ARMAV based prediction
(or forecasting by leading indicator – but it's a misnomer)

In many cases, the forecasting of a series of interest can be improved by using information from a related series -- leading indicator

For example,

$$X_{2t} = \phi_{211} X_{1t-1} + \phi_{221} X_{2t-1} + a_{2t}$$

Using conditional expectation,

$$\hat{X}_{2t}(1) = \phi_{211} X_{1t} + \phi_{221} X_{2t}$$

$$\hat{X}_{2t}(k) = \phi_{211} \hat{X}_{1t}(k-1) + \phi_{221} \hat{X}_{2t}(k-1), \quad k \geq 2$$

case (i) If X_{1t} is a white noise series,

$$X_{1t} = a_{1t}$$

then

$$\hat{X}_{1t}(k) = 0 \quad k \geq 1$$

$$\hat{X}_{2t}(1) = \phi_{211} X_{1t} + \phi_{221} X_{2t}$$

$$\hat{X}_{2t}(k) = \phi_{221} \hat{X}_{2t}(k-1), \quad k \geq 2$$

In this case, the leading indicator does not contribute to the forecast of X_{2t} .

case (ii) If X_{1t} is another autoregressive model

$$X_{1t} = \phi_{111} X_{1t-1} + \phi_{121} X_{2t-1} + a_{1t}$$

$$\hat{X}_{1t}(k-1) = \phi_{111} \hat{X}_{1t}(k-2) + \phi_{121} \hat{X}_{2t}(k-2)$$

$$= \phi_{111} [\phi_{111} \hat{X}_{1t}(k-3) + \phi_{121} \hat{X}_{2t}(k-3)] + \phi_{121} \hat{X}_{2t}(k-2)$$

$$= \ldots$$

$$= \phi_{111}^{k-1} X_{1t} + \phi_{121} [\hat{X}_{2t}(k-2) + \phi_{111} \hat{X}_{2t}(k-3) + \ldots + \phi_{111}^{k-2} X_{2t}]$$

Thus,

$$\hat{X}_{2t}(1) = \phi_{211} X_{1t} + \phi_{221} X_{2t}$$

$$\hat{X}_{2t}(k) = \phi_{211}\{ \phi_{111}^{k-1} X_{1t} + \phi_{121}[\hat{X}_{2t}(k-2) + \phi_{111}\hat{X}_{2t}(k-3) + \ldots + \phi_{111}^{k-2} X_{2t}] \} + \phi_{221}\hat{X}_{2t}(k-1), \quad k \geq 2$$

In practice, forecasting by a leading indicator is well worth exploring, in particular, the number of observations in a given series is too small. The additional information provided by the leading indicator may improve the parameter estimates and as well the forecasting.

3. Example of Forecasting with the help of leading indicator

e.g, The Dow-Jones and the Australian All-ordinaries Indices data (shown in Fig 7-1)

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0.0295 \\ 0.0309 \end{bmatrix} = \begin{bmatrix} -0.0148 & 0.0357 \\ 0.06584 & 0.0998 \end{bmatrix} \left(\begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} - \begin{bmatrix} 0.0295 \\ 0.0309 \end{bmatrix} \right) + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

where

$$\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \propto NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.3653 & 0.0224 \\ 0.0224 & 0.6016 \end{bmatrix} \right)$$

One-step ahead mean squared error for the prediction of X2t is thus 0.6016 assuming the above model is adequate. This is a substantial reduction from the variance of X2t of 0.7712 when the sample mean of 0.0309 is used as the one-step ahead predictor.

If we fit a univariate model to X2t,

$$X_{2t} = 0.0273 + 0.1180 X_{2t-1} + a_{2t} \qquad \text{where} \quad a_{2t} \propto NID(0, 0.7604)$$

One-step ahead mean squared error for the prediction of X2t is 0.7604, which is larger than the prediction error of the vectorial model.

4. Conditional Expectation from Orthogonal Decomposition

The techniques learned in Chapter 5 can be applied directly.

For example,

$$X_{2t} = (1 - \phi_{221} B)^{-1} (\phi_{211} X_{1t-1} + a_{2t})$$

$$using$$

$$X_{1t} = a_{1t}$$

We will have,

$$
\begin{aligned}
X_{2t} &= (1-\phi_{221}B)^{-1}(\phi_{211}a_{1t-1}+a_{2t}) \\
&= (1+\phi_{221}B+\phi_{221}^2B^2+\ldots)(\phi_{211}a_{1t-1}+a_{2t}) \\
&= \sum_{j=0}^{\infty} G_j(\phi_{211}a_{1t-j-1}+a_{2t-j})
\end{aligned}
$$

$$
\hat{X}_{2t}(\ell) = G_{\ell-1}\phi_{211}a_{1t} + G_\ell(\phi_{211}a_{1t-1}+a_{2t}) + G_{\ell+1}(\phi_{211}a_{1t-2}+a_{2t-1}) + \ldots, \ell \geq 1
$$

Forecasting errors,

$$
\begin{aligned}
e_{2t}(\ell) = X_{2t+\ell} - \hat{X}_{2t}(\ell) &= (\phi_{211}a_{1t+\ell-1}+a_{2t+\ell}) + G_1(\phi_{211}a_{1t+\ell-2}+a_{2t+\ell-1}) + \\
&\ldots + G_{\ell-2}(\phi_{211}a_{1t+1}+a_{2t+2}) + G_{\ell-1}a_{2t+1}, \ell \geq 2
\end{aligned}
$$

$$
V[e_{2t}(\ell)] = (1+G_1^2+G_2^2+\ldots+G_{\ell-2}^2)(\phi_{211}^2\gamma_{a11}+\gamma_{a22}) + G_{\ell-1}^2\gamma_{a22}, \ell \geq 2
$$

Optimal Stochastic Regulation (control that keeps a constant output) Based on ARMAV Models

Minimum mean squared error control strategy

Goal: to keep the output at the target value, which may be taken as zero.

Problems: due to noise and disturbance, it is very difficult to maintain the output exactly at zero level; to keep the deviation or errors from the zero target values as small as possible.

Smallness: for random variables with zero mean, the measure of their smallness is given by their variance.

Optimal control: adjustments in the manipulable input values that yield minimum variance (minimum mean squared error) of the output.

MMSE control strategy: to adjust the input X_{1t} such that the forecast of X_{2t+L} made at time t is zero (or target value). [the earliest time at which the input can affect the output is t+L].

$$\hat{X}_{2t}(L) = 0$$

$$X_{2t+L} = \hat{X}_{2t}(L) + e_{2t}(L)$$

After implementing the control equation,

$$X_{2t+L} = e_{2t}(L)$$

This optimally controlled output, or output error or deviation from the mean, is the L-step ahead forecasting error, which is a MA(L-1) model.

Example 1 — First order model with lag 1

a) model before control

$$X_{2t}=0.25X_{1t-1}+0.7X_{2t-1}+a_{2t} \qquad \gamma_{a22}=0.0062$$

for papermaking process data

b) control equation

$$\hat{X}_{2t}(1)=0.25X_{1t}+0.7X_{2t}=0 \quad \rightarrow \quad 0.25X_{1t}=-0.7X_{2t} \quad \text{or} \quad X_{1t}=-2.8X_{2t}$$

where,

we used zero target value since X_{2t} is the deviation of basis weight from its mean.

$$X_{1t}-X_{1t-1}=-2.8X_{2t}+2.8X_{2t-1} \qquad \nabla X_{1t}=-2.8\nabla X_{2t}$$

This is an equivalent P-control.

c) model after control