Time series analysis: ARIMA models

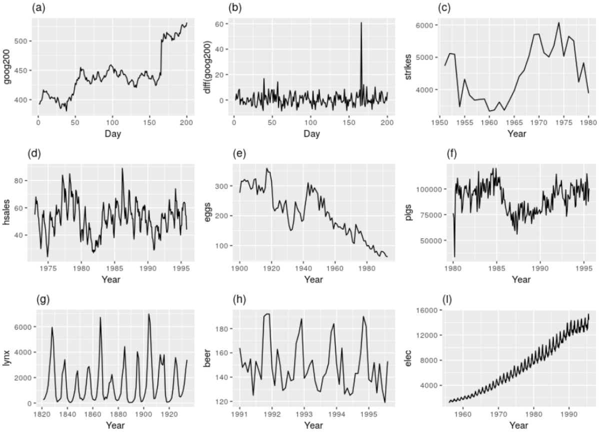
ARIMA models: introduction

- ARIMA models provide a typical approach to time series forecasting.
- Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem.
- While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

Stationarity and differencing

A stationary time series is one whose properties do not depend on the time at which the series is observed.

Thus, time series with trends, or with seasonality, are not stationary.



Differencing

- Differencing can help stabilise the mean of a time series by removing changes in the level of a time series, and therefore eliminating (or reducing) trend and seasonality.

$$y'_t = y_t - y_{t-1}.$$

- Seasonal differencing (for monthly data)

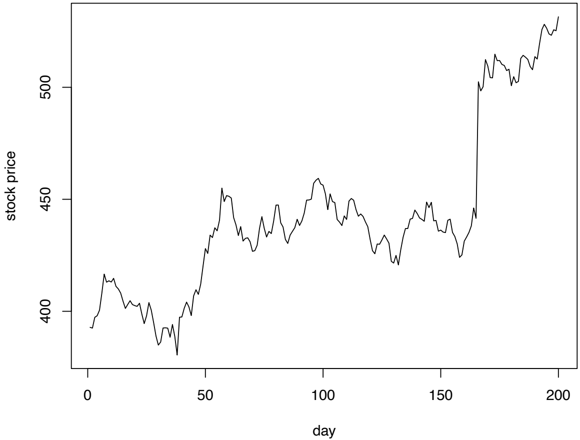
$$y'_t = y_t - y_{t-12}.$$

- A further differencing may be performed

$$y^*_t = y'_t - y'_{t-1} = (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}).$$

