Seasonal models

Seasonal time series exhibits cyclical or periodic behavior.

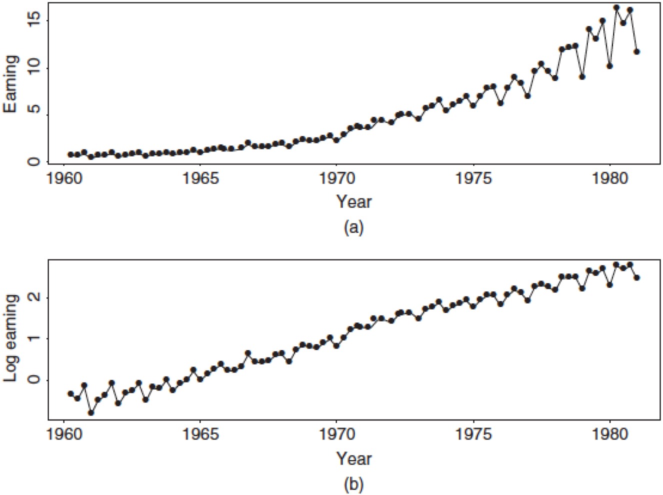


Figure 2.13 of Tsay (2010): Time plots of quarterly earnings per share of Johnson & Johnson from 1960 to 1980: (a) observed earnings and (b) log earnings.

Analysis of seasonal time series has a long history.

In some applications, seasonality is of secondary importance and is removed from the data, resulting in a seasonally adjusted time series that is then used to make inference.

The procedure to remove seasonality from a time series is referred to as seasonal adjustment.

Most economic data published by the U.S. government are seasonally adjusted (e.g., the growth rate of gross domestic product and the unemployment rate).

In other applications such as forecasting, seasonality is as important as other characteristics of the data and must be handled accordingly.

Seasonal Differencing

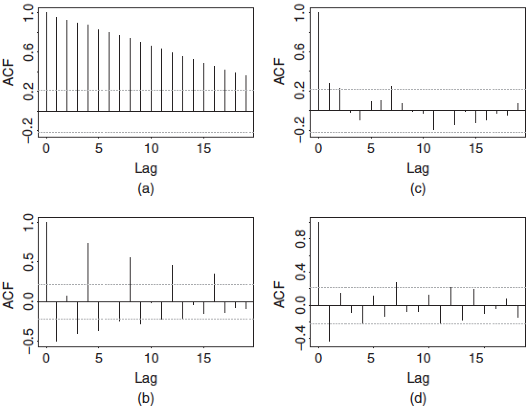
Denote the log earnings by x_t.

(a) ACF of x_t: strong serial correlations.

(b) ACF of Δx_t: strong periodicity.

(c) ACF of $\Delta_4 x_t = (1 - B^4) x_t$

(d) ACF of $\Delta_4(\Delta x_t) = x_t - x_{t-1} - x_{t-4} + x_{t-5}$



Multiplicative Seasonal Models

The seasonal time series model

$$(1 - B^s)(1 - B)x_t = (1 - \theta B)(1 - \Theta B^s)a_t$$

where s is the periodicity of the series, a_t is a white noise series, $|\theta| < 1$, and $|\Theta| < 1$, is referred to as the *airline model* in the literature (Box-Jenkins-Reinsel, 1994, Chapter 9).

It is easy to see that

$$w_t \equiv (1 - B^s)(1 - B)x_t = a_t - \theta a_{t-1} - \Theta a_{t-s} + \theta\Theta a_{t-s-1}$$

and that

$$E(w_t) = 0 \quad \text{and} \quad V(w_t) = (1 + \theta^2)(1 + \Theta^2)\sigma_a^2.$$

ACF

The ACF of the w_t series is given by

$$\begin{aligned}\rho_1 &= \frac{-\theta}{1+\theta^2} \\ \rho_s &= \frac{-\Theta}{1+\Theta^2} \\ \rho_{s-1} &= \rho_{s+1} = \rho_1\rho_s = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}\end{aligned}$$

and $\rho_l = 0$ for $l > 0$ and $l \neq 1, s-1, s, s+1$.

