1. Estimation of Model Parameters

For completeness, some portions are repeated from the previous lecture

1.1 AR models

If the average $\overline{x}$ is subtracted from observed data,

N observations: $X_1, X_2, \ldots, X_N$

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_n X_{t-n} + a_t$$

$$Y = \begin{bmatrix} X_{n+1} \\ X_{n+2} \\ \cdot \\ \cdot \\ X_N \end{bmatrix} \quad X = \begin{bmatrix} X_n & X_{n-1} & \cdots & X_1 \\ X_{n+1} & X_n & \cdots & X_2 \\ \cdots & & & \\ \cdots & & & \\ X_{N-1} & X_{N-2} & \cdots & X_{N-n} \end{bmatrix} \quad \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \cdot \\ \cdot \\ \phi_n \end{bmatrix}$$

$$\hat{\phi} = (X'X)^{-1}X'Y$$

$$\hat{\sigma}_a^2 = \frac{1}{(N-n)} \sum_{t=n+1}^{N} (X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} - \ldots - \hat{\phi}_n X_{t-n})^2$$

linear least squares estimation methods

1.2 ARMA models

It requires nonlinear least squares method.

The general procedure is as follows:

(i) Starts with some initial or "guess" values of the parameters ϕ_i and θ_i;

(ii) Search new values in the parameter space $\{\phi_1, \phi_2, \ldots, \phi_n, \theta_1, \theta_2, \ldots, \theta_m\}$ so that a direction of smaller sum of squares of a_t's is obtained.

(iii) Use the obtained values as new initial values of the parameters and repeat step (ii) until certain criteria can be met. (e.g.,relative reduction in sum of squares, or the maximum change in the parameter values is below a given level)

Many nonlinear search algorithms are available in the form of standard library routines.

In the final estimation, the mean μ should also be estimated by

$$(\dot{X}_t - \mu) - \phi_1(\dot{X}_{t-1} - \mu) - \phi_2(\dot{X}_{t-2} - \mu) - \ldots - \phi_n(\dot{X}_{t-n} - \mu) =$$
$$a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_m a_{t-m}$$

n+m + 1 parameters and the initial guess value for μ is the average $\overline{x}$.

1.3 Initial guess values by inverse function

ARMA(n,m) model:

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_n B^n)\, X_t = (1-\theta_1 B - \ldots - \theta_m B^m) a_t$$

$$a_t = (1 - I_1 B - I_2 B^2 - \ldots)\, X_t$$

$$(1-\phi_1 B - \phi_2 B^2 - \ldots - \phi_n B^n) = (1-\theta_1 B - \ldots - \theta_m B^m)\,(1 - I_1 B - I_2 B^2 - \ldots)$$

By matching the coefficients of equal powers of B,

$$\phi_1 = \theta_1 + I_1$$

$$\phi_2 = \theta_2 - \theta_1 I_1 + I_2$$

$$\phi_3 = \theta_3 - \theta_1 I_2 - \theta_2 I_1 + I_3$$

$$\ldots$$

$$\phi_j = \theta_j - \theta_1 I_{j-1} - \theta_2 I_{j-2} - \ldots - \theta_{j-1} I_1 + I_j$$

and

$$(1-\theta_1 B - \ldots - \theta_m B^m)\, I_j = 0 \qquad \text{for } j > \max(n, m)$$

<u>Remarks:</u>

- The relations between ϕ_i and θ_i expressed by the inverse function I_i are linear.
- The I_i's are autoregressive parameters of the infinite expansion of an ARMA model and thus can be estimated by the linear least squares methods. AR(∞)
- If the estimates of the inverse function I_i's are known, the initial values for the autoregressive as well as the moving average parameters can be estimated.

<u>Estimates of I_i's</u>:

Since for an AR(p) model,

$$a_t = (1 - \phi_1 B - \ldots - \phi_p B^p)\, X_t$$

$$I_j = \phi_j \quad j=1,2,\ldots,p \quad I_j=0 \text{ for } j>p$$

good estimates of I_i's for an ARMA(n,m) model can be obtained from ϕ_i's of an AR(p) model for sufficiently large p.

$$p = \max(n, m) + m$$

m initial values of θ_i can be estimated from m equations with j>max(n,m).

Steps:

i) Get I_i's from the data by fitting an AR(p) model

ii) Estimate ϕ_i and θ_i from these estimated I_i's.

Sometimes, the initial value of $\underline{\theta_i}$ does not satisfy the invertibility conditions then take its reciprocal.

Example:

Sunspot activity data: ARMA(2,1)

$p = \max(n,m) + m = 2 + 1 = 3$

AR(3) is fitted using linear least squares method:

$I_1 = \phi_1 = 1.27$

$I_2 = \phi_2 = -0.5$

$I_3 = \phi_3 = -0.11$ RSS = 40843

$(1-\theta_1 B)\, I_j = 0 \quad j > 2$

$I_3 - \theta_1 I_2 = 0 \qquad \theta_1 = \dfrac{I_3}{I_2} = 0.22$

$\phi_1 = I_1 + \theta_1 = 1.49$

$\phi_2 = \theta_2 - \theta_1 I_1 + I_2 = 0 - 0.22\text{x}1.27 - 0.5 = -0.78$

Chapter 5. Forecasting

1. Geometric interpretation of prediction as orthogonal projection

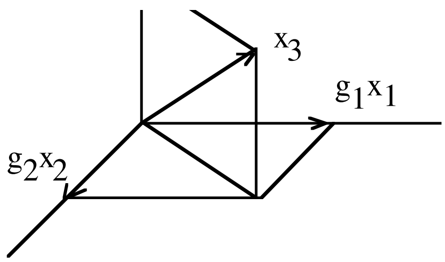
(1) Simple orthogonal prediction

Given two vectors x_1 and x_2, approximate a 3rd vector x_3 .

Find the vector

$$\hat{x}_3 = g_1 x_1 + g_2 x_2$$

so that it is closest to the vector x_3.



The orthogonal projection of x_3 on a plane formed by the vectors x_1 and x_2, gives the vector $\hat{x}_3 = g_1 x_1 + g_2 x_2$ that is closest to x_3.

The prediction error is also orthogonal to the vectors x_1 and x_2.

(2) Prediction of a stationary stochastic time-series

Given observations at time t:

Predict the value of X_{t+l} (l step ahead prediction.

The best linear prediction of X_{t+l} at time t will be given by a linear combination

$$\hat{X}_t(l) = g_0^* X_t + g_1^* X_{t-1} + g_2^* X_{t-2} + \dots$$

which minimizes the mean square of the prediction error

$$e_t(l) = X_{t+l} - \hat{X}_t(l)$$

$$E[e_t(l)]^2 = E[X_{t+l} - \hat{X}_t(l)]^2$$

is given by the orthogonal projection of X_{t+l} on the "plane" formed by the linear combination of

$$X_t, X_{t-1}, X_{t-2}, \dots.$$

Problems become how to find all g*'s.

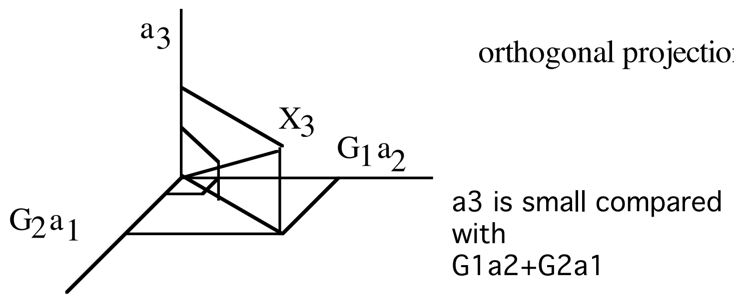
But if the vectors are not necessarily orthogonal, the problems may seem not quite apparent.

If the coordinates of X_3 in an orthogonal frame of reference

$$X_3 = a_3 + G_1 a_2 + G_2 a_1$$

then, the orthogonal projection on a plane formed by a_1 and a_2 will be

$$X_3 = G_1 a_2 + G_2 a_1$$



2. Prediction by conditional expectation

Rules:

The conditional expectation at time t

$$\underset{t}{E}(X_{t-j}) = X_{t-j} \qquad j=0,1,2,\ldots$$

$$\underset{t}{E}(X_{t+j}) = \hat{X}_t(j) \qquad j=1,2,3,\ldots$$

$$\underset{t}{E}(a_{t-j}) = a_{t-j} \qquad j=0,1,2,\ldots$$

$$\underset{t}{E}(a_{t+j}) = 0 \qquad j=1,2,3,\ldots$$

Example of an AR(1) model:

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

$$\hat{X}_t(1) = \underset{t}{E}(X_{t+1}) = \underset{t}{E}(\phi_1 X_t + a_{t+1}) = \phi_1 X_t$$

$$\hat{X}_t(2) = \underset{t}{E}(X_{t+2}) = \underset{t}{E}(\phi_1 X_{t+1} + a_{t+2}) = \phi_1 \hat{X}_t(1) = \phi_1^2 X_t$$

$$\hat{X}_t(l) = \underset{t}{E}(X_{t+l}) = \phi_1 \hat{X}_t(l-1) = \phi_1^l X_t$$

3. Eventual forecasts and stability

For large enough l,

$\hat{X}_t(l) = \phi_1 \hat{X}_t(l-1) + \ldots + \phi_n \hat{X}_t(l-n) \quad l \geq \max(n+1, m+1)$

it is the autoregressive parameters that control the forecasts.

Asymptotically stable system, the eventual forecast tends to be zero, which is the mean.

e.g.

ARMA(1, 1) with $\phi_1 = 0.5$

$\hat{x}_t(1) = 0.5\, x_t - \theta_1 a_t$

$\hat{x}_t(l) = 0.5\, \hat{x}_t(l-1)$ for $l \geq 2$

$= 0.5^{l-1}\, \hat{x}_t(1) \to 0$ as $l \to \infty$

Stable system (at least one of the roots is one in absolute value), the eventual forecast will be constant.

Unstable system, the eventual forecast will be infinite.