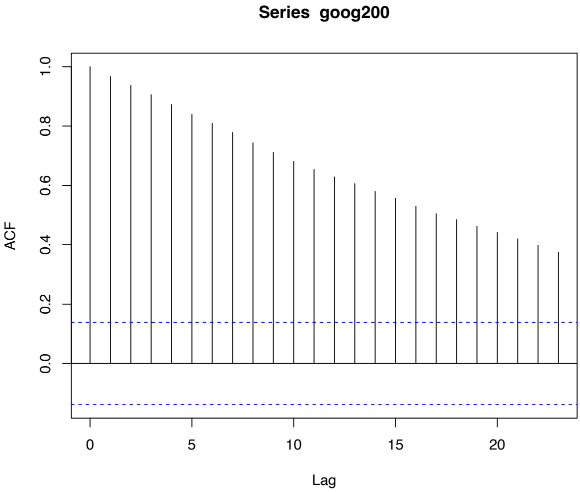
Stationarity and differencing

Google stock price for 200 consecutive days



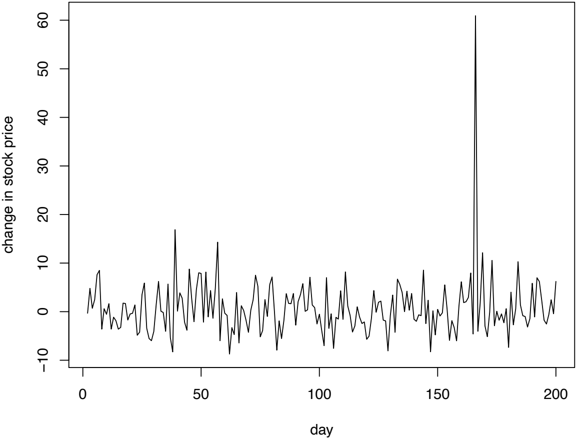
Stationarity and differencing

ACF for Google stock price



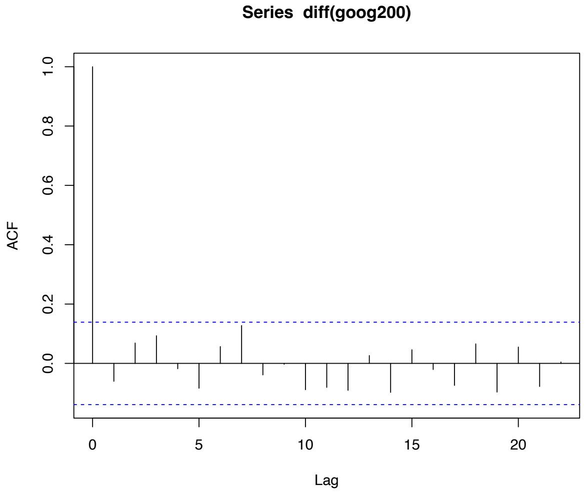
Stationarity and differencing

Daily change in Google stock price for 200 consecutive days



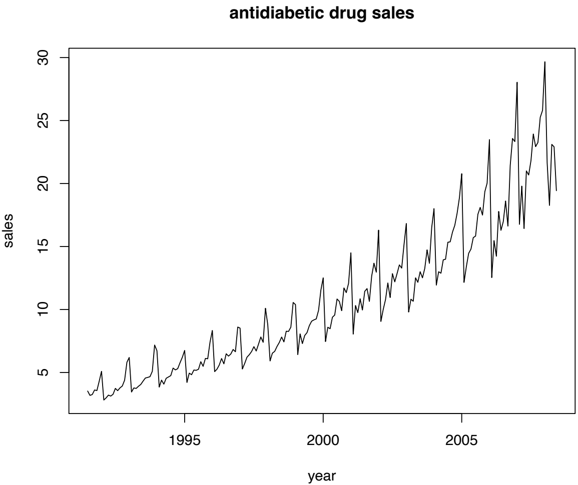
Stationarity and differencing

ACF for daily change in Google stock price

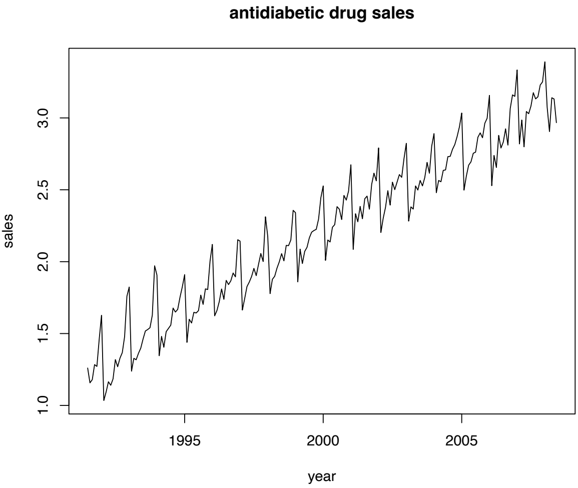


Stationarity and differencing

Monthly sales of antidiabetic drugs

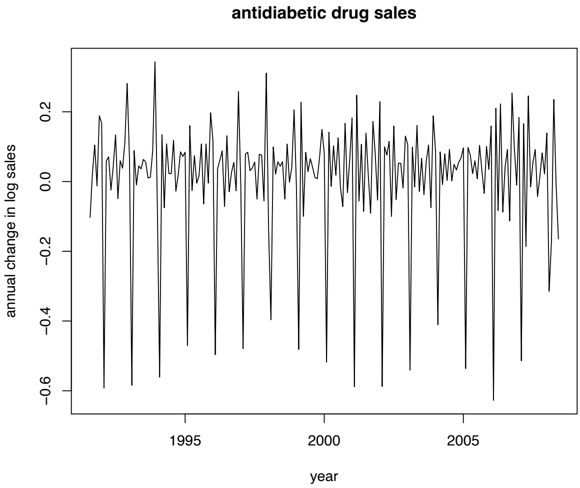


Stationarity and differencing
Logarithmic transformation



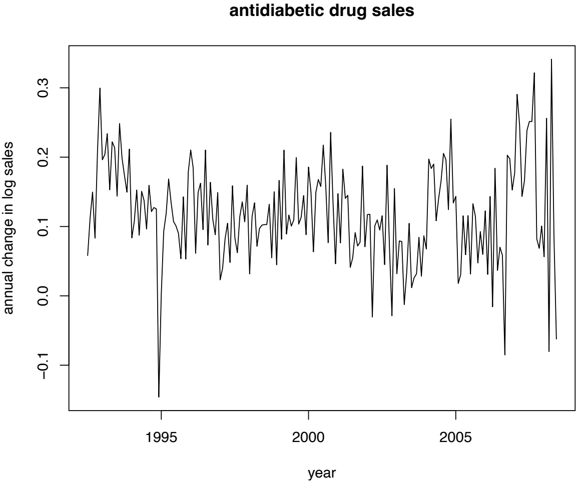
Stationarity and differencing

First differencing



Stationarity and differencing

Seasonal differencing



Backshift notation

The backward shift operator B is a useful notational device when working with time series lags:

$$By_t = y_{t-1}.$$

In other words, B has the effect of shifting the data back one period. Two applications of B shifts the data back two periods

$$B(By_t) = B^2 y_t = y_{t-2}.$$

Backshift notation

The backward shift operator is convenient for describing the process of differencing. A first difference can be written as

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$$

Similarly, if second-order differences have to be computed, then

$$\begin{aligned} y''_t &= (y'_t - y'_{t-1}) \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \\ &= (1 - 2B + B^2)y_t \\ &= (1 - B)^2 y_t \end{aligned}$$

Autoregressive models

- In a multiple regression model, we forecast the variable of interest using a linear combination of predictors.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon.$$

