ISMT S-136 Time Series Analysis with Python

Harvard Summer School

Dmitry Kurochkin

Summer 2021
Lecture 9

Contents

Contents

Motivation

Recall: Let $\phi(B) = 1 - \phi_1 B - \ldots - \phi_m B^m$, then

$$\text{AR}(m)\text{-ARCH}(p): \quad \phi(B)x_t = r_t$$
$$r_t = \sigma_t \varepsilon_t, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j r_{t-j}^2.$$

$$\text{AR}(m)\text{-GARCH}(p, q): \quad \phi(B)x_t = r_t$$
$$r_t = \sigma_t \varepsilon_t, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j r_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

Note that, given $x_{t-1}, x_{t-2}, \ldots$, volatility σ_t^2 is conditionally nonstochastic. Can GARCH be modified to also include (conditional) stochastisity of σ_t^2?

Answer: Stochastic Volatility (SV) Model.

Contents

Definition of Stochastic Volatility (SV) Model

Let's consider GARCH$(1, 1)$:

$$r_t = \sigma_t \varepsilon_t, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1),$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

or equivalently:

$$\ln r_t^2 = \ln \sigma_t^2 + \ln \varepsilon_t^2, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1),$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

Let's modfiy the model and introduce stochastic volatility:

Def.
Stochastic volatility (SV) model for r_t is defined as follows:

$$\ln r_t^2 = \ln \sigma_t^2 + \ln \varepsilon_t^2, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1),$$
$$\ln \sigma_t^2 = \phi_0 + \phi_1 \ln \sigma_{t-1}^2 + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2).$$

Definition of Stochastic Volatility (SV) Models

Remark:
The stochastic volatility model,

$$\ln r_t^2 = \ln \sigma_t^2 + \ln \varepsilon_t^2, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$
$$\underbrace{\ln \sigma_t^2}_{v_t} = \phi_0 + \phi_1 \underbrace{\ln \sigma_{t-1}^2}_{v_{t-1}} + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma_w^2),$$

can be rewritten as follows:

$$r_t = e^{\frac{v_t}{2}} \varepsilon_t, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0,1),$$
$$v_t = \phi_0 + \phi_1 v_{t-1} + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma_w^2),$$

where we introduced $v_t = \ln \sigma_t^2$.

Contents

Parameter Estimation: Likelihood for SV Model

We notice that in

$$r_t = e^{\frac{v_t}{2}} \varepsilon_t, \quad \text{where} \quad \varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1),$$

$$v_t = \phi_0 + \phi_1 v_{t-1} + w_t, \quad \text{where} \quad w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2),$$

both v_t and ε_t are unobserved.

In order to obtain estimates of the parameters, one needs to maximize the likelihood function, given observations $r_1, r_2, \ldots, r_n$.

Contents

Cross-Covariance and Joint Stationary

Def.
Given two time series x_t and y_t,

$$\gamma_{xy}(s,t) = \mathrm{Cov}(x_s, y_t) = \mathrm{E}[(x_s - \mu_{xs})(y_t - \mu_{yt})]$$

is called *cross-covariance function* between the two series x_t and y_t.

Def.
Two time series x_t and y_t are called *jointly stationary* if
they are stationary and the cross-covariance function depends on $s-t$ only.

Notation:
$\gamma_{xy}(s,t)$ will be denoted by $\gamma_{xy}(s-t)$ in case of jointly stationary x_t and y_t.

Remark:
$\gamma_{xy}(h)$ is generally not an even function.

Cross-Correlation Function (CCF) between x_t and y_t

Def.

Given two time series x_t and y_t,

$$\begin{aligned}\rho_{xy}(s,t) =& \mathrm{Corr}(x_s, y_t)\\ =& \frac{\mathrm{Cov}(x_s, y_t)}{\sqrt{\mathrm{Var}(x_s)\mathrm{Var}(y_t)}}\\ =& \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}}\end{aligned}$$

is called *cross-correlation function* (CCF) between the two series x_t and y_t.

Notation:

If x_t and y_t are jointly stationary, $\rho_{xy}(s,t)$ depends on $s-t$ only and will be denoted by $\rho_{xy}(s-t)$ then.

Remark:

$\rho_{xy}(h)$ is generally not an even function.

Example 1.23 Joint Stationarity

Let

$$
\begin{aligned}
x_t =& w_t + w_{t-1}, \\
y_t =& w_t - w_{t-1},
\end{aligned}
$$

where $w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$.

Then

$$
\rho_{xy}(h) = \begin{cases} 0, & h = 0, \\ \frac{1}{2}, & h = 1, \\ -\frac{1}{2}, & h = -1, \\ 0, & |h| \geq 2. \end{cases}
$$

Example 1.24 Prediction Using Cross-Correlation

Let x_t be a stationary time series.
Assume that x_t "leads" y_t (equivalently, y_t "lags" x_t):

$$y_t = ax_{t-5} + w_t,$$

where $w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$.

