Syllabus.

- Chapter 1. Concepts in time series.
- Chapter 2. Univariate ARIMA models.
- Chapter 3. Model fitting and checking.
- Chapter 4. Prediction and model selection.
- Chapter 5. Outliers and influential observations.
- Chapter 6. Heterocedastic models.
- Chapter 7. Multivariate time series.

Chapter 4. Prediction and model selection.

- 4.1. Integrated models. I(d).
- 4.2. Seasonal models. SARIMA(p,d,q)(P,D,Q).
- 4.3.Forecasting with time series
- 4.4.Model selection.

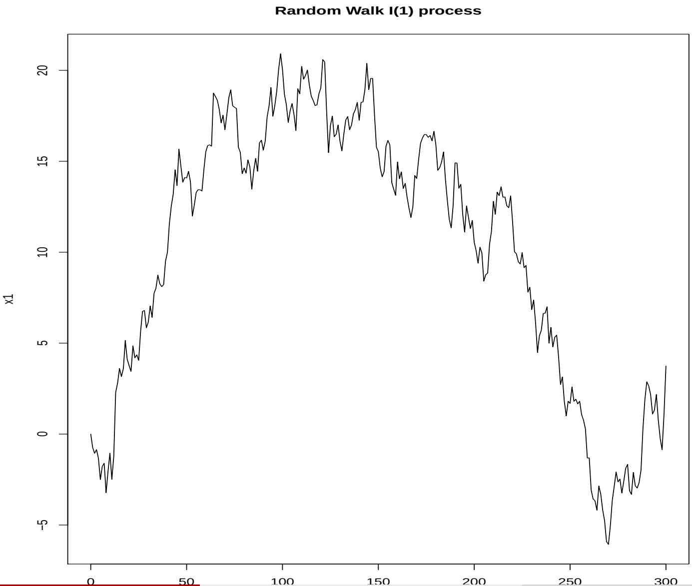
Random walk. I(1)

A random walk process can be expresed as:

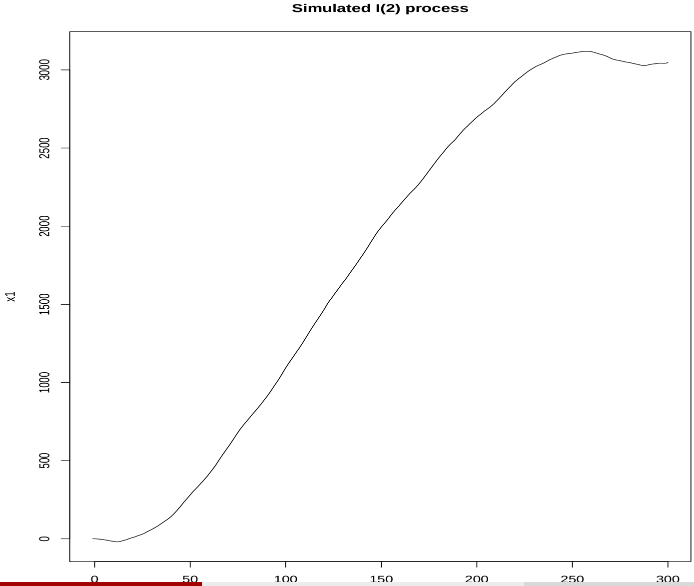
$$X_t = X_{t-1} + a_t$$

- Unconditional mean not constant. Long memory
- Unconditional variance not constant.
- ACF slow decaying pattern.
- PACF one significant peak at $k = 1$.

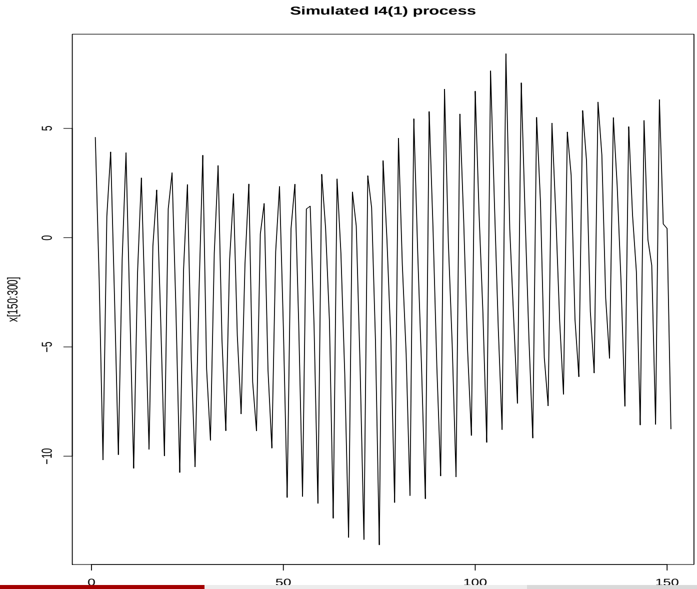
Random Walk process I(1).



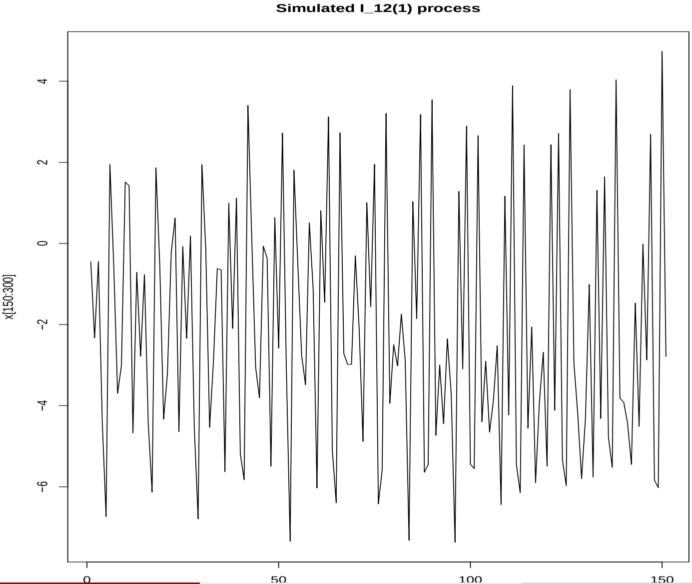
Cuadratic trend I(2).



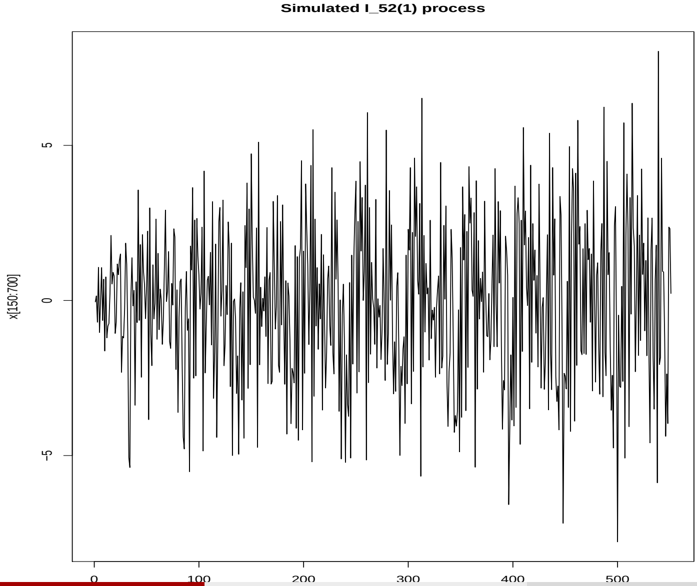
Seasonality $I_s(1)$.



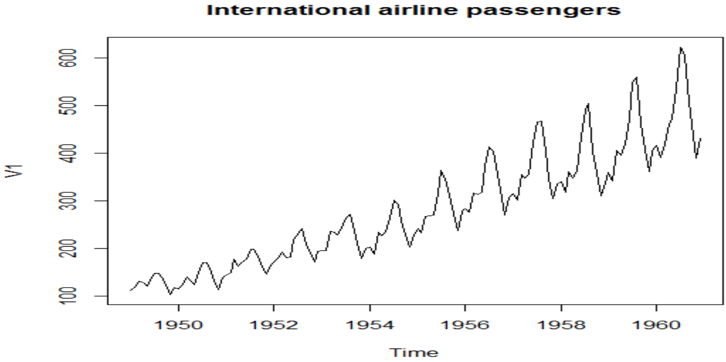
Seasonality $I_s(1)$.



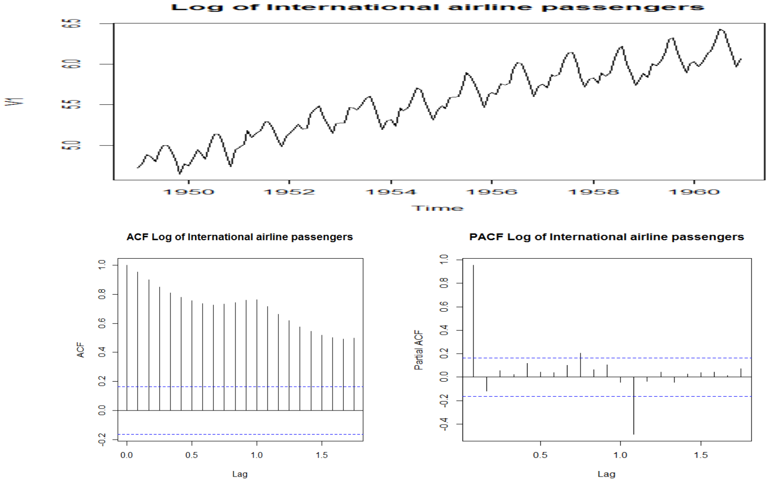
Seasonality $I_s(1)$.