- In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \varepsilon_t$$

- We refer to this as an $AR(p)$, an *autoregressive* model of order p.
- This is like a multiple regression but with lagged values of y_t as predictors.

Autoregressive models

An autoregressive model of order p is normally written as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

For an $AR(1)$ model:

- when $\phi_1 = 0$, y_t is a white noise
- when $\phi_1 = 1$ and $c = 0$, y_t is a random walk
- when $\phi_1 = 1$ and $c \neq 0$, y_t is a random walk with drift
- when $\phi < 0$, y_t tends to oscillate between positive and negative values

Moving-average models

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

$$y_t = c + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

We refer to this as an $MA(q)$, a Moving Average of order q.

ARIMA models

- If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model.
- ARIMA is an acronym for AutoRegressive Integrated Moving Average, ARIMA (p, d, q) where p refers to the AR part, q refers to the MA part and d is the degree of first differencing involved.
- A *White Noise* model $y_t = c + \varepsilon_t$ is an ARIMA$(0, 0, 0)$,
- A *Random Walk* $y_t = y_{t-1} + \varepsilon_t$, is an ARIMA (0,1,0)
- Autoregressive model is ARIMA$(p, 0, 0)$
- Moving average model is ARIMA$(0, 0, q)$

ARMA(p,q) models

An ARMA (p, q) may be expressed as

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_q y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - ... - \theta_q \varepsilon_{t-q}$$

or, by using backshift notation,

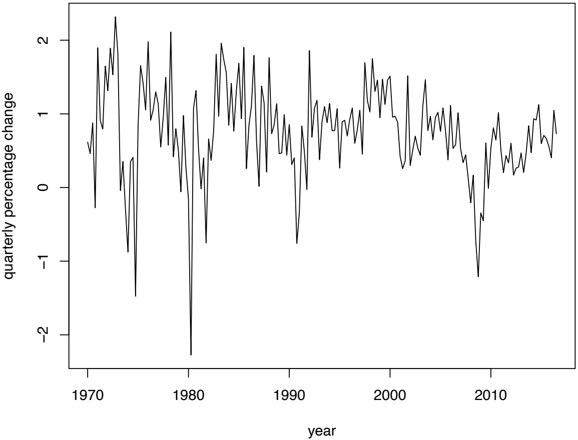
$$(1 - \phi_1 B - ... - \phi_p B^p) y_t = c + (1 - \theta_1 B - ... - \theta_q B^q) \varepsilon_t$$

