STAT 443: Time Series and Forecasting

Chapter 4

Prediction for Time Series

"It is far better to foresee even without certainty than not to foresee at all."

/Henri Poincare/

Introduction

- One of the principle aims of studying and modelling time series is to make predictions of future values

Forecasting methods can be categorized in several ways:

- Subjective vs. model-based (not necessarily mutually exclusive)
- "Automatic" vs. "non-automatic" wrt. the user input
- Purpose of the forecast, for example,
  - Point vs. interval forecasts
  - Expected vs. extreme values

Framework and notation

- We restrict attention to univariate forecasting methods, i.e. we consider a single time series in isolation
  - Clearly, more can be gained by using cross-correlated series but that requires modelling of multivariate time series

Notation:

- Given a time series $\{x_1, \ldots, x_n\}$ up to time n, the aim is to predict the value $x_{n+\ell}$ which is ℓ time steps in the future
- The integer value ℓ is known as the lead time
- The predicted value is denoted $\hat{x}_n(\ell)$

Outline

We will next discuss several commonly used forecasting methods:

1. Extrapolation of trend curves
2. Exponential smoothing
3. Holt-Winters forecasting
4. Box-Jenkins forecasting

Extrapolation of trend curves

- Advantage: simplicity
- Disadvantages:
  - The method is clearly non-dynamic
  - Assumes the fitted curve is acceptable not just for the observed values but also for future values
  - Choice of the curve is not straightforward: two competing curves fitting the data almost equally well may give very different predictions when extrapolated to future times
  - Observations at the start of the series exert undue influence on the curve choice, which for obvious reasons should have least impact on future behaviour
- Can be used if a series is too short for anything more sophisticated, or it is clear that a simple parametric curve captures the key features in the data and this likely to hold for future values

Exponential smoothing

- Suitable only for stationary time series, hence if there are non-stationarity effects they need to be removed prior to applying this procedure

- Basic idea: take as an estimate of future value x_{n+1} a weighted sum of past observations:

$$\hat{x}_n(1) = \omega_0 x_n + \omega_1 x_{n-1} + \cdots$$

  for some weights $\omega_0,\ \omega_1, \ldots$

- It is sensible to choose a decreasing sequence of weights to give more weight to the recent past

Exponential smoothing (cont'd)

Geometric smoothing: $\omega_i = \alpha(1-\alpha)^i$ for $i = 0, 1, \ldots$ and $0 < \alpha < 1$

- This gives:

$$\hat{x}_n(1) = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots$$

- Since only a finite number of observations is available, it is usual to write:

$$\begin{aligned}\hat{x}_n(1) &= \alpha x_n + (1-\alpha)\{\alpha x_{n-1} + \alpha(1-\alpha)x_{n-2} + \cdots\} \\ &= \alpha x_n + (1-\alpha)\hat{x}_{n-1}(1) \qquad\qquad (*)\end{aligned}$$

- Eq. $(*)$ can be used recursively to compute forecasts by setting

$$\hat{x}_1(1) = x_1$$

- Eq. $(*)$ also facilitates computations, since the forecast is simply a weighted sum of the past observation and the previous forecast

Exponential smoothing (cont'd)

Alternatively , Eq. $(*)$ can be re-written as:

$$\begin{aligned}\hat{x}_n(1) &= \alpha x_n + (1-\alpha)\hat{x}_{n-1}(1) \qquad (*)\\ &= \alpha(x_n - \hat{x}_{n-1}(1)) + \hat{x}_{n-1}(1)\\ &= \alpha e_n + \hat{x}_{n-1}(1)\end{aligned}$$

where $e_n := x_n - \hat{x}_{n-1}(1)$ is the prediction error at time n

Exponential smoothing - the choice of parameter α

$$\boxed{\hat{x}_n(1) = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots}$$