Then

$$\begin{aligned}\gamma_{yx}(h) =& \mathrm{Cov}(y_{t+h}, x_t)\\ =& \mathrm{Cov}(ax_{t+h-5} + w_{t+h}, x_t)\\ =& \mathrm{Cov}(ax_{t+h-5}, x_t)\\ =& a\gamma_x(h-5),\end{aligned}$$

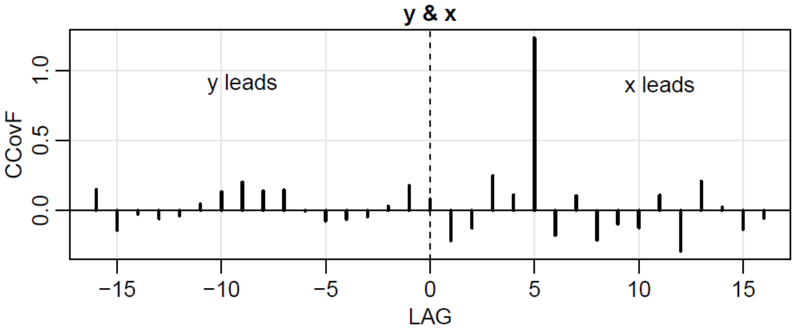
i.e. $\gamma_{yx}(h)$ is "shifted" $\gamma_x(h-5)$.

Notice that x_t and y_t are jointly stationary.

Example 1.24 Prediction Using Cross-Correlation (continued)

Stationary x_t and $y_t = a x_{t-5} + w_t$, where $w_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$.
Then $\gamma_{yx}(h) = a\gamma_x(h-5)$.

Example:



Source: *Time Series Analysis and Its Applications: With R Examples*
by R. Shumway and D. Stoffer

Sample Cross-Covariance Function

Def.

Given a realization $(x_1, y_t), (x_2, y_2), \ldots, (x_n, y_n)$ of jointly stationary time series x_t and y_t, the sample cross-covariance function is defined as follows:

$$\hat{\gamma}_{xy}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(y_t - \bar{y})$$

for $h = 0, 1, \ldots, n-1$, and

$$\hat{\gamma}_{xy}(-h) = \hat{\gamma}_{yx}(h).$$

Sample Cross-Correlation Functions (Sample CCF)

Def.

Given $\hat{\gamma}_{xy}(h)$ for $h = 0, \pm 1, \ldots, \pm(n-1)$ for a realization $(x_1, y_t), (x_2, y_2), \ldots, (x_n, y_n)$ of jointly stationary time series x_t and y_t, the corresponding *sample CCF* is:

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}},$$

where $\hat{\gamma}_x(\cdot)$ and $\hat{\gamma}_y(\cdot)$ are ACFs of x_t and y_t, respectively.

Example 1.29 Prewhitening and Cross Correlation Analysis

Let

$$
\begin{aligned}
x_t =& 2\cos(2\pi t \frac{1}{12}) + w_{t,1},\\
y_t =& 2\cos(2\pi [t+5] \frac{1}{12}) + w_{t,2},
\end{aligned}
$$

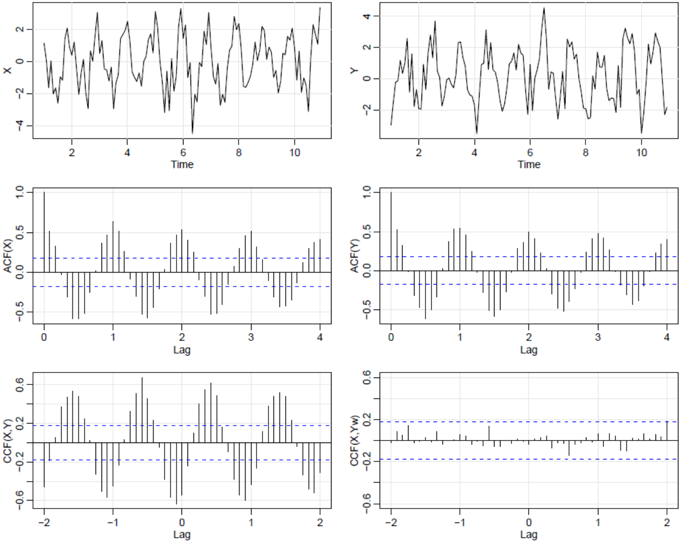
where $w_{t,1}, w_{t,2} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

Then

$$
\begin{aligned}
\gamma_{xy}(h) =& \mathrm{Cov}(x_{t+h}, y_t)\\
=& \mathrm{Cov}(2\cos(2\pi[t+h]\frac{1}{12}) + w_{t+h,1}, 2\cos(2\pi[t+5]\frac{1}{12}) + w_{t,2})\\
=& \mathrm{Cov}(w_{t+h,1}, w_{t,2})\\
=& 0,
\end{aligned}
$$

for all h.