An ARMA$(1, 1)$ is defined

$$(1 - \phi_1 B) y_t = c + (1 - \theta_1 B) \varepsilon_t$$

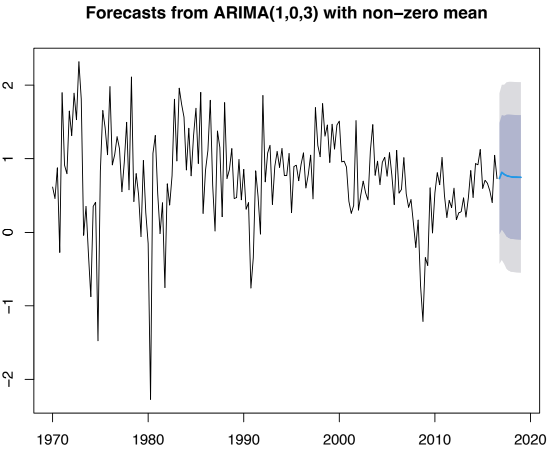
Example

US percentage consumption



Example

US percentage consumption



ARIMA(p,d,q) models

- If an ARMA (p, q) model is non stationary, we obtain an ARIMA (p, d, q) model.
- the simplest case, ARIMA (1,1,1), is defined as

$$(1 - \phi_1 B)(1 - B)y_t = c + (1 - \theta_1 B)\varepsilon_t$$

Note: In practice, it is seldom necessary to deal with values of p, d and q other than $0, 1, 2$. Such a small range of values can cover a large range of practical forecasting situations.

ARIMA(p,d,q) models

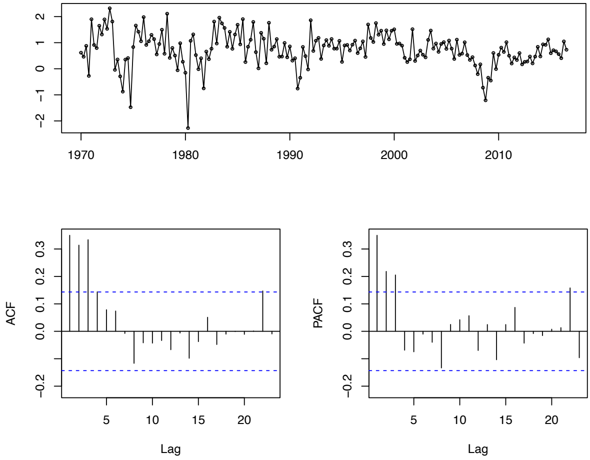
- It is usually not possible to tell, simply from a time plot, what values of p and q are appropriate for the data
- However, it is sometimes possible to use ACF and PACF plots to determine them
- PACF: measures the partial autocorrelations, i.e. the relationship between y_t and y_{t-k} after removing the effects of other time lags $(1, 2, 3, ..., k-1)$

ARIMA(p,d,q) models

- the data may follow an ARIMA$(p, d, 0)$ if the ACF and PACF of differenced data show these patterns
  - ACF exponentially decaying or sinusoidal
  - significant spike at the lag p in PACF and nothing else beyond
- the data may follow an ARIMA$(0, d, q)$ if the ACF and PACF of differenced data show these patterns
  - PACF exponentially decaying or sinusoidal
  - significant spike at the lag q in ACF and nothing else beyond

Example

US percentage consumption, time series display



ARIMA(p,d,q) models

The constant c has an important effect on the long-term forecasts

- if $c = 0$ and $d = 0$, the long-term forecast will go to zero
- if $c = 0$ and $d = 1$, the long-term forecast will go to a non-zero constant
- if $c = 0$ and $d = 2$, the long-term forecast will follow a straight line
- if $c \neq 0$ and $d = 0$, the long-term forecast will go to the mean of the data
- if $c \neq 0$ and $d = 1$, the long-term forecast will follow a straight line
- if $c \neq 0$ and $d = 2$, the long-term forecast will follow a quadratic trend

Modelling procedure

A useful general approach is the following:

1. plot the data, and identify unusual observations
2. if necessary, transform the data
3. if the data are non-stationary, take first differences
4. examine ACF and PACF: how do they behave?
5. try chosen models and use the AIC to search for the best
6. check the residuals: do they behave like WN?
7. if residuals look like WN, calculate forecasts

ARIMA(p,d,q) and seasonality

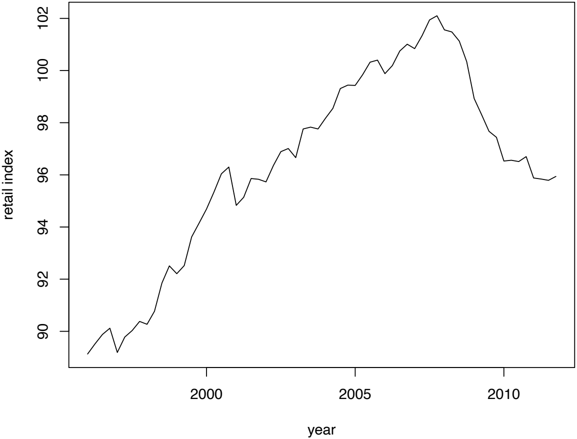
A further extension to ARMA models concerns seasonality.

- an ARIMA model with seasonal components is an ARIMA $(p, d, q)(P, D, Q)_s$, where (p, d, q) indicates the non-seasonal part of the model, while (P, D, Q) indicates the seasonal part of order s.
- the seasonal ARIMA model $(1, 1, 1)(1, 1, 1)_4$ is

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 - \theta_1 B)(1 - \Theta_1 B^4)\varepsilon_t$$

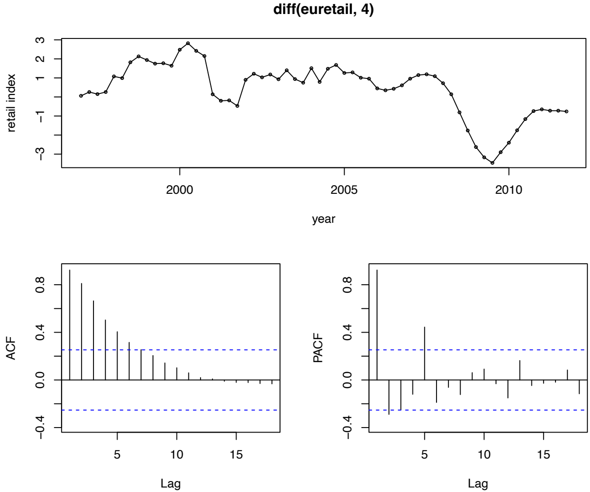
Example

European quarterly retail trade

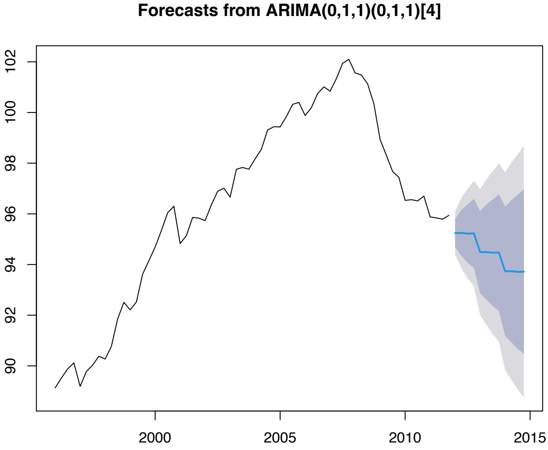


Example

European quarterly retail trade

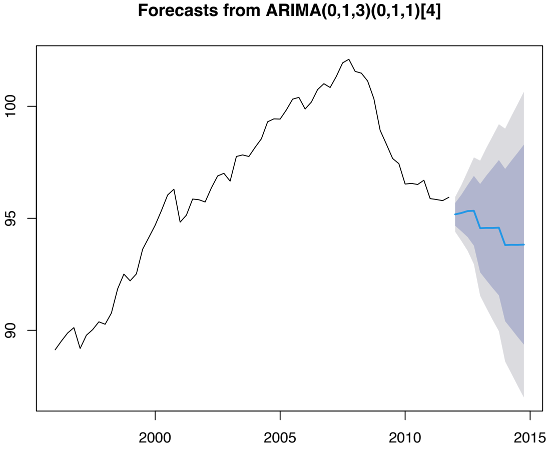


Example
Manual model selection



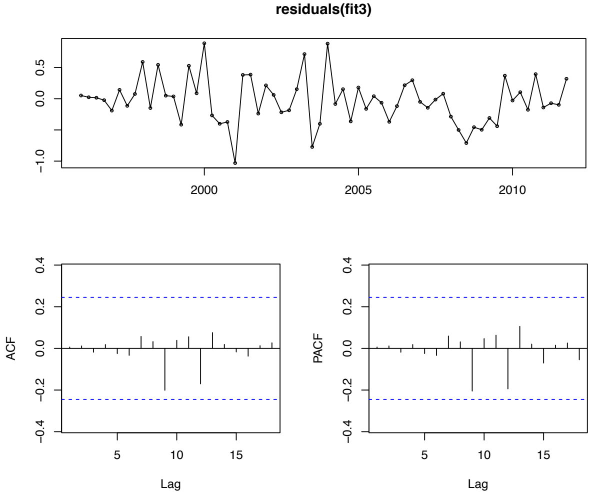
Example

Automatic model selection with auto.arima



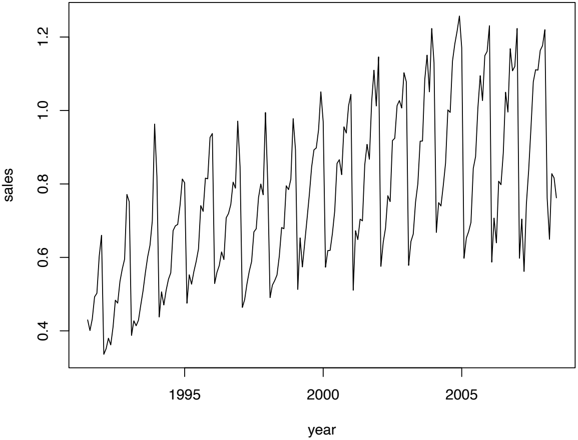
Example

Residuals from 'fit3'



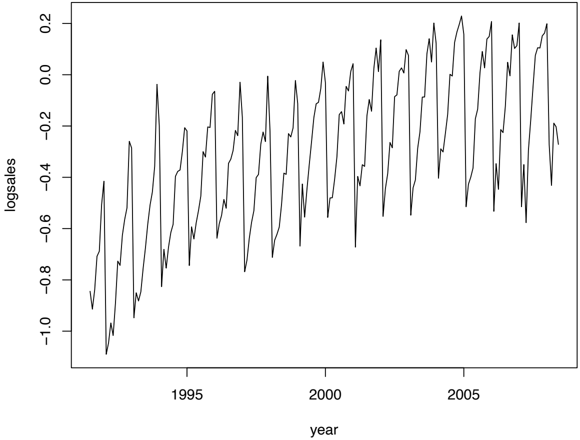
Example

Cortecosteroid drug sales in Australia
Plot the data, and identify unusual observations



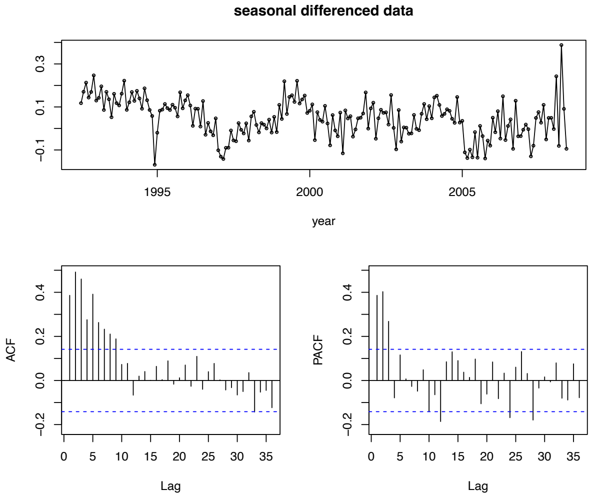
Example

If necessary, transform the data
Logarithmic transformation to stabilize variance



Example

If the data are non-stationary, take first differences
examine ACF and PACF: how do they behave?



Example

Try chosen models and use the AIC to search for the best

```
Series: h02
ARIMA(3,0,1)(0,1,2)[12]
Box Cox transformation: lambda= 0

Coefficients:
          ar1     ar2     ar3     ma1      sma1     sma2
      -0.1603  0.5481  0.5678  0.3827   -0.5222  -0.1768
s.e.   0.1636  0.0878  0.0942  0.1895    0.0861   0.0872

AIC=-486.08   AICc=-485.48   BIC=-463.28

Training set error measures:
                      ME       RMSE        MAE       MPE     MAPE
Training set 0.001736226 0.0493495 0.03596218 0.2515195 4.621055
                    ACF1
Training set -0.01964279
```

Example

Model selection

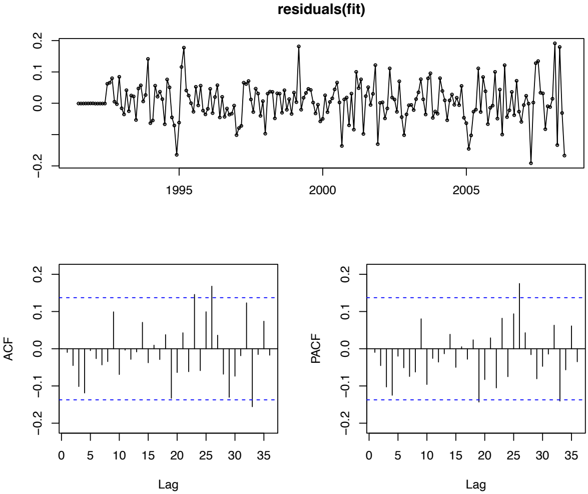
Akaike's Information Criterion is useful for determining the order of an ARIMA model.

$$\text{AIC} = T \log\left(\frac{SSE}{T}\right) + 2p$$

Suitable modifications of the AIC are the AIC_c and the BIC.
Good models are obtained by minimizing either the AIC, AIC_c or BIC.

Example

Check the residuals: do they behave like WN?



Example

Check the residuals: do they behave like WN?

In addition to looking at the ACF plot, we can do a more formal test for autocorrelation by considering a whole set of r_k values as a group.
We may consider the Box-Pierce test based on

$$Q = T \sum_{k=1}^{h} r_k^2$$

where h is the maximum lag being considered and T is the number of observations. If each r_k is close to zero then Q will be small.
A related test is the Ljung-Box test based on

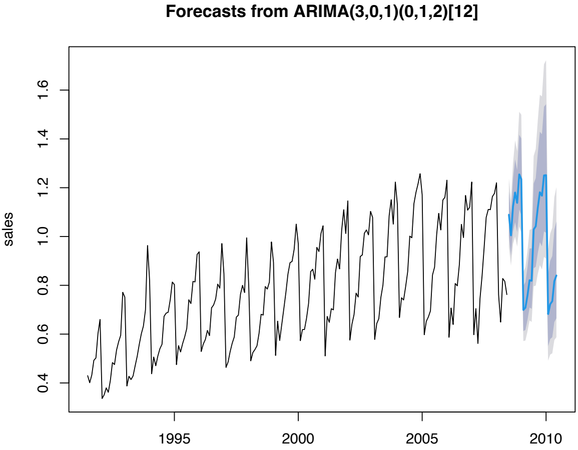
$$Q* = T(T+2) \sum_{k=1}^{h} (T-k)^{-1} r_k^2$$

'Large' values of Q and $Q*$ suggest that the autocorrelations do not come from a white noise.

Example

Calculate forecasts.

None of the models considered pass all the residual tests. In practice, we would normally use the best model we could find even if it did not pass all the tests.



Example

A note on prediction intervals

- A prediction interval gives an interval within which we expect y_t to lie with a specified probability (80%, 95%)
- the value of prediction intervals is that they express the uncertainty in the forecasts
- a common feature of prediction intervals is that the further ahead we forecast, the more uncertainty is associated with the forecast, and so the prediction intervals grow wider
- the forecast intervals for ARIMA models are based on assumptions of uncorrelation and normality of residuals: for this reason it is important to check the behavior of residuals.