Vectorial (Multivariate) ARMA Models (ARMAV Models)

$\vec{X}_t$, $t \in \mathbb{Z}$ is a wss vectorial random process if:

* $E[\vec{X}_t] = \text{const.}$ (without loss of generality, let's assume it's zero)

* $$E[\vec{X}_t \, \vec{X}_{t-\ell}^T] = \begin{bmatrix} E[X_{1_t} X_{1_{t-\ell}}] & E[X_{1_t} X_{2_{t-\ell}}] & \dots & E[X_{1_t} X_{d_{t-\ell}}] \\ E[X_{2_t} X_{1_{t-\ell}}] & E[X_{2_t} X_{2_{t-\ell}}] & \dots & E[X_{2_t} X_{d_{t-\ell}}] \\ - & - & - & - \\ E[X_{d_t} X_{1_{t-\ell}}] & E[X_{d_t} X_{2_{t-\ell}}] & \dots & E[X_{d_t} X_{d_{t-\ell}}] \end{bmatrix}$$

Note: $\vec{X}_t = \begin{bmatrix} X_{1_t} \\ X_{2_t} \\ \vdots \\ X_{d_t} \end{bmatrix}$

$= \Gamma_\ell$ (matrices of 2nd order moments are origin independent)

Vectorial Wold's decomposition theorem

Any wss vectorial random process X_t, $t \in \mathbb{Z}$ can be represented as

$$\vec{X}_t = \vec{\varepsilon}_t + \Psi_1 \vec{\varepsilon}_{t-1} + \Psi_2 \vec{\varepsilon}_{t-2} + \dots + \Psi_\ell \varepsilon_{t-\ell} + \dots \qquad (*)$$

where

* $\sum_{\ell=0}^{\infty} \| \Psi_\ell \|^2 < \infty \qquad (\Psi_0 = I)$

* $E[\vec{\varepsilon}_t \vec{\varepsilon}_{t-\ell}^T] = \begin{cases} \Sigma , & \ell = 0 \\ 0 , & \text{otherwise} \end{cases}$

If we introduce

$$\vec{a}_t = \mathbb{T} \cdot \vec{\varepsilon}_t$$

so that

$$E[\vec{a}_t \vec{a}_t^T] = D \qquad (\text{diagonalizing } \Sigma)$$

then $(*)$ can be expressed as

$$\vec{X}_t = \Theta_0 \vec{a}_t + \Theta_1 a_{t-1} + \Theta_2 \vec{a}_{t-2} + \dots + \Theta_\ell \vec{a}_{t-\ell} + \dots$$

where $E[\vec{a}_t] = \vec{0} \qquad E[\vec{a}_t \vec{a}_{t-\ell}^T] = D \delta_\ell$

Since $\sum_{i=0}^{\infty} \|\Theta_i\|^2 < \infty$

we can approximate $\vec{X}_t$ to within an arbitrarily small accuracy ε using the form

$$\vec{X}_t = \underbrace{\begin{bmatrix} R_{1,1}(B) & R_{1,2}(B) & \dots & R_{1,d}(B) \\ R_{2,1}(B) & R_{2,2}(B) & \dots & R_{2,d}(B) \\ - & - & - & - \\ R_{d,1}(B) & R_{d,2}(B) & \dots & R_{d,d}(B) \end{bmatrix}}_{\mathbb{R}(B)} \vec{a}_t \qquad (1)$$

where $R_{i,j}(B)$ are rational functions of backshift operators of sufficiently high order.