In practice, a multiplicative seasonal model says that the dynamics of the regular and seasonal components of the series are approximately orthogonal.

The model

$$w_t = (1 - \theta B - \Theta B^s)a_t$$

where $|\theta| < 1$ and $|\Theta| < 1$, is a nonmultiplicative seasonal MA model.

A multiplicative model is more parsimonious than the corresponding nonmultiplicative model because both models use the same number of parameters, but the multiplicative model has more nonzero ACFs.

Seasonal ARIMA Model

The airline model

$$(1 - B^s)(1 - B)x_t = (1 - \theta B)(1 - \Theta B^s)a_t$$

is a special case of the general class of

$$\text{ARIMA}(p, d, q) \times (P, D, Q)_s \text{ models.}$$

The seasonal autoregressive integrated moving average model of Box and Jenkins (1970) is given by

$$\Phi_P(B^s)\phi(B)\Delta_s^D \Delta^d x_t = \alpha + \Theta_Q(B^s)\theta(B)a_t$$

and is denoted as an ARIMA$(p, d, q) \times (P, D, Q)_s$.

The airline model is a ARIMA$(0, 1, 1) \times (0, 1, 1)_s$ model.

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

The model

$$(1-\phi L)(1-\Phi L^{12})\Delta\Delta_{12} y_t = (1-\theta L)(1-\Theta L^{12})\epsilon_t$$

can be decomposed into

$$\begin{aligned}
\Delta\Delta_{12} y_t &= y_t - y_{t-1} - y_{t-12} + y_{t-13} \\
(1-\phi L)(1-\Phi L^{12}) &= (1-\phi L - \Phi L^{12} + \phi\Phi L^{13}) \\
(1-\theta L)(1-\Theta L^{12}) &= (1-\theta L - \Theta L^{12} + \theta\Theta L^{13})
\end{aligned}$$

Therefore,

$$\begin{aligned}
y_t &= (1+\phi) y_{t-1} - \phi y_{t-2} + (1-\phi+\Phi) y_{t-12} \\
&- (1+\phi\Phi+\Phi) y_{t-13} + (1+\Phi)\phi y_{t-14} \\
&- \Phi y_{t-24} + (1+\phi)\Phi y_{t-25} - \phi\Phi y_{t-26} \\
&+ \epsilon_t - \theta\epsilon_{t-1} - \Theta\epsilon_{t-12} + \theta\Theta\epsilon_{t-13}
\end{aligned}$$

is an ARMA$(26,13)$ with severe parameter constraints.

In this example we apply the airline model to the log series of quarterly earnings per share of Johnson & Johnson from 1960 to 1980.

Based on the exact-likelihood method, the fitted model is

$$(1 - B)(1 - B^4)x_t = (1 - 0.678B)(1 - 0.314B^4)a_t,$$

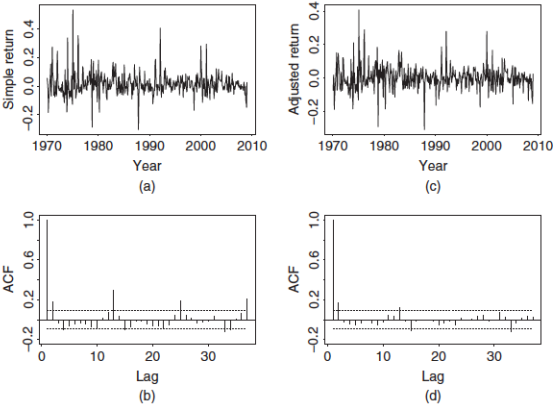
with $\hat{\sigma}_a = 0.089$.

The Ljung-Box statistics of the residuals show $Q(12) = 10.0$ with a p-value of 0.44. The model appears to be adequate.

