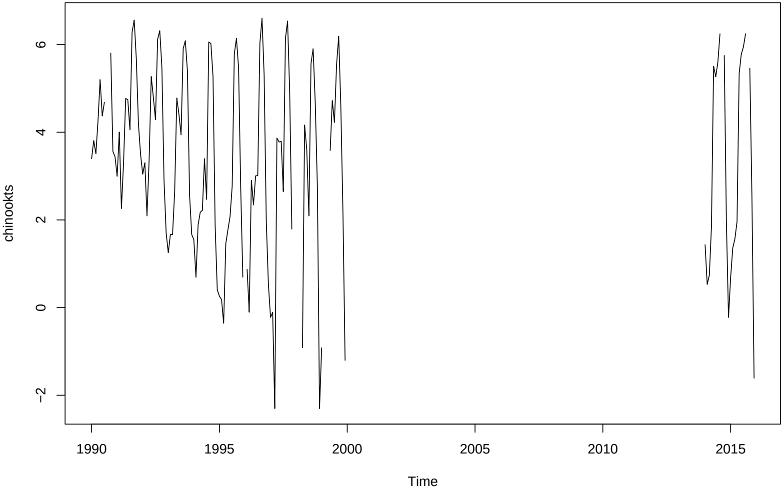
Forecasting with a Seasonal model

Load the chinook salmon data

```
load("chinook.RData")
chinookts <- ts(chinook$log.metric.tons, start=c(1990,1),
                frequency=12)
```

Plot seasonal data

```
plot(chinookts)
```



Seasonal ARIMA model

Seasonally differenced data, e.g. chinook data January 1990 - chinook data January 1989.

$$z_t = x_t - x_{t+s} - m$$

Basic structure of a seasonal AR model

z_t = AR(p) + AR(season) + AR(p+season)

e.g. z_t = AR(1) + AR(12) + AR(1+12)

Example AR(1) non-seasonal part + AR(1) seasonal part

$$z_t = \phi_1 z_{t-1} + \Phi_1 z_{t-12} - \phi_1 \Phi_1 z_{t-13}$$

Notation

ARIMA (p,d,q)(ps,ds,qs)S

ARIMA (1,0,0)(1,1,0)[12]

auto.arima() for seasonal ts

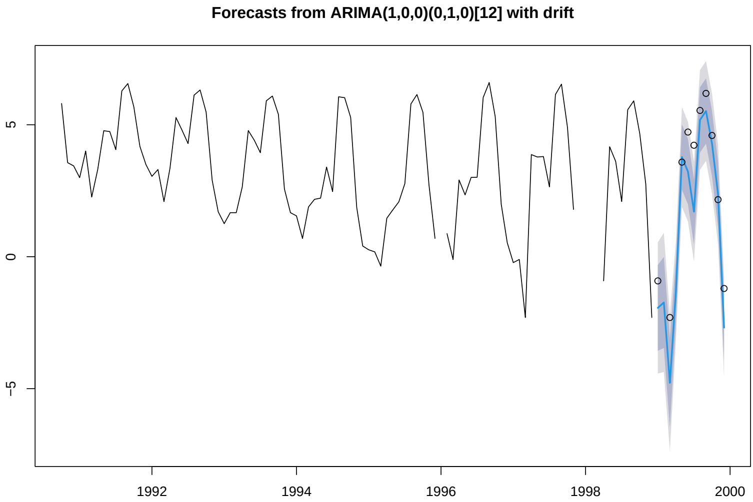
auto.arima() will recognize that our data has season and fit a seasonal ARIMA model to our data by default. We will define the training data up to 1998 and use 1999 as the test data.

```
traindat <- window(chinookts, c(1990,10), c(1998,12))
testdat <- window(chinookts, c(1999,1), c(1999,12))
fit <- forecast::auto.arima(traindat)
fit
```

```
## Series: traindat
## ARIMA(1,0,0)(0,1,0)[12] with drift
##
## Coefficients:
##          ar1    drift
##       0.3676  -0.0320
## s.e.  0.1335   0.0127
##
## sigma^2 = 0.8053:  log likelihood = -107.37
## AIC=220.73   AICc=221.02   BIC=228.13
```

Forecast using seasonal model

```
fr <- forecast::forecast(fit, h=12)
plot(fr)
points(testdat)
```



Missing values

Missing values are ok when fitting a seasonal ARIMA model