Activity: Exponential Smoothing - Questions 1 and 2

$$\boxed{\hat{x}_n(1) = \alpha x_n + \alpha(1-\alpha)x_{n-1} + \alpha(1-\alpha)^2 x_{n-2} + \cdots}$$

- Note:
  - Small α: dependence on a large number of past observations
  - Large α: forecasts are greatly influenced by recent observations

- In general, the choice of α is not an exact science

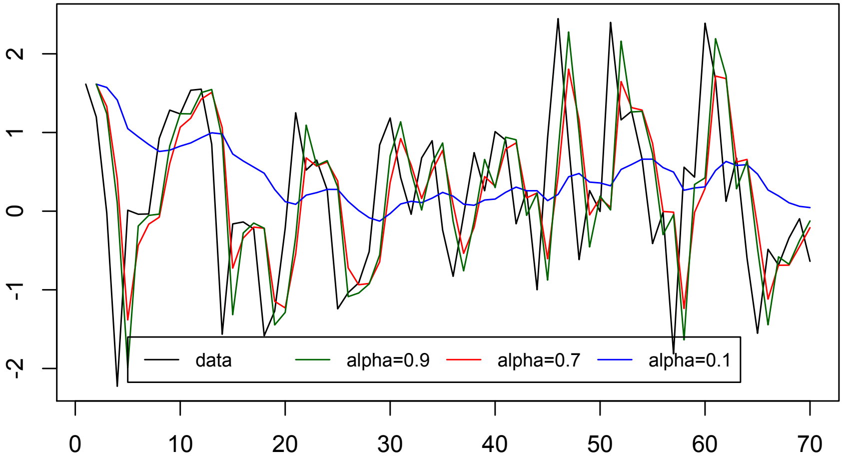
- Possible approaches:
  - Choose a value (close to one) which has worked well for similar data in the past
  - Choose α that minimizes the sum of squared errors:

$$\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (x_t - \hat{x}_{t-1}(1))^2$$

    - has to be done numerically; e.g. using grid search or iterative search
    - In most cases, the sum of squares curve is relatively flat, so the exact choice of α is not crucial

Exponential smoothing - the choice of parameter α (cont'd)

The value of α that minimizes the sum of squared errors here is 0.68 (red line)



Exponential smoothing - concluding remarks

- This forecasting technique has been shown to work quite well in practice
- The only underlying assumption: the series has to have "memory"
  - so that the future depends on the past
- The shortcoming is that it is not possible to indicate the distribution of the prediction errors $\{e_t\}$ and hence to construct prediction intervals
- Exponential smoothing can be extended to series with a trend and seasonal variation, known as Holt-Winters forecasting

The Holt's method, or double exponential smoothing

- Extension of exponential smoothing to series with a trend
- Recall exponential smoothing: $\hat{x}_t(1) = \alpha x_t + (1-\alpha)\hat{x}_{t-1}(1)$
  - can be thought of as giving a forecast for the expected "level" of the series at the next time point
- Let L_t be the "level" of the series at time t
- In analogy with the exponential smoothing, we have:

$$L_t = \alpha x_t + (1-\alpha) L_{t-1} \qquad \text{for some } \alpha$$

- Suppose the series has a trend with the expected change per unit time at time t being T_t, likely dependent on the level at that time

The Holt's method, or double exponential smoothing (cont'd)

Holt(1957) suggested the following relationships:

1. $L_t = \alpha x_t + (1-\alpha)(L_{t-1} + T_{t-1})$
2. $T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

The model-based forecast is of the form: $\hat{x}_t(\ell) = L_t + \ell T_t$ for $\ell = 1, 2, \ldots$

Remarks:

- Two parameters to be estimated α and β, usually restricted to (0,1)
- $\beta = 0$ defaults to simple exponential smoothing

Activity: Exponential Smoothing - Question 3

Holt-Winters forecasting

- Extension to series with trend and seasonal variation
- Let I_t denote the seasonal effect at time t
  - The seasonal effect can be either additive or multiplicative
  - The de-seasonalized series is then either $x_t - \mathrm{I}_t$ or x_t/I_t, resp.
- In case of the multiplicative seasonal effect of period p, the updating equations are

$$
\begin{aligned}
L_t &= \alpha\left(\frac{x_t}{\mathrm{I}_{t-p}}\right) + (1-\alpha)(L_{t-1} + T_{t-1}) \\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \\
\mathrm{I}_t &= \gamma\left(\frac{x_t}{L_t}\right) + (1-\gamma)\mathrm{I}_{t-p}
\end{aligned}
$$

and the forecast is given by

$$\hat{x}_t(\ell) = (L_t + \ell T_t)\mathrm{I}_{t-p+\ell}$$

Holt-Winters forecasting (cont'd)

- Similarly, in case of the additive seasonal effect, the updating equations are

$$
\begin{aligned}
L_t &= \alpha\big(x_t - \mathrm{I}_{t-p}\big) + (1-\alpha)\big(L_{t-1} + T_{t-1}\big)\\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}\\
\mathrm{I}_t &= \gamma\big(x_t - L_t\big) + (1-\gamma)\mathrm{I}_{t-p}
\end{aligned}
$$

  and the forecast is given by

$$\hat{x}_t(\ell) = L_t + \ell T_t + \mathrm{I}_{t-p+\ell}$$

Holt-Winters forecasting (cont'd)

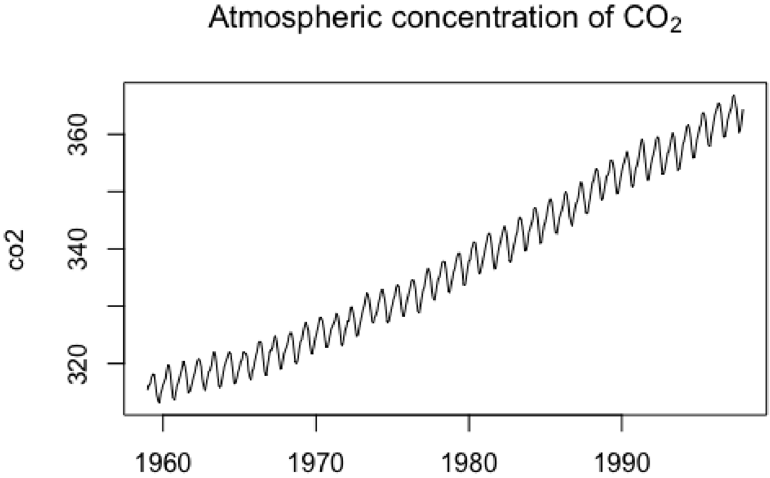
Summary of the steps:

1. Determine whether a seasonal effect is present, and if so whether it is additive or multiplicative
2. Consider what starting values for L, T and $\mathrm{I}_1, \ldots, \mathrm{I}_p$ might sensibly be assigned
3. Decide on the values of α, β and γ – typically this is done by minimising $\sum_t e_t^2$ over either the whole data set or a reasonable subset

Concluding remark:

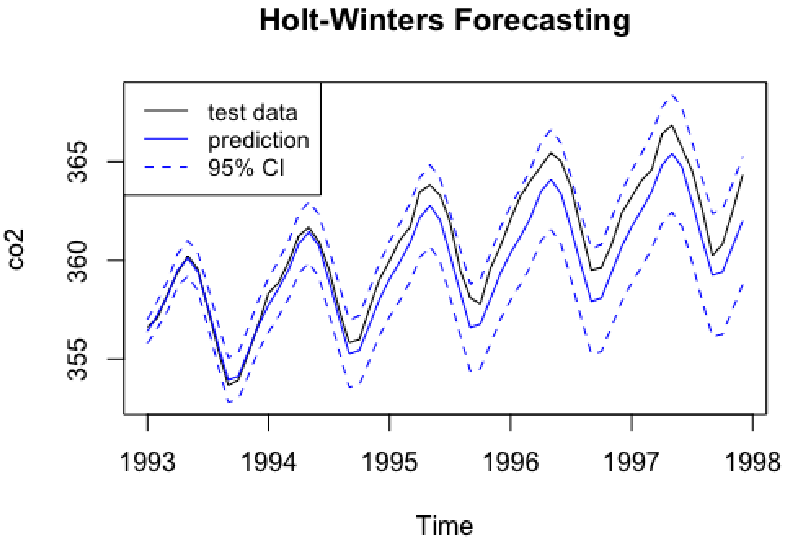
- The method is favoured in practice due to being both fairly straightforward to implement and comparatively accurate in its forecasts

Example



Atmospheric concentration of CO2
co2
320
340
360
1960
1970
1980
1990

Example (cont'd)



Box-Jenkins forecasting

Historical note:

- Based on the pioneering work of Box and Jenkins in 1960's
- Their main contributions were
  - showing that various types of non-stationary effects can be removed by differencing
  - formulating a coherent strategy for forecasting using ARIMA models
- Due to the popularity of their approach, ARIMA models became also known as "Box-Jenkins" models

Box-Jenkins forecasting (cont'd)

A five-step model fitting procedure:

1. If necessary, reduce the observed series to stationarity (usually by differencing and/or trend/seasonal effect removal)
2. Having examined facets of the data (such as the acf and pacf), select an appropriate ARMA model for the time series
3. Estimate the parameters for the fitted model
4. Perform diagnostic measures to assess the goodness–of–fit of the model fitted
5. If necessary, examine alternative models for the data from the ARMA family

Box-Jenkins forecasting (cont'd)

The final step, forecasting, can be performed in one of three different ways, each having its merits:

1. Using the model equation
   - a natural method to obtain point forecasts
2. Using the MA representation of the model
   - useful for computing prediction intervals
3. Using the AR representation of the model
   - gives computational efficiency via an iterative procedure

Box-Jenkins forecasting using the model equation

Given a satisfactory fitted model for the time series in hand,
the forecasting procedure uses:

1. previous, observed values of X and Z
2. zero for future values of Z which have not been observed
3. the expectation of future values of X for the prediction
   - This refers to the conditional expectation of the value to be predicted given the observed data

$$\mathbb{E}[X_{n+\ell} \mid X_n, X_{n-1}, \ldots]$$

Box-Jenkins forecasting using the model equation (cont'd)

Example

- Consider the AR(1) process: $X_t = \alpha X_{t-1} + Z_t$
- The value at time $t+1$ is then $X_{t+1} = \alpha X_t + Z_{t+1}$
- Hence, the one-step ahead forecast is

$$\begin{aligned}\hat{x}_t(1) &= \mathbb{E}[X_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots] \\ &= \mathbb{E}[\alpha X_t + Z_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots] \\ &= \alpha x_t\end{aligned}$$

- The two-step ahead forecast is

$$\begin{aligned}\hat{x}_t(2) &= \mathbb{E}[X_{t+2} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots] \\ &= \mathbb{E}[\alpha X_{t+1} + Z_{t+2} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots] \\ &= \alpha \mathbb{E}[X_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots] \\ &= \alpha \hat{x}_t(1) = \alpha^2 x_t\end{aligned}$$

Box-Jenkins forecasting using the model equation (cont'd)

Example (cont'd)

- Forecasts for larger lead times can be computed similarly
- Of course, in practice an estimate of parameter α is required first

$* * *$

Box-Jenkins forecasting using the model equation (cont'd)

- Models with an MA component require that we have "observed" values of the noise process Z_t for values of $t = 2, \ldots, n$ in order to predict future values of X_t
- The usual approach is to take the residual at time t as the estimate of z_t

Example

- *Janacek and Swift (1993, p. 151–152)* examine a time series which is a record of daily temperature readings
- The data are initially de-seasonalized, and then differenced to attain stationarity; this gives a series $\{x_2, \ldots, x_{360}\}$

Example (cont'd)

- An MA(4) model has been fitted to these data:

$$\boxed{X_t = Z_t + 0.07Z_{t-1} - 0.31Z_{t-2} - 0.13Z_{t-3} - 0.20Z_{t-4}}$$

- Below is a sample of the data, fitted values and residuals:

t	x_t	$\hat{x}_t$	$\hat{z}_t$
2	–0.13	0.12	–0.25
3	6.82	0.92	5.90
4	1.03	0.88	0.15
5	–0.02	–1.35	1.33
⋮	⋮	⋮	⋮
357	4.94	0.63	4.31
358	0.85	0.41	0.44
359	–0.64	–1.41	0.77
360	–4.62	–0.58	–4.04

- Exercise: predict the next value in the series, x_{361}, using the Box-Jenkins method

Box-Jenkins forecasting

Recall: Box-Jenkins forecasting can be performed in one of three different ways, each having its merits:

1. Using the model equation [done!]
   - a natural method to obtain point forecasts
2. Using the MA representation of the model [NEXT!]
   - useful for computing prediction intervals
3. Using the AR representation of the model
   - gives computational efficiency via an iterative procedure

In-class activity: MA representations

Box-Jenkins forecasting using the MA representation of the model

- Recall: any ARMA process $\phi(B)X_t = \theta(B)Z_t$ can be written as an MA process of possibly infinite order

$$X_t = \frac{\theta(B)}{\phi(B)} Z_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$$

- So

$$X_{n+\ell} = Z_{n+\ell} + \psi_1 Z_{n+\ell-1} + \psi_2 Z_{n+\ell-2} + \cdots$$

- Replacing future values of Z, i.e. all Z_t with $t > n$, by zero gives

$$\hat{X}_n(\ell) = \sum_{i=\ell}^{\infty} \psi_i Z_{n+\ell-i}$$

- The forecast error is hence

$$e_n(\ell) := X_{n+\ell} - \hat{X}_n(\ell) = Z_{n+\ell} + \psi_1 Z_{n+\ell-1} + \cdots + \psi_{\ell-1} Z_{n+1}$$

Prediction intervals

$$\boxed{e_n(\ell) := X_{n+\ell} - \hat{X}_n(\ell) = \sum_{i=0}^{\ell-1} \psi_i Z_{n+\ell-i}}$$

- $\mathbb{E}(e_n(\ell)) = \sum_{i=0}^{\ell-1} \psi_i \mathbb{E}(Z_{n+\ell-i}) = 0$
- $Var(e_n(\ell)) = \sum_{i=0}^{\ell-1} \psi_i^2 Var(Z_{n+\ell-i}) = \sigma^2 \sum_{i=0}^{\ell-1} \psi_i^2$
- If $Z_t \sim \mathcal{N}(0, \sigma^2)$ then $e_n(\ell) \sim \mathcal{N}\left(0, \sigma^2 \sum_{i=0}^{\ell-1} \psi_i^2\right)$
- Estimates $\hat{\sigma}^2$ and $\{\hat{\psi}_1, \ldots, \hat{\psi}_{\ell-1}\}$ can be obtained from the data and the fitted model

Prediction intervals (cont'd)