Another example

Consider the monthly simple returns of the CRSP Decile 1 Index from Jan/1970 to Dec/2008 for 468 observations.

No clear pattern of seasonality.
ACF contains significant lags at 1,12, 24, and 36.



If seasonal ARMA models are entertained, a model in the form

$$(1 - \phi_1 B)(1 - \phi_{12} B^{12}) R_t = (1 - \theta_{12} B^{12}) a_t$$

is identified, where R_t denotes the monthly simple return.

Using the conditional-likelihood method, the fitted model is

$$(1 - 0.18B)(1 - 0.87B^{12}) R_t = (1 - 0.74B^{12}) a_t,$$

with $\tilde{\sigma}_a = 0.069$.

Using the exact-likelihood method, the fitted model is

$$(1 - 0.188B)(1 - 0.951B^{12}) R_t = (1 - 0.997B^{12}) a_t,$$

with $\tilde{\sigma}_a = 0.063$.

The cancellation between seasonal AR and MA factors is clearly seen.

This highlights the usefulness of using the exact-likelihood method and, the estimation result suggests that the seasonal behavior might be deterministic.

January effect: To further confirm this assertion, we define the dummy variable for January, that is,

$$\text{Jan}_t = 1 \quad \text{if } t \text{ is January,}$$

and $\text{Jan}_t = 0$ otherwise, and employ the simple linear regression $R_t = \beta_0 + \beta_1 \text{Jan}_t + e_t$.

The fitted model is $R_t = 0.0029 + 0.1253\text{Jan}_t + e_t$, where the standard errors of the estimates are 0.0033 and 0.0115, respectively.

We will use Rob J Hyndman's R package `forecast`

```
Description: forecast is a generic function for forecasting from time series
or time series models.  The function invokes particular methods which
depend on the class of the first argument.

For example, the function forecast.Arima makes forecasts based on the
results produced by arima.

The function forecast.ts makes forecasts using ets models (if the data are
non-seasonal or the seasonal period is 12 or less) or stlf (if the seasonal
period is 13 or more).

Usage

forecast(object,...)
## S3 method for class 'ts'
forecast(object, h = ifelse(frequency(object) > 1, 2 * frequency(object), 10) ,
         level=c(80,95), fan=FALSE, robust=FALSE, lambda=NULL, find.frequency=FALSE,
         allow.multiplicative.trend=FALSE, ...)
```

Hyndman and Khandakar (2008) Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, **27**(3). https://www.jstatsoft.org/article/view/v027i03

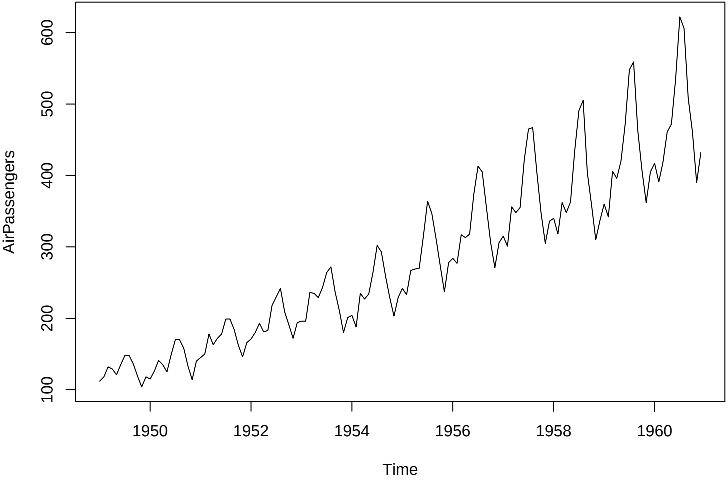
Airline data

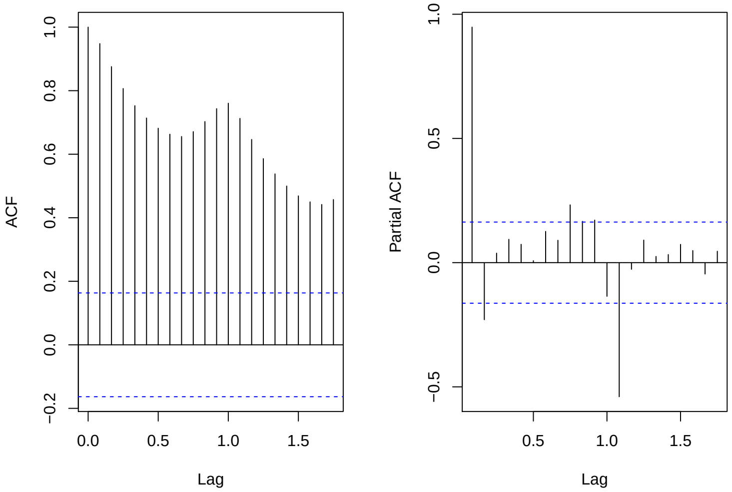
Monthly Airline Passenger Numbers (in thousands) 1949-1960[1].

AirPassengers

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

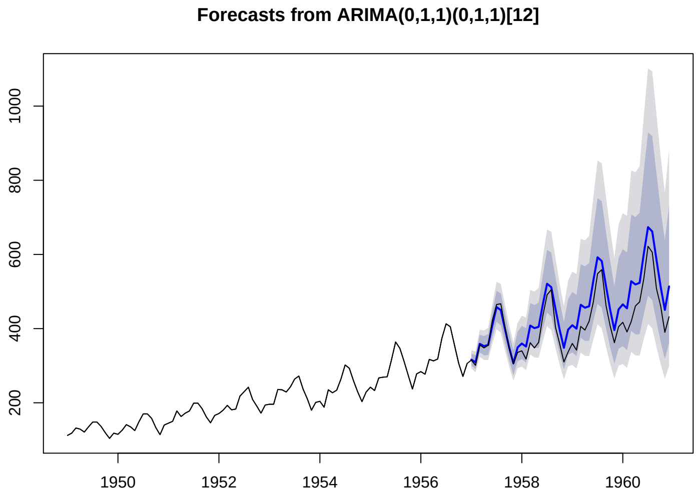
[1] Box, Jenkins, Reinsel (1976) *Time Series Analysis, Forecasting and Control (3rd edition)*. Holden-Day. Series G.


AirPassengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Time

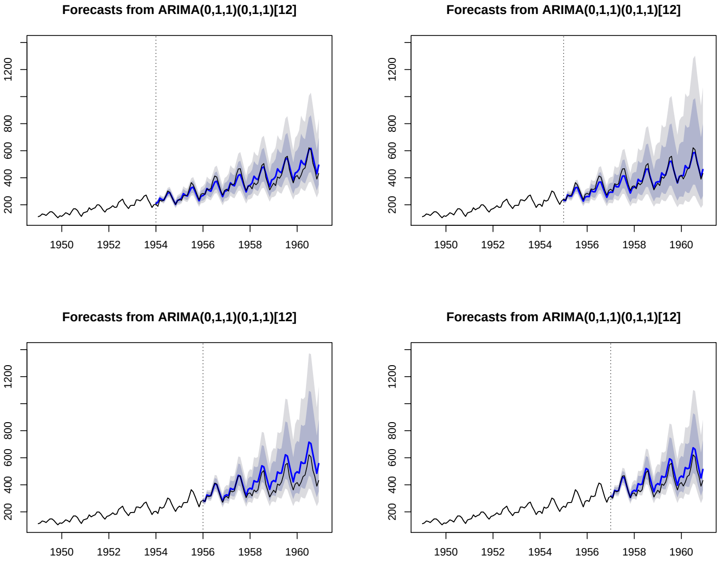

ACF
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.5
1.0
1.5
Lag
Partial ACF
1.0
0.5
0.0
−0.5
0.5
1.0
1.5
Lag

```
install.packages("forecast")

library(forecast)

air.model = Arima(window(AirPassengers,end=1956+11/12),
order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12),
lambda=0)

> air.model
Series: window(AirPassengers, end = 1956 + 11/12)
ARIMA(0,1,1)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
          ma1     sma1
      -0.3941  -0.6129
s.e.   0.1173   0.1076

sigma^2 estimated as 0.001556:  log likelihood=148.76
AIC=-291.53   AICc=-291.22   BIC=-284.27
```

```
plot(forecast(air.model,h=48))
lines(AirPassengers)
```



```
par(mfrow=c(2,2))
for (t in 1953:1956){
  air.model = Arima(window(AirPassengers,end=t+11/12),
  order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12),lambda=0)
  plot(forecast(air.model,h=84-12*(t-1953)),xlim=c(1949,1961),
  ylim=c(100,1400))
  lines(AirPassengers)
  abline(v=t+1,lty=3)
 }
```


Forecasts from ARIMA(0,1,1)(0,1,1)[12]
Forecasts from ARIMA(0,1,1)(0,1,1)[12]
Forecasts from ARIMA(0,1,1)(0,1,1)[12]
Forecasts from ARIMA(0,1,1)(0,1,1)[12]
200
400
600
800
1200
1950
1952
1954
1956
1958
1960

h-STEPS AHEAD RMSE

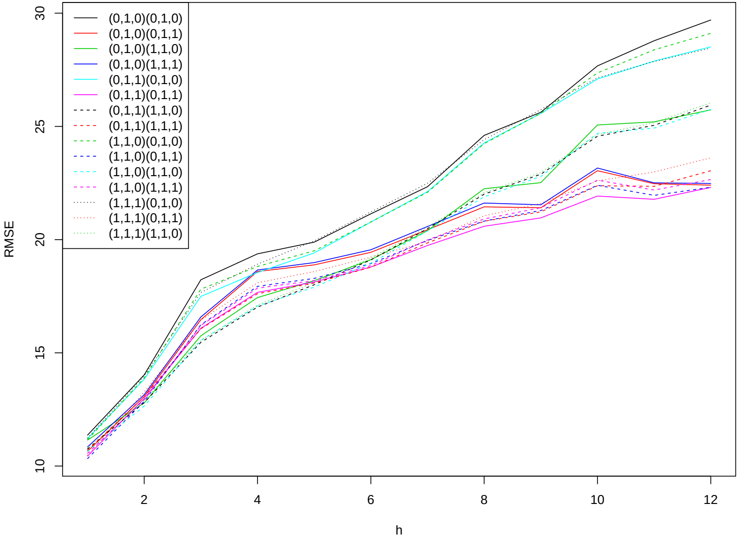
For each t from t_0 to $n-12$, we fit a seasonal ARIMA model based on observations $\{1, \ldots, t\}$ and compute the h-steps ahead forecast:

$$y_t(h) \qquad h = 1, \ldots, 12,$$

and the h-step head root mean square error (RMSE):

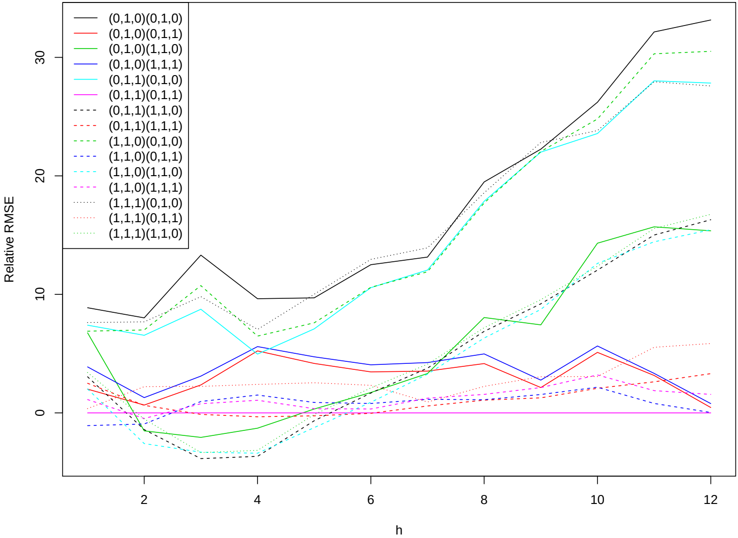
$$\mathrm{RMSE}(h) = \sqrt{\frac{1}{n-11-t_0} \sum_{t=t_0}^{n-12} (y_{t+h} - y_t(h))^2}$$

RMSE



(0,1,0)(0,1,0)
(0,1,0)(0,1,1)
(0,1,0)(1,1,0)
(0,1,0)(1,1,1)
(0,1,1)(0,1,0)
(0,1,1)(0,1,1)
(0,1,1)(1,1,0)
(0,1,1)(1,1,1)
(1,1,0)(0,1,0)
(1,1,0)(0,1,1)
(1,1,0)(1,1,0)
(1,1,0)(1,1,1)
(1,1,1)(0,1,0)
(1,1,1)(0,1,1)
(1,1,1)(1,1,0)
RMSE
10
15
20
25
30
2
4
6
8
10
12
h

Relative RMSE (%)



```
ARIMA(0,1,1)(0,1,1)[12]
Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001371:  log likelihood=244.7
AIC=-483.4   AICc=-483.21   BIC=-474.77

ARIMA(0,1,1)(1,1,1)[12]
Coefficients:
          ma1     sar1     sma1
      -0.4143  -0.1116  -0.4817
s.e.   0.0899   0.1547   0.1363

sigma^2 estimated as 0.001375:  log likelihood=244.96
AIC=-481.91   AICc=-481.6   BIC=-470.41

ARIMA(1,1,0)(1,1,1)[12]
Coefficients:
          ar1     sar1     sma1
      -0.3451  -0.0760  -0.5108
s.e.   0.0828   0.1548   0.1347

sigma^2 estimated as 0.001399:  log likelihood=243.86
AIC=-479.73   AICc=-479.41   BIC=-468.23

ARIMA(1,1,0)(0,1,1)[12]
Coefficients:
          ar1     sma1
      -0.3395  -0.5619
s.e.   0.0822   0.0748

sigma^2 estimated as 0.001391:  log likelihood=243.74
AIC=-481.49   AICc=-481.3   BIC=-472.86
```

```
install.packages("forecast")
library(forecast)

n = length(AirPassengers)
error = array(0,c(15,n-12-48+1,12))
model = matrix(0,15,6)
for (t in 48:(n-12)){
  test = AirPassengers[(t+1):(t+12)]
  ll = 0
  for (i in 0:1) for (j in 0:1) for (k in 0:1)for (l in 0:1){
    ll = ll + 1
    if (ll<16){
      air.model = Arima(AirPassengers[1:t],order=c(i,1,j),
                        seasonal=list(order=c(k,1,l),period=12),lambda=0)
      forec = forecast(air.model,h=12)$mean
      error[ll,t-47,] = test-forec
      model[ll,] = c(i,1,j,k,1,l)
    }
  }
}
rmse = matrix(0,15,12)
rrmse = matrix(0,15,12)
for (i in 1:15) rmse[i,] = sqrt(apply(error[i,,]^2,2,mean))
for (i in 1:15) rrmse[i,] = 100*(rmse[i,]/rmse[6,]-1)

cols = c(1:6,1:6,1:3)
ltys = c(rep(1,6),rep(2,6),rep(3,3))

ts.plot(t(rmse),col=cols,lty=ltys,xlab="h",ylab="Root MSE")
legend("topleft",legend=legends,col=cols,lty=ltys)
ts.plot(t(rrmse),col=cols,lty=ltys,xlab="h",ylab="Root MSE")
legend("topleft",legend=legends,col=cols,lty=ltys)
```

Two best models