Example 1.29 Prewhitening and Cross Correlation Analysis (continued)



Contents

Vector-valued Time Series

Def.
We define a *vector-valued* time series x_t as the following collection $x_{t,1}, x_{t,2}, \ldots, x_{t,k}$ of k time series:

$$x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T = \begin{bmatrix} x_{t,1} \\ x_{t,2} \\ \vdots \\ x_{t,k} \end{bmatrix}.$$

Stationarity of Vector-valued Time Series

Def.

A vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ is called *stationary* if

$$x_i \text{ and } x_j \text{ are jointly stationary}$$

for all $i \neq j$, $i, j \in \{1, 2, \ldots, k\}$.

Contents

Autocovariance Matrix

Notation:
Let $x_t = (x_{t,1}, x_{t,2}, \dots, x_{t,k})^T$ be a stationary vector-valued time series.
We introduce an expectation $\mu = (\mu_1, \mu_2, \dots, \mu_k)^T$ (i.e. $k \times 1$ matrix) of x_t as follows:

$$\mu = \mathrm{E}[x_t] = \begin{bmatrix} \mathrm{E}[x_{t,1}] \\ \mathrm{E}[x_{t,2}] \\ \vdots \\ \mathrm{E}[x_{t,k}] \end{bmatrix}$$

Autocovariance Matrix

Def.
Let $x_t = (x_{t,1}, x_{t,2}, \dots, x_{t,k})^T$ be a stationary vector-valued time series. The $k \times k$ matrix

$$\Gamma(h) = \mathrm{E}[(x_{t+h} - \mu)(x_t - \mu)^T]$$

$$= \mathrm{E}\left[\begin{bmatrix} x_{t+h,1} - \mu_1 \\ x_{t+h,2} - \mu_2 \\ \vdots \\ x_{t+h,k} - \mu_p \end{bmatrix} \begin{bmatrix} x_{t,1} - \mu_1 & x_{t,2} - \mu_2 & \dots & x_{t,k} - \mu_p \end{bmatrix}\right]$$

$$= \mathrm{E}\begin{bmatrix} (x_{t+h,1} - \mu_1)(x_{t,1} - \mu_1) & (x_{t+h,1} - \mu_1)(x_{t,2} - \mu_2) & \dots & (x_{t+h,1} - \mu_1)(x_{t,k} - \mu_k) \\ (x_{t+h,2} - \mu_2)(x_{t,1} - \mu_1) & (x_{t+h,2} - \mu_2)(x_{t,2} - \mu_2) & \dots & (x_{t+h,2} - \mu_2)(x_{t,k} - \mu_k) \\ \vdots & & \ddots & \\ (x_{t+h,k} - \mu_k)(x_{t,1} - \mu_1) & (x_{t+h,k} - \mu_k)(x_{t,2} - \mu_2) & \dots & (x_{t+h,k} - \mu_k)(x_{t,k} - \mu_k) \end{bmatrix}$$

is called *autocovariance matrix*.

Notice:

1. Off-diagonal elements, $\gamma_{ij}(h) = \mathrm{E}[(x_{t+h,i} - \mu_i)(x_{t,j} - \mu_j)]$, $i \neq j$, of $\Gamma(h)$ are the cross-covariance functions between $x_{t,i}$ and $x_{t,j}$.
2. Diagonal elements, $\gamma_{ij}(h) = \mathrm{E}[(x_{t+h,i} - \mu_i)(x_{t,i} - \mu_i)]$, of $\Gamma(h)$ are the autocovariance functions of $x_{t,i}$.

Sample Autocovariance Matrix

Def.

The *sample autocovariance matrix* is defined as

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})^T,$$

for $h = 0, 1, \ldots, n-1$, and

$$\hat{\Gamma}(-h) = \hat{\Gamma}^T(h).$$

Contents

Vector Autoregressive Model of First Order (VAR(1))

Def.

We say that a vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \dots, x_{t,k})^T$ follows *Vector Autoregressive* model of first order, abbreviated VAR(1), if

$$x_t = \Phi x_{t-1} + w_t,$$

where

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \dots & \Phi_{1k} \\ \Phi_{21} & \Phi_{22} & \dots & \Phi_{2k} \\ \vdots & & \ddots & \\ \Phi_{k1} & \Phi_{k2} & \dots & \Phi_{kk} \end{bmatrix}$$

and w_t is the *vector Gaussian white noise*, that is, multivariate normal with zero mean, $\mathrm{E}[w_s w_t^T] = 0$ for all $s \neq t$, and some covariance matrix,

$$\Sigma_w = \mathrm{E}[w_t w_t^T],$$

which is generally not diagonal.