Inverting (1) gives

$$\vec{a}_t = \mathbb{R}^{-1}(B) \cdot \vec{X}_t \qquad (2)$$

and since $R^{-1}(B)$ is also a matrix of rational functions of "B", (2) gives

$$a_{1_t} = \frac{P_{11}(B)}{Q_{11}(B)} X_{1_t} + \frac{P_{12}(B)}{Q_{12}(B)} X_{2_t} + \dots + \frac{P_{1d}(B)}{Q_{1d}(B)} X_{d_t}$$

$$- - - - - - -$$

$$a_{d_t} = \frac{P_{d,1}(B)}{Q_{d,1}(B)} X_{1_t} + \dots + \frac{P_{d,d}(B)}{Q_{d,d}(B)} X_{d_t} \quad \text{---} \quad (3)$$

This now leads to

$$(1 - \theta_{1_1} B - \dots - \theta_{1_{m^{(1)}}} B^{m^{(1)}}) a_{1_t} =$$

$$(\phi_{0_{11}} - \phi_{1_{11}} B - \dots - \phi_{n^{(1)}_{11}} B^{n^{(1)}}) X_{1_t} + (\phi_{0_{1,2}} - \phi_{1_{1,2}} B - \dots - \phi_{n^{(1)}_{1,2}} B^{n^{(1)}} X_{2_t} +$$

$$\dots (\phi_{0_{1,d}} + \phi_{1_{1,d}} B + \dots + \phi_{n^{(1)}_{1,d}} B^{n^{(1)}}) X_{d_t} \quad (i)$$

$$1 - \theta_{2_1} B - \dots - \theta_{2_{m^{(2)}}} B^{m^{(2)}}) a_{2_t} =$$

$$(\phi_{0_{2,1}} - \phi_{1_{2,1}} B - \dots - \phi_{n^{(2)}_{2,1}} B^{n^{(2)}}) X_{1_t} +$$

$$(\phi_{0_{2,2}} - \phi_{1_{2,2}} B - \dots - \phi_{n^{(2)}_{2,2}} B^{n^{(2)}}) X_{2_t} + \dots$$

$$+ (\phi_{0_{2,d}} - \phi_{1_{2,d}} B - \dots - \phi_{n^{(2)}_{2,d}} B^{n^{(2)}}) X_{d_t} \quad (ii)$$

$$\vdots$$

$$(1-\theta_{d_1}B-\theta_{d_2}B^2-\dots-\theta_{d_{m^{(d)}}}B^{m^{(d)}})\,a_{d_t} =$$

$$=(\phi_{0_{d,1}}-\phi_{1_{d,1}}B-\dots-\phi_{n^{(d)}_{d,1}}B^{n^{(d)}})X_{1_t} +$$

$$+(\phi_{0_{d,2}}-\phi_{1_{d,2}}B-\dots-\phi_{n^{(d)}_{d,2}}B^{n^{(d)}})X_{2_t} + \dots$$

$$\dots+(\phi_{0_{d,d}}-\phi_{1_{d,d}}B-\dots-\phi_{n^{(d)}_{d,d}}B^{n^{(d)}})X_{d_t} \qquad (d)$$

Note, since all rational functions in (1) must be strictly rational (degree of the numerator poly is smaller than the degree of the denominator poly), we must have that $m^{(i)} < n^{(i)}$, $i=1,2,\dots,d$!

So all this summarizes down to an elegant matrix formulation saying that any WSS vectorial random process $\vec{X}_t$, $t\in\mathbb{Z}$ can be approximated to within arbitrary accuracy using models of the form

$$\vec{a}_t - \Theta_1 a_{t-1} - \dots - \Theta_m \vec{a}_{t-m} =$$

$$= \Phi_0 \vec{X}_t - \Phi_1 \vec{X}_{t-1} - \dots - \Phi_n X_{t-n} \qquad \star$$

where $E[\vec{a}_t] = \vec{0}$, $E[\vec{a}_t a_{t-e}^T] = \begin{cases} \mathbb{D}, & e=0 \\ 0, & \text{othws} \end{cases}$

and Θ_i are diagonal!

Note: eqns i), ii), ..., d, are independent optimization problems in the sense that their parameters can be obtained via independent optimizations of RSS!

i) $\to$ get $\theta_1, \dots, \theta_{m^{(1)}}, \phi_{0,1,j}, \phi_{1,j}, \dots, \phi_{n^{(1)},1,j}$ $\quad j = 1, 2, \dots, d$

by minimizing the corresponding RSS

(max likelihood estimates of those params)

ii) $\rightarrow$ get $\theta_{2_1}, \theta_{2_2}, \dots, \theta_{2_{m^{(2)}}}$,

$\phi_{0_{2,j}}, \phi_{1_{2,j}}, \dots, \phi_{n^{(2)}_{2,j}}$ $\quad j = 1, 2, \dots, d$

by minimizing the corresponding RSS
(max likelihood estimates of those params)

etc ...

$\Rightarrow$ Each row in $\bigstar$ can be identified via
independent optimizations!

How can we do modeling? Just like in the
scalar case $\rightarrow$ only must repeat the process
many times (d times)

$$X_{1t} = -\phi_{0\,12} X_{2t} + \phi_{1\,11} X_{1\,t-1} + \phi_{1\,12} X_{2\,t-1} + \phi_{2\,11} X_{1\,t-2} + \phi_{2\,12} X_{2\,t-2} +$$

$$\dots + \phi_{n\,11} X_{1t} + \phi_{n\,12} X_{2t} + a_{1t} - \theta_{11} a_{1\,t-1} - \dots - \theta_{m\,11} a_{1\,t-m} \qquad (*)$$

$$X_{2t} = -\phi_{0\,21} X_{1t} + \phi_{1\,21} X_{1\,t-1} + \phi_{1\,22} X_{2\,t-1} + \phi_{2\,21} X_{1\,t-2} + \phi_{2\,22} X_{t-2} + \dots +$$

$$+ \phi_{n\,21} X_{1\,t-n} + \phi_{n\,22} X_{2\,t-n} + a_{2t} - \theta_{1\,22} a_{2\,t-1} - \theta_{2\,22} a_{2\,t-2} \dots - \theta_{m\,22} a_{2\,t-m} \qquad (**)$$

Modeling strategy:

Once I know the structure of the model ARMAV(n, m), I can define a minimization procedure – but how do I get the structure?

Again we can use F-test.

① Increase order of model (*) untill RSS_1 does not decrease significantly

② Increase order of model (**) untill RSS_2 does not decrease significantly

③ The order of the vectorial model is the maximal order found in (*) and (**)

From T 11.2 (X_{1t} series)

Test 1

$RSS_1 = 0.1051$ $s = 3$

$RSS_0 = 0.08464$ $r = 6$

$$F = \frac{(RSS_1 - RSS_0)/s}{RSS_0/(N-r)} = 3.552$$

$F^{95\%}_{3,44} = 2.84 < F$ => keep going

Test 2

$RSS_1 = 0.08464$ $s = 3$

$RSS_0 = 0.03525$ $r = 9$

$F = 19.14$ $F^{95\%}_{3,41} = 2.84$ => keep going

Test 3

$RSS_1 = 0.03525$ $s = 3$

$RSS_0 = 0.02892$ $r = 12$

$F = 2.845$ $F^{95}_{3,38} = 2.845$ (roughly)

=> We can stop here

=> again (4,3)

Table 11.2 Modeling Papermaking Process Input; Gate Opening X_1, (Mean = 18.302, Variance = 0.0024)

Parameters	Model Order			
	(1,0)	(2,1)	(3,2)	(4,3)
ϕ_{120}	0.104 ± 0.168	0.121 ± 0.164	0.123 ± 0.150	0.112 ± 0.113
ϕ_{111}	0.002 ± 0.276	-0.290 ± 0.569	0.300 ± 0.228	0.265 ± 0.537
ϕ_{112}		0.023 ± 0.277	-0.553 ± 0.056	-0.455 ± 0.392
ϕ_{113}			-0.053 ± 0.191	-0.059 ± 0.241
ϕ_{114}				0.022 ± 0.135
ϕ_{121}	0.191 ± 0.164	0.006 ± 0.176	0.151 ± 0.212	0.126 ± 0.144
ϕ_{122}		0.260 ± 0.173	0.136 ± 0.220	0.199 ± 0.155
ϕ_{123}			-0.0121 ± 0.163	-0.036 ± 0.055
ϕ_{124}				-0.013 ± 0.058
θ_{111}		-0.221 ± 0.643	0.591 ± 0.319	0.494 ± 0.195
θ_{112}			-1.451 ± 0.291	-1.523 ± 0.103
θ_{113}				-0.089 ± 0.453
Residual sum of squares	0.1051	0.08464	0.03525	0.02892
γ_{a11}	0.00214	0.00176	0.00075	0.00063
F	—	3.552	19.14	2.845

The adequate model is (3,2).

Example from the book (Tables 11.1 and 11.2);
N=50 observations of time-series X_{1t} and X_{2t}

ever, the external inputs such as step or impulse are still commonly used because of the difficulties of obtaining the dynamics from the normal operating data. An attempt by Goh (1973) to model this data by empirical cross-correlation methods did not succeed because of large disturbance.

Modeling of this data by the procedure of Section 11.2.1 will now be illustrated. We first fit a first order model to the output X_{2t} series. As in Step 1, the initial values of the parameters ϕ_{211} and ϕ_{221} are computed using the procedure illustrated in Appendix A 11.1. The initial values also indicate that there is no significant delay between input and output, and hence the lag was chosen as one. The nonlinear least squares method gives the results for the output X_{2t} and the input X_{1t} presented in Tables 11.1 and 11.2, respectively.

It is seen from Table 11.1 that the first order model fitted to the output-input model is

$$X_{2t} = 0.247X_{1t-1} + 0.696X_{2t-1} + a_{2t}, \quad \gamma_{a22} = 0.00624$$

The improvement in the residual sum of squares from the first order to second order is significant since

$$F = \frac{0.3060 - 0.1950}{3} \div \frac{0.1950}{50 - 5} = 8.538$$

compared with $F_{0.95}(3,45) = 2.84$. Comparing the (2,1) model with the (3,2) gives $F = 3.14$. On the other hand, ARMA(4,3) model (not given

Table 11.1 Modeling Papermaking Process Output, Basis Weight, X_{2t} (Mean = 38.969, Variance = .0134)

Parameters	Model Order		
	(1,0)	(2,1)	(3,2)
ϕ_{211}	0.247 ± 0.461	0.275 ± 0.457	0.454 ± 0.422
ϕ_{212}		-0.593 ± 0.471	-0.416 ± 0.378
ϕ_{213}			-0.536 ± 0.428
ϕ_{221}	0.696 ± 0.195	1.453 ± 0.245	0.470 ± 0.308
ϕ_{222}		-0.523 ± 0.241	0.923 ± 0.251
ϕ_{223}			-0.518 ± 0.267
θ_{221}		1.125 ± 0.239	0.029 ± 0.323
θ_{222}			1.304 ± 0.349
Residual sum of squares	0.3060	0.1950	0.1593
γ_{a22}	0.00624	0.00406	0.00339
F		8.538	3.140

The adequate model is (3,2).
Number of observations = 50.

From 11.1 (t-series X_{2t})

<u>Test 1</u> $RSS_1 = 0.306$ $s = 3$ (restricted params)

$RSS_0 = 0.195$ $r = 5$ (estimated pars)

$$F = \frac{(RSS_1 - RSS_0)/s}{RSS_0/(N-r)} = 8.538$$

$$F^{95\%}_{3,45} = 2.84 < F \Rightarrow \text{keep going}$$

Test 2 $RSS_1 = 0.155$. $s = 3$

$RSS_0 = 0.1593$ $r = 8$

$$F = \frac{(RSS_1 - RSS_0)/s}{RSS_0/(N-r)} = 3.140$$

$$F^{95}_{3,42} = 2.84 < F \Rightarrow \text{keep going}$$

They went to (4,3) which gave $F = 0.693$ & the F test indicated we should stop

$\Rightarrow X_{1t}$ model order was (3,2)