- Assume normality for $e_n(\ell)$: $e_n(\ell) \sim \mathcal{N}\Big(0, Var(e_n(\ell))\Big)$
- Note: $X_{n+\ell} = \hat{X}_n(\ell) + e_n(\ell)$
- Hence, conditionally on the observed data we have

$$X_{n+\ell} \sim \mathcal{N}(\hat{x}_n(\ell), \widehat{Var}(e_n(\ell))) \;\Rightarrow\; \frac{X_{n+\ell} - \hat{x}_n(\ell)}{\sqrt{\widehat{Var}(e_n(\ell))}} \sim \mathcal{N}(0,1)$$

so that

$$0.95 = \mathbb{P}\Big\{ -1.96 \leq \frac{X_{n+\ell} - \hat{x}_n(\ell)}{\sqrt{\widehat{Var}(e_n(\ell))}} \leq 1.96 \Big\}$$

$$\begin{aligned} 0.95 &= \mathbb{P}\Big\{ -1.96 \leq \frac{X_{n+\ell} - \hat{x}_n(\ell)}{\sqrt{\widehat{Var}(e_n(\ell))}} \leq 1.96 \Big\} \\ &= \mathbb{P}\Big\{ \hat{x}_n(\ell) - 1.96\sqrt{\widehat{Var}(e_n(\ell))} \leq X_{n+\ell} \leq \hat{x}_n(\ell) + 1.96\sqrt{\widehat{Var}(e_n(\ell))} \Big\} \end{aligned}$$

Prediction intervals (cont'd)

In general, assuming approximate normality for $e_n(\ell)$,
an approximate $(1-\alpha)100\%$ prediction interval for $X_{n+\ell}$ is given by

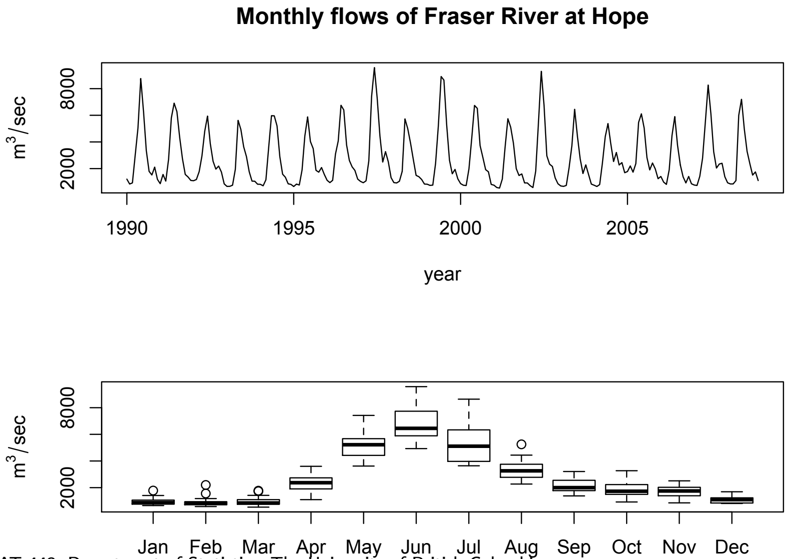
$$\hat{x}_n(\ell) \pm \Phi^{-1}(1-\alpha/2)\ \hat{\sigma}\sqrt{\sum_{i=0}^{\ell-1} \hat{\psi}_i^2}$$

where $\Phi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution

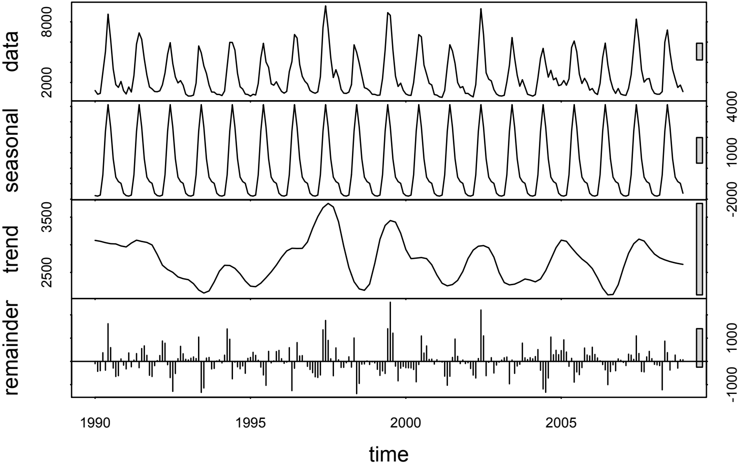
Remarks:

- The larger ℓ, the wider the prediction interval becomes
- If Z_t's are normally distributed, the forecast errors are exactly normally distributed

Case study: prediction of monthly mean flows of Fraser River at Hope



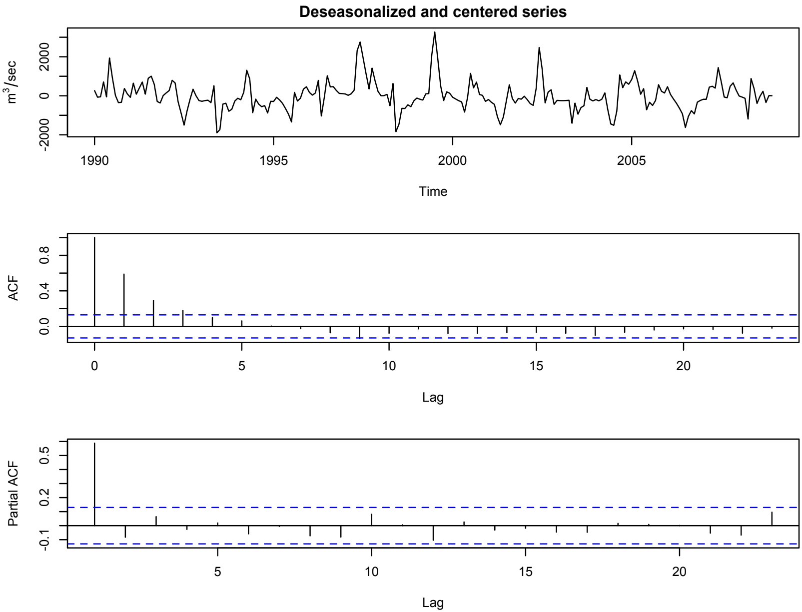
Seasonal decomposition by Loess:



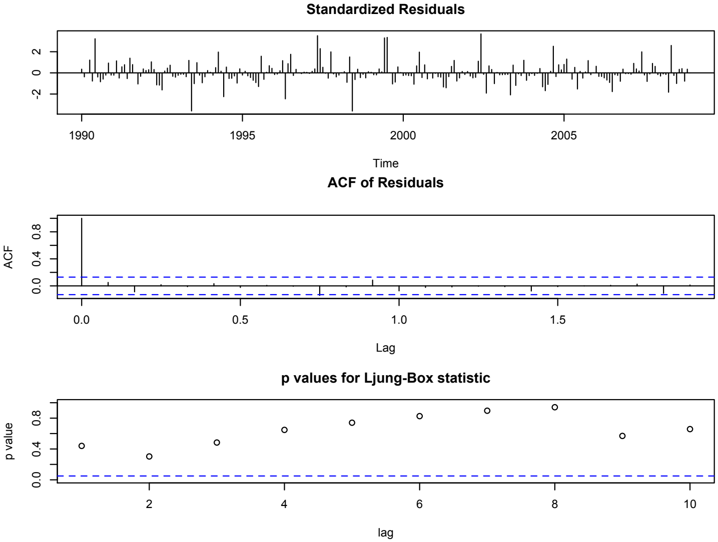
R code to de-seasonalize the series:

```
flow.stl = stl(flow, "period")

# de-seasonalizing
flow.ds = flow - flow.stl$time.series[,"seasonal"]

# re-centering
mu.hat = mean(flow.ds)
flow1 = flow.ds - mu.hat
```

Model selection from the ARMA family:

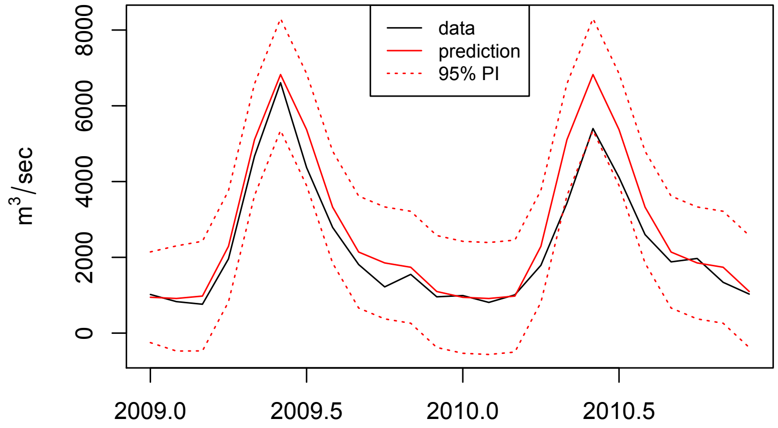


AR(1) seems to be a very clear choice here with the sample acf decaying and the sample pacf "cutting-off" sharply at lag 1



Prediction

```
# predicting de-seasonalized series
flow.pred = predict(fm, 24)

# adjusting for the mean and seasonal effect
flow.pred$pred = flow.pred$pred + mu.hat
                    + flow.stl$time.series[,"seasonal"][1:24]

# prediction intervals
lb = flow.pred$pred - 1.96*flow.pred$se
ub = flow.pred$pred + 1.96*flow.pred$se
```


data
prediction
95% PI
m^3/sec
0
2000
4000
6000
8000
2009.0
2009.5
2010.0
2010.5

Box-Jenkins forecasting

Recall: Box-Jenkins forecasting can be performed in one of three different ways, each having its merits:

1. Using the model equation [done!]
   - a natural method to obtain point forecasts
2. Using the MA representation of the model [done!]
   - useful for computing prediction intervals
3. Using the AR representation of the model [NEXT!]
   - gives computational efficiency via an iterative procedure

Box-Jenkins forecasting and the AR representation

- Recall: An ARMA(p,q) process can be expressed as a pure AR process of the form

$$\boxed{\pi(B)X_t = Z_t, \qquad \pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i}$$

  for some coefficients π_i

- This gives

$$X_{N+\ell} = \pi_1 X_{n+\ell-1} + \pi_2 X_{n+\ell-2} + \cdots + \pi_\ell X_N + \cdots + Z_{N+\ell}$$

- Hence, the ℓ-step ahead forecast can be written as

$$\begin{aligned} \hat{x}_N(\ell) &= \hat{\mathbb{E}}\big(X_{N+\ell} \mid X_N = x_N, \ldots, X_1 = x_1\big) \\ &= \pi_1 \hat{x}_N(\ell-1) + \cdots + \pi_\ell x_N + \pi_{\ell+1} x_{N-1} + \ldots, \end{aligned}$$

  which can be used to compute forecasts recursively

Activity: Box-Jenkins forecasting

Optimality of the Box-Jenkins forecasting procedure

Forecasting methods - Summary

- We have discussed two types of forecasting procedures
- "**Model-free**" methods include:
  - Exponential smoothing: suitable for stationary time series
  - Holt's method: applied to time series with a trend
  - Holt-Winters method: applied to time series with a trend and seasonal variation
- Time series **model-based** method:
  - Box-Jenkins forecasting: makes use of an ARMA model fitted to a series that has been reduced to stationarity

Concluding remarks

- Neither model fitting nor forecasting can be described as "automatic" procedures, where the user simply follows a set of rules in order to arrive at results
- Rather, a little knowledge of the data, a grasp of the type of models which are candidates to fit to the data along with at least some experience of analyzing time series can lead to fruitful modelling useful for prediction
- Prediction is a dangerous business, and no great faith should be put in any predictions, even those based on a seemingly good model which fits well to a lengthy series
- It is also beyond the scope of a model to predict outside events which would substantially influence future data