Vector Autoregressive Model of Order p (VAR(p))

Def.

We say that a vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ follows *Vector Autoregressive* model of first order, abbreviated VAR(p), if

$$x_t = \Phi_1 \, x_{t-1} + \ldots + \Phi_p \, x_{t-p} + w_t,$$

where for each $j \in \{1, 2, \ldots, p\}$

$$\Phi_j = \begin{bmatrix} \Phi_{j,11} & \Phi_{j,12} & \ldots & \Phi_{j,1k} \\ \Phi_{j,21} & \Phi_{j,22} & \ldots & \Phi_{j,2k} \\ \vdots & & \ddots & \\ \Phi_{j,k1} & \Phi_{j,k2} & \ldots & \Phi_{j,kk} \end{bmatrix}$$

and w_t is the *vector Gaussian white noise*, that is, multivariate normal with zero mean, $\mathrm{E}[w_s w_t^T] = 0$ for all $s \neq t$, and some covariance matrix,

$$\Sigma_w = \mathrm{E}[w_t w_t^T],$$

which is generally not diagonal.

Contents

Parameter Estimation: Yule–Walker Equations for VAR

If we have a VAR(p) process:

$$x_t = \Phi_1 \, x_{t-1} + \ldots + \Phi_p \, x_{t-p} + w_t,$$

where $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$, Φ_p are $k \times k$ matrices, and w_t is the vector Gaussian white noise.

Then similarly to the univariate AR(p) case, one show that the following Yule-Walker equations hold:

$$\begin{aligned} \Gamma(h) &= \Phi_1 \, \Gamma(h-1) + \ldots + \Phi_p \, \Gamma(h-p), \quad \text{for } h = 1, 2, \ldots \\ \Gamma(0) &= \Phi_1 \, \Gamma(-1) + \ldots + \Phi_p \, \Gamma(-p) + \Sigma_w, \end{aligned}$$

where

$$\Sigma_w = \mathrm{E}[w_t w_t^T].$$

Parameter Estimation: Yule–Walker Equations for VAR

If we replace $\Gamma(h)$ by the sample autocovariance matrix $\hat{\Gamma}(h)$ (i.e. use method of moments),
we get the estimates (namely Yule-Walker estimators) for unknown parameters as a solution to the following linear system of equations:

$$\hat{\Gamma}(h) = \Phi_1 \hat{\Gamma}(h-1) + \ldots + \Phi_p \hat{\Gamma}(h-p), \quad \text{for } h = 1, 2, \ldots$$
$$\hat{\Gamma}(0) = \Phi_1 \hat{\Gamma}(-1) + \ldots + \Phi_p \hat{\Gamma}(-p) + \Sigma_w.$$

Contents

Parameter Estimation: Likelihood for VAR(p)

If we have a VAR(p) process:

$$x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + w_t,$$

where $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$, Φ_j are $k \times k$ matrices, and w_t is the vector Gaussian white noise.

Then similarly to the univariate AR(p) case, one can write the likelihood function explicitly.

Contents

Vector Autoregressive Moving Average (VARMA)

Def.

We say that a vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ follows *Vector Autoregressive Moving Average* model of order p, q, abbreviated VARMA(p, q), if

$$x_t = \sum_{j=1}^{p} \Phi_j \, x_{t-j} + \sum_{i=1}^{q} \Theta_i w_{t-i} + w_t,$$

where for each $j \in \{1, 2, \ldots, p\}$ and $i \in \{1, 2, \ldots, q\}$ we have

$$\Phi_j = \begin{bmatrix} \Phi_{j,11} & \Phi_{j,12} & \ldots & \Phi_{j,1k} \\ \Phi_{j,21} & \Phi_{j,22} & \ldots & \Phi_{j,2k} \\ \vdots & & \ddots & \\ \Phi_{j,k1} & \Phi_{j,k2} & \ldots & \Phi_{j,kk} \end{bmatrix}, \; \Theta_i = \begin{bmatrix} \Theta_{i,11} & \Theta_{i,12} & \ldots & \Theta_{i,1k} \\ \Theta_{i,21} & \Theta_{i,22} & \ldots & \Theta_{i,2k} \\ \vdots & & \ddots & \\ \Theta_{i,k1} & \Theta_{i,k2} & \ldots & \Theta_{i,kk} \end{bmatrix},$$

respectively, and w_t is the *vector Gaussian white noise*, that is, multivariate normal with zero mean, $\mathrm{E}[w_s w_t^T] = 0$ for all $s \neq t$, and some covariance matrix,

$$\Sigