Analysis of Time Series, L6

Rolf Larsson

Uppsala University

4 april 2025

Today

3.5: Forecasting

- The best linear prediction
- The Durbin-Levinson algorithm
- ARMA processes

The best linear prediction

AR(p):

- We observe $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t$ for $t = 1, 2, ..., n$.
- Assume that $\phi_1, \phi_2, ..., \phi_p$ are *known*.
- Predict x_{n+1}.
- A natural predictor is

$$\phi_1 x_n + \phi_2 x_{n-1} + ... + \phi_p x_{n-p+1}.$$

- Is it optimal in some sense?

The best linear prediction

- Assume that $\{x_t\}$ is stationary with known model parameters.
- We observe $x_1, x_2, ..., x_n$.
- Predict x_{n+m} by the linear predictor

$$\begin{aligned} x_{n+m}^n &= \alpha_0 + \alpha_n x_n + \alpha_{n-1} x_{n-1} + ... + \alpha_1 x_1 \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k . \end{aligned}$$

- Mean square error $E\{(x_{n+m} - x_{n+m}^n)^2\}$.

Theorem (Property 3.3)

The linear predictor that minimizes the mean square error is found by solving

$$E\{(x_{n+m} - x_{n+m}^n)x_k\} = 0, \quad k = 0, 1, ..., n$$

where $x_0 = 1$*, for* $\alpha_0, \alpha_1, ..., \alpha_n$.

The best linear prediction

- Assume WLOG that $E(x_t) = \mu = 0$, i.e. $\alpha_0 = 0$.
- Find the optimal one-step predictor

$$x_{n+1}^n = \phi_{n1}x_n + \phi_{n2}x_{n-1} + \ldots + \phi_{nn}x_1.$$

- Solve (why?)

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix}$$

- i.e. $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$.

The best linear prediction

- $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$.
- For ARMA models, it can be shown that Γ_n is positive definite.
- Hence, $\boldsymbol{\phi}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n$, i.e.

$$\begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \vdots \\ \phi_{nn} \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}^{-1} \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix}$$

- Derive the one-step predictors for zero mean causal AR(1) and AR(2) processes with one and two observations, respectively.

The best linear prediction

- Problem 3.45: For a zero mean causal AR(p),

$$x_{n+1}^n = \phi_1 x_n + \phi_2 x_{n-1} + \ldots + \phi_p x_{n+1-p}.$$

- In general, for a zero mean causal process,

$$\begin{aligned} x_{n+1}^n &= \phi_{n1} x_n + \phi_{n2} x_{n-1} + \ldots + \phi_{nn} x_1 \\ &= \begin{pmatrix} \phi_{n1} & \phi_{n2} & \ldots & \phi_{nn} \end{pmatrix} \begin{pmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{pmatrix} = \boldsymbol{\phi}_n' \boldsymbol{x}. \end{aligned}$$

- Mean square prediction error (see the book, p.102)

$$P_{n+1}^n = E\{(x_{n+1} - x_{n+1}^n)^2\} = \gamma(0) - \boldsymbol{\gamma}_n' \Gamma_n^{-1} \boldsymbol{\gamma}_n.$$

The Durbin-Levinson algorithm

- We have derived $x_{n+1}^n = \phi_n' \boldsymbol{x} = \gamma_n' \Gamma_n^{-1} \boldsymbol{x}$.
- Is it possible to calculate x_{n+1}^n without inverting Γ_n?

Theorem (Property 3.4, Problem 3.13)

ϕ_n and P_{n+1}^n may be found iteratively as follows:

$$\phi_{00} = 0, \quad P_1^0 = \gamma_0.$$

For $n \geq 1$,

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}, \quad P_{n+1}^n = P_n^{n-1}(1 - \phi_{nn}^2),$$

where for $n \geq 2$,

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn}\phi_{n-1,n-k}, \quad k = 1, 2, ..., n-1.$$

The Durbin-Levinson algorithm

Theorem (Property 3.5, Problem 3.13)

The PACF of a stationary process $\{x_t\}$ is given by ϕ_{nn}.

Example 1:
Use the Durbin-Levinson algorithm to obtain the PACF of an AR(2) process!

The Durbin-Levinson algorithm

- Generalization to prediction more than one step is straightforward, see the book.
- For predicting MA processes it is easier to use *the innovations algorithm* (property 3.6).
- Read about it by yourself!

ARMA processes

- Recall: For a zero mean causal AR(p), the optimal one-step predictor is
$$x_{n+1}^{n} = \phi_1 x_n + \phi_2 x_{n-1} + \ldots + \phi_p x_{n+1-p}.$$
- Idea: For causal ARMA processes (including MA), why not just use the AR representation $x_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \ldots = w_t$, i.e.
$$x_t = w_t - \pi_1 x_{t-1} - \pi_2 x_{t-2} - \ldots$$
for prediction?
- Problem: This requires an infinite past!
- *But* for large samples, by truncation it can be used as an approximation.
- In particular, it can also be used as an approximation for pure MA processes.

ARMA processes

One step ahead, forecast:

- Replace the minimum mean square error predictor

$$x_{n+1}^n = E(x_{n+1}|x_n, ..., x_1)$$

  by

$$\tilde{x}_{n+1} = E(x_{n+1}|x_n, ..., x_1, x_0, x_{-1}, ...).$$

- The AR form

$$x_{n+1} = w_{n+1} - \pi_1 x_n - \pi_2 x_{n-1} - ... = w_{n+1} - \sum_{j=1}^{\infty} \pi_j x_{n+1-j}$$

- implies (take $E(\cdot|x_n, ..., x_1, x_0, x_{-1}, ...)$)

$$\tilde{x}_{n+1} = -\pi_1 x_n - \pi_2 x_{n-1} - ... = -\sum_{j=1}^{\infty} \pi_j x_{n+1-j}.$$

ARMA processes

One step ahead, prediction error:

- For invertible processes, the MA representation

$$x_{n+1} = w_{n+1} + \psi_1 w_n + \psi_2 w_{n-1} + \ldots = w_{n+1} + \sum_{j=1}^{\infty} \psi_j w_{n+1-j}$$

- implies (take $E(\cdot | x_n, \ldots, x_1, x_0, x_{-1}, \ldots)$)

$$\tilde{x}_{n+1} = \psi_1 w_n + \psi_2 w_{n-1} + \ldots = \sum_{j=1}^{\infty} \psi_j w_{n+1-j}.$$

- Hence, $x_{n+1} - \tilde{x}_{n+1} = w_{n+1}$, and the mean square prediction error is

$$P_{n+1}^n = E\{(x_{n+1} - \tilde{x}_{n+1})^2\} = E(w_{n+1}^2) = \sigma_w^2.$$

ARMA processes

$m \geq 2$ steps ahead, forecast:

- Calculate

$$\tilde{x}_{n+m} = E(x_{n+m} | x_n, \ldots, x_1, x_0, x_{-1}, \ldots).$$

- The AR form

$$\begin{aligned} x_{n+m} &= w_{n+m} - \pi_1 x_{n+m-1} - \pi_2 x_{n+m-2} - \ldots \\ &= w_{n+m} - \sum_{j=1}^{m-1} \pi_j x_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j} \end{aligned}$$

- implies (take $E(\cdot | x_n, \ldots, x_1, x_0, x_{-1}, \ldots)$)

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}.$$

ARMA processes

$m \geq 2$ steps ahead, prediction error:

- For invertible processes, the MA representation ($\psi_0 = 1$)

$$\begin{aligned} x_{n+m} &= w_{n+m} + \psi_1 w_{n+m-1} + \psi_2 w_{n+m-2} + \dots \\ &= \sum_{j=0}^{m-1} \psi_j w_{n+m-j} + \sum_{j=m}^{\infty} \psi_j w_{n+m-j} \end{aligned}$$

- $E(w_{n+m-j} | x_n, \dots, x_1, x_0, x_{-1}, \dots) = 0$ if $m > j$ and w_{n+m-j} if $m \leq j$
- implies $\tilde{x}_{n+m} = \sum_{j=m}^{\infty} \psi_j w_{n+m-j}$.
- Hence, the mean square prediction error (MSPE) is

$$\begin{aligned} P_{n+m}^n &= E\{(x_{n+m} - \tilde{x}_{n+m})^2\} \\ &= E\left\{ \left(\sum_{j=0}^{m-1} \psi_j w_{n+m-j} \right)^2 \right\} = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2. \end{aligned}$$

ARMA processes

To summarize:

- Forecast

$$\tilde{x}_{n+m} = -\sum_{j=1}^{m-1} \pi_j \tilde{x}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j x_{n+m-j}$$

- MSPE

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2$$

- Prediction interval

$$\tilde{x}_{n+m} \pm c_{\alpha/2} \sqrt{P_{n+m}^n}$$

- For normal (Gaussian) processes with $\alpha = 0.05$, $c_{\alpha/2} = 1.96$.

ARMA processes

Example 2: Let

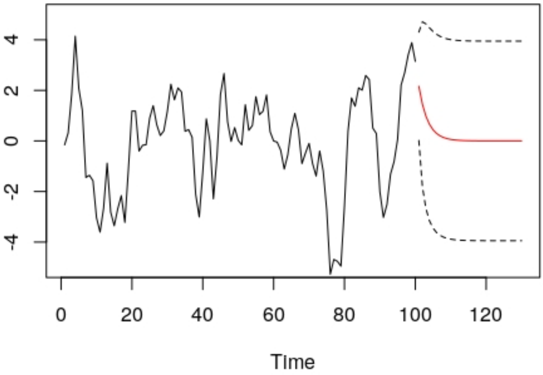
$$x_t = 0.8x_{t-1} + w_t + 0.4w_{t-1}.$$

We observe $x_1, x_2, ..., x_n$, where $n = 100$, $x_{95} = 0.06$, $x_{96} = 2.21$, $x_{97} = 2.70$, $x_{98} = 3.40$, $x_{99} = 3.89$, $x_{100} = 3.15$.

1. Calculate the forecast $\tilde{x}_{n+m}$ for $m = 1, 2, 3$.
2. Calculate the MSPE.
3. Calculate the $1 - \alpha = 0.95$ prediction interval assuming that $\{w_t\}$ is a normal process with $\sigma_w^2 = 1$.

ARMA processes

Example 2: Plot of x_t with forecast horizon 30 and pointwise two standard deviations prediction intervals (obtained from estimated parameters).



In R:

```
> x=arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.4),100)
> m=arima(x,order=c(1,0,1),include.mean=FALSE)
> m

Call:
arima(x = x, order = c(1, 0, 1), include.mean = FALSE)

Coefficients:
         ar1     ma1
      0.6721  0.4730
s.e.  0.0838  0.0969

sigma^2 estimated as 1.148:  log likelihood = -149.51,  aic = 305.02

> fore=predict(m,n.ahead=30)
> ts.plot(x,fore$pred,col=1:2,ylim=c(-5,5))
> lines(fore$pred+2*fore$se,lty='dashed')
> lines(fore$pred-2*fore$se,lty='dashed')
```

News of today

- Optimal prediction of AR processes
  - by matrix inversion
  - by iteration
- Optimal prediction of ARMA processes with infinite past (useful as approximations even if not infinite past).
- Note that all of this required *known* parameters.