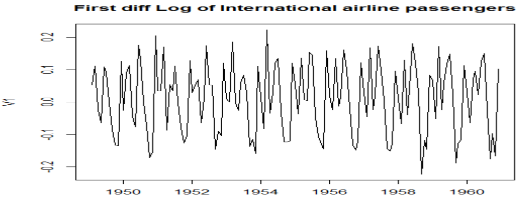
Airline passengers example.

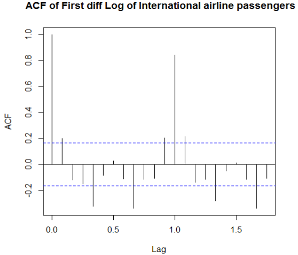


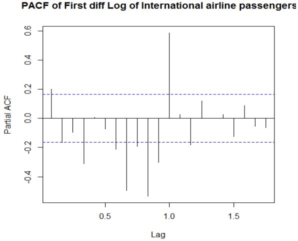
Airline passengers example.



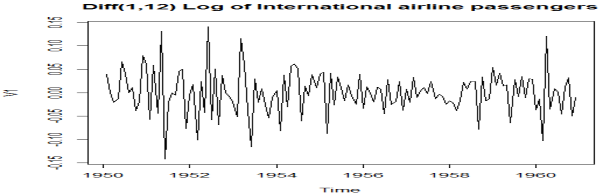
Airline passengers example.

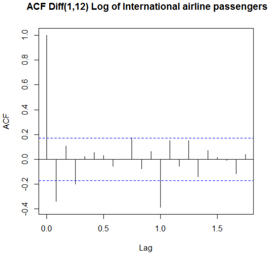


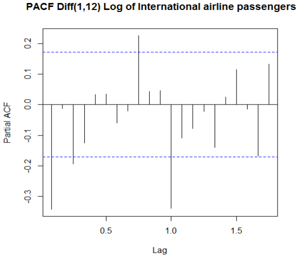




Airline passengers example.







Stylized Facts in ACF and PACF

Regular and seasonal differenced data show typical behavior in ACF/PACF: there is usually some significant short term dependence over the first couple of lags, as well as significant autocorrelation at multiples of the period S, the frequency of observations.

- This suggests that large p, q are required for describing the data with ARMA models, making them non-parsimonious.
- We may overcome this problem by using the *Airline model*:

$$\nabla \nabla_{12} X_t = (1 + \theta_1 B)(1 + \theta_{12} B^{12}) a_t$$

This is a $SARIMA(0,1,1)(0,1,1)^{12}$.

SARIMA (p,d,q)(P,D,Q) models.

A series X_t follows a SARIMA (p,d,q,)(P,D,Q) process if the following equation holds:

$$\Phi(B)\Phi_S(B^S)\nabla^d\nabla_S^D X_t = (1+\theta_1 B)(1+\theta_S B^S)a_t$$

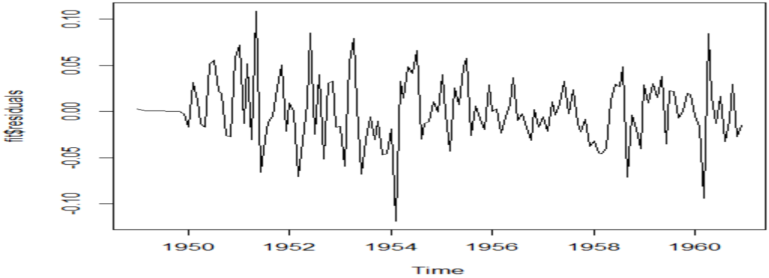
In most practical cases, using a differencing order $d = D = 1$ will be sufficient. Choosing of p, q, P, Q happens via ACF/PACF or via aic-based parsimonious decisions.

Airline passengers example.

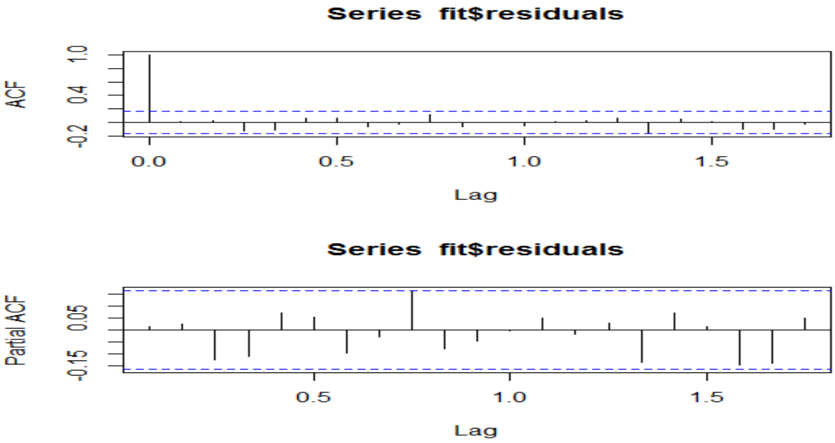
```
> fit<-auto.arima(log(passengers),ic="aic")
> fit
Series: log(passengers)
ARIMA(0,1,1)(0,1,1)[12]

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
> ts.plot(fit$residuals)
```



Airline passengers example.



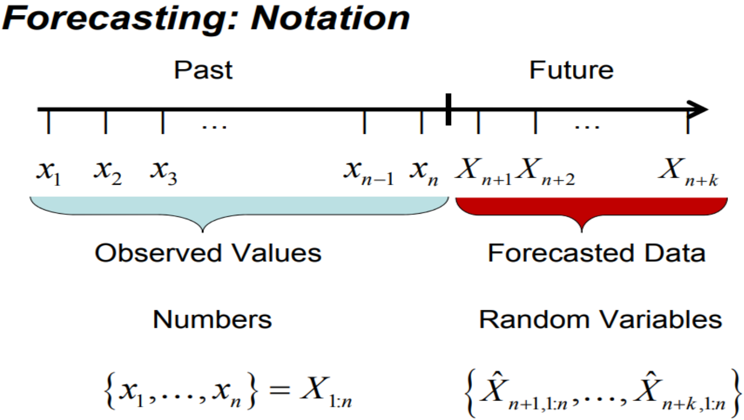
Forecasting with time series.

Goal: Point predictions for future observations with a measure of uncertainty, i.e. a 95 % prediction interval.

Note:

- will be based on a stochastic model.
- builds on the dependency structure and past data.
- is an extrapolation and, therefore, we should be cautious.

Forecasting notation.



Sources of uncertainty in forecasting.

There are four principal sources of uncertainty:

- Does the data generating model from the past also apply in the future? Or are there any breaks?
- Is the ARMA(p,q) model we fitted to the data correctly chosen? What is the true order?
- Are the parameters accurately estimated?
- The stochastic variability coming from the innovations a_t.

How to forecast?

1. Probabilistic principle for deriving point forecast:

$$\hat{X}_{n+k/1:n} = E[X_{n+k}|x_1, \ldots, x_n]$$

$\longrightarrow$ The point forecast will be based on the conditional mean.

2. Probabilistic principle for deriving prediction intervals:

$$\sigma^2_{X_{n+k}-\hat{X}_{n+k}} = Var(X_{n+k} - \hat{X}_{n+k}|x_1, \ldots, x_n)$$

$\longrightarrow$ An approximation 95 % prediction interval will be obtained via :

$$\hat{X}_{n+k/1:n} \pm 1{,}96\sigma_{X_{n+k}\hat{X}_{n+k}}$$

Forecasting white noise with constant.

Suppose the white noise plus constant process:

$$X_t = c + a_t$$

The one-step prediction will be:

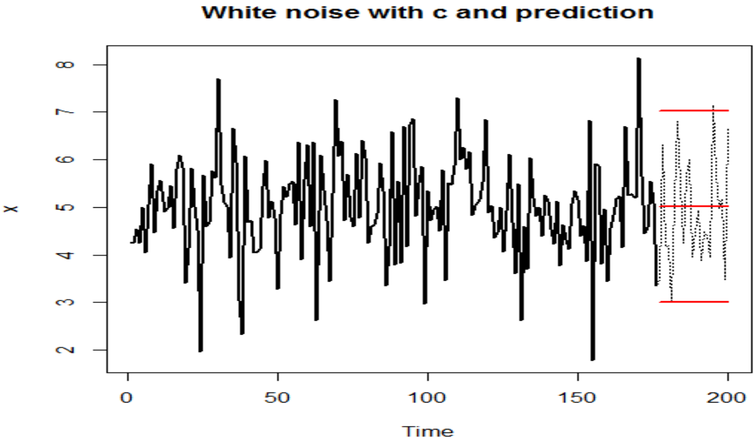
$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \dots, x_n] = c$$

The k-step prediction will be:

$$\hat{X}_{n+k/1:n} = E[X_{n+1}|x_1, \dots, x_n] = c$$

And,therefore, assumming a correct estimation of c all the uncertainty in the prediction is given by σ_a^2.

Forecasting white noise with constant.



Forecasting random walk.

Suppose the random walk I(1) process:

$$X_t = X_{t-1} + a_t$$

The one-step prediction will be:

$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = x_n$$

The k-step prediction will be:

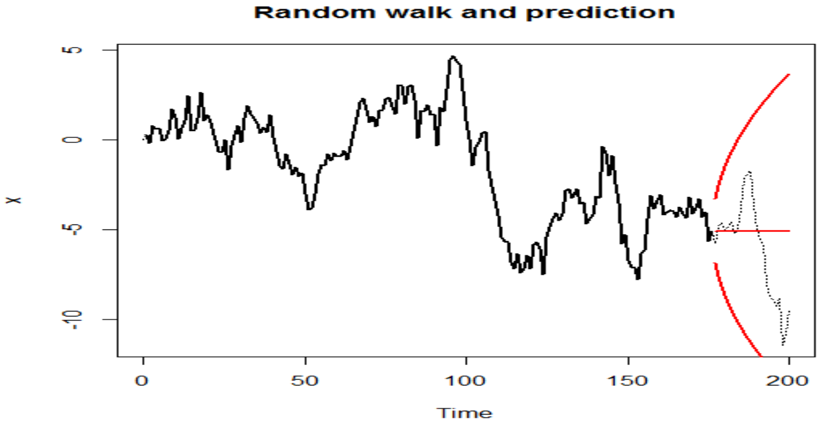
$$\hat{X}_{n+k/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = x_n$$

and error variance equal to:

$$\sigma^2_{X_{n+k}-\hat{X}_{n+k}} = Var(X_{n+k} - \hat{X}_{n+k}|x_1, \ldots, x_n) = (k-1) \cdot \sigma^2_a$$

The long term prediction is equal to a constant equal to the last observation.The variance does not converge and increases with the horizon k.

Forecasting random walk.



Forecasting random walk plus drift.

Suppose now the random walk I(1) plus drift process:

$$X_t = X_{t-1} + c + a_t$$

The one-step prediction will be:

$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = x_n + c$$

The k-step prediction will be:

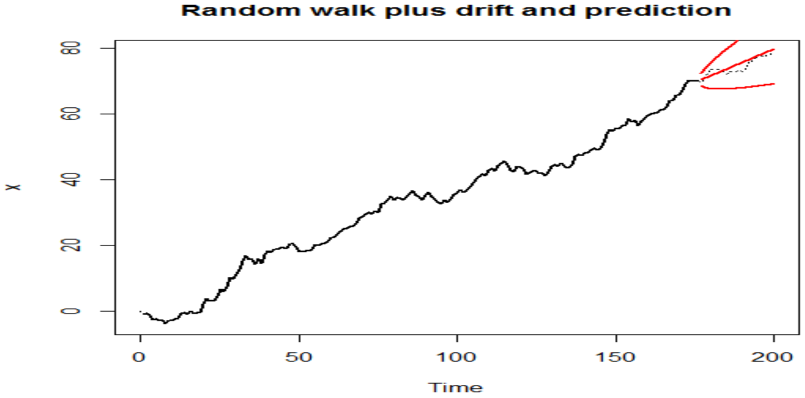
$$\hat{X}_{n+k/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = x_n + k \cdot c$$

and variance equal to:

$$\sigma^2_{X_{n+k} - \hat{X}_{n+k}} = Var(X_{n+k} - \hat{X}_{n+k}|x_1, \ldots, x_n) = (k-1) \cdot \sigma^2_a$$

The long term prediction is equal to a linear trend with origin in the last observation and slope c.The variance does not converge and increases with the horizon k.

Forecasting random walk plus drift.



Forecasting an AR(1).

Suppose the AR(1) process:

$$X_t = \phi_1 X_{t-1} + a_t$$

The one-step prediction will be:

$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = \phi_1 x_n$$

The k-step prediction will be:

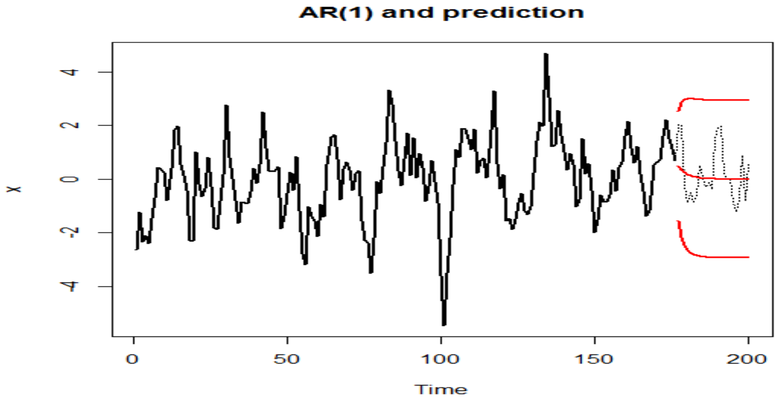
$$\hat{X}_{n+k/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = \phi_1 \hat{X}_{n+k-1} = \phi_1^k x_n$$

and the variance equal to:

$$\sigma^2_{X_{n+k}-\hat{X}_{n+k}} = Var(X_{n+k} - \hat{X}_{n+k}|x_1, \ldots, x_n) = \left(1 + \sum_{j=1}^{k-1} \phi_1^{2j}\right) \sigma_a^2$$

The long term prediction converges to the global mean at a ratio depending on the size of ϕ_1. The variance converges to the variance of the process.

Forecasting an AR(1).



Forecasting an AR(p).

Suppose the AR(p) process:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t$$

The one-step prediction will be:

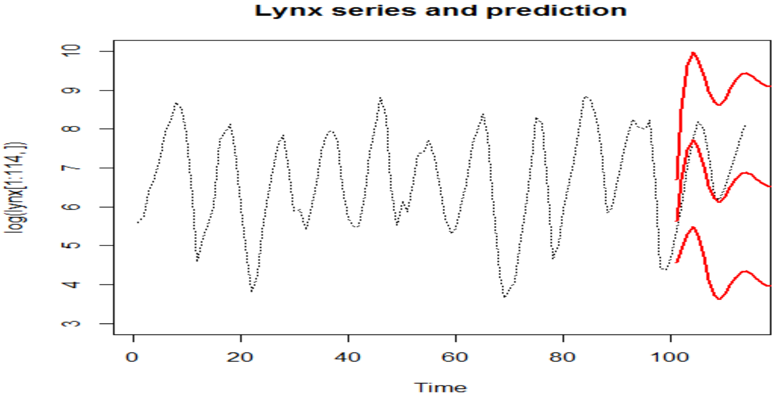
$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = \phi_1 x_n + \ldots + \phi_p x_{n+1-p}$$

The k-step prediction will be:

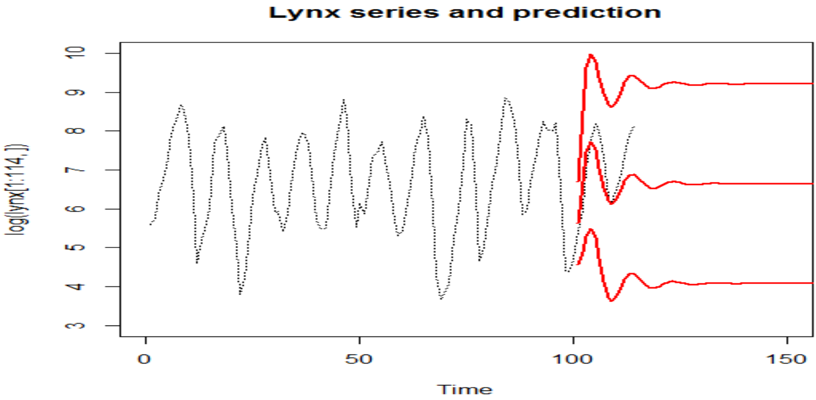
$$\hat{X}_{n+k/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = \phi_1 \hat{X}_{n+k-1} + \ldots + \phi_p \hat{X}_{n+k-p}$$

If an observed value for $\hat{X}_{n+k-t}$ is available, we plug it in. else, the forecasted value is used. Hence, the forecast for horizons $k < 1$ are determined recursively.

Forecasting the lynx series.



Forecasting the lynx series.



Forecasting an MA(1).

Suppose the invertible MA(1) process:

$$X_t = a_t + \theta_1 a_{t-1}$$

The k-step prediction will be:

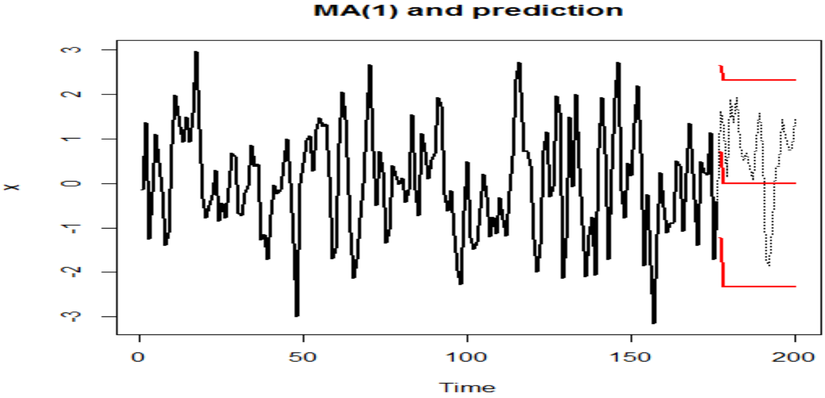
$$\hat{X}_{n+k/1:n} = E[X_{n+k}|x_1, \ldots, x_n] = E[a_{n+k} + \theta_1 a_{n+k-1}|x_1, \ldots, x_n] = 0$$

The best forecast for horizons 2 and up is zero. The one-step forecast is more problematic:

$$\hat{X}_{n+1/1:n} = E[X_{n+1}|x_1, \ldots, x_n] = \theta_1 E[a_n|x_1, \ldots, x_n] = \theta_1 \sum_{j=0}^{n-1} \theta_1^j x_{n-j}$$

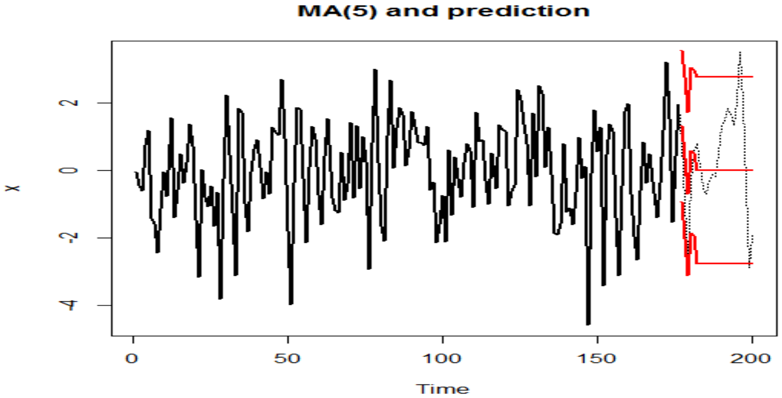
where we have made use of the AR(∞) representation of an MA(1). Therefore, the one-step ahead forecast of a MA(1) is the sum of all observed values, with exponentially decaying weights.

Forecasting an MA(1).



Forecasting an MA(q).

When forecasting from MA(q) processes, we encounter the same difficulties as above. The predictions for horizons exceeding q are all zero, but anything bellow contains terms that are a combination of all observations.



Forecasting from ARMA(p,q).

Suppose the stationary and invertible ARMA(p,q) process:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$$

The one-step prediction will be:

$$\begin{aligned}
\hat{X}_{n+1/1:n} &= E[X_{n+1}|x_1, \ldots, x_n] \\
&= \sum_{i=1}^{p} \phi_i E[X_{n+1-i}|x_1^n,] + E[a_{n+1}|x_1^n] + \sum_{j=1}^{q} \theta_j E[a_{n+1-j}|x_1^n] \\
&= \sum_{i=1}^{p} \phi_i x_{n+1-i} + \sum_{i=1}^{q} \theta_j E[a_{n+1-j}|x_1^n]
\end{aligned}$$

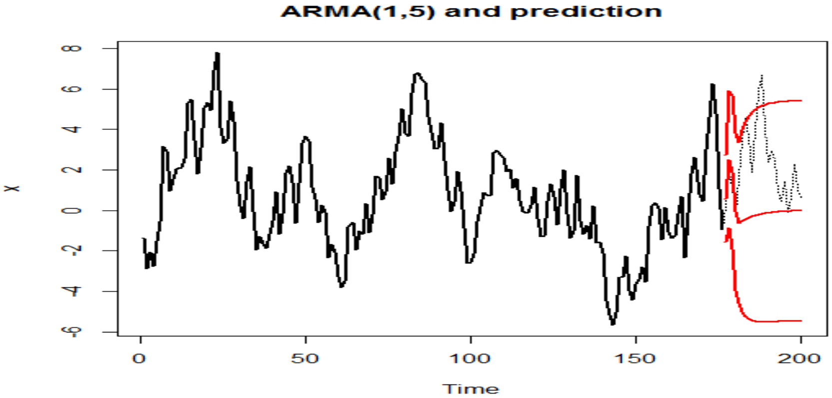
where $E[a_{n+1-j}|x_1^n]$ can be obtained using the AR(∞) representation.

Forecasting from ARMA(p,q).

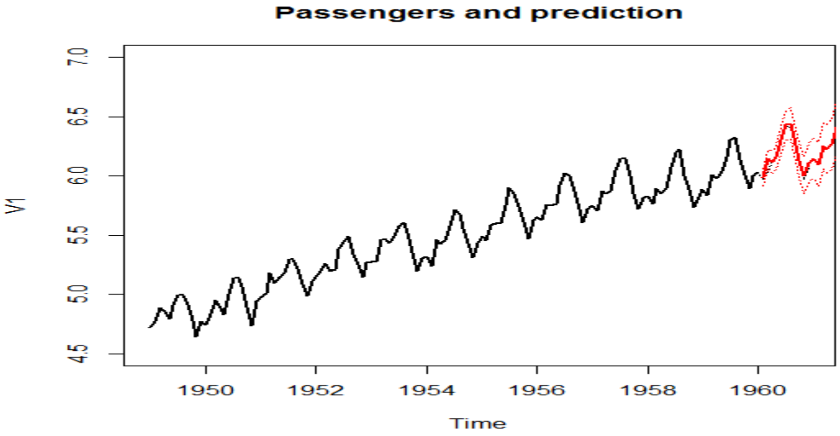
The k-step prediction will be:

$$
\begin{aligned}
\hat{X}_{n+k/1:n} &= E[X_{n+k}|x_1,\ldots,x_n] \\
&= \sum_{i=1}^{p} \phi_i E[X_{n+k-i}|x_1^n,] + \sum_{j=1}^{q} \theta_j E[a_{n+k-j}|x_1^n]
\end{aligned}
$$

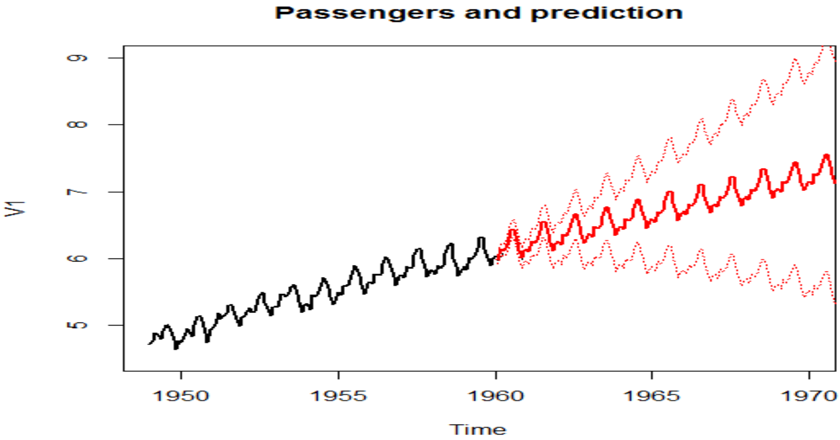
Forecasting from ARMA(p,q).



Forecasting from Airline model.



Forecasting from Airline model.



Model selection.

- Goal: We want to select the order of an AR(p) model in such a way that the one-step prediction mean square error is minimized.
- Assume the AR(p) model given by:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t$$

  where a_t are gaussian white noise with finite variance σ_a^2.
- The conditional one-step prediction mean square error can then be written as:

$$e_{n+1} = E[(X_{n+1} - \hat{X}_{n+1})^2 | x_1^n]$$

  and, under the Gaussian assumption, one can show that

$$\hat{e}_{n+1} = \hat{\sigma}^2(1 + pn^{-1})$$

Model selection.

- Inserting this, we have an estimation of the one-step forecast error. If we want to minimize this value, it implies that the order p must be chosen by minimizing the Final Prediction Error (FPE)

$$FPE = \frac{\hat{\sigma}^2(n+p)}{n-p}$$

  The FPE combines fitting with parsimony, due to the penalty introduced by the term $(n+p)(n-p)$.

- An equivalent form of this criterion is:

$$log(FPE) = log\hat{\sigma}^2 + logn(1+p/n) - logn(1-p/n) \approx log\hat{\sigma}^2 + 2p/n$$

Model selection.

- Multiplying for n, we obtain the Akaike Information Criteria:

$$AIC = nlog\hat{\sigma}^2 + 2p$$

- or the Bayesian Information Criteria, with greater penalty:

$$BIC = nlog\hat{\sigma}^2 + (logn)p$$

- BIC tends to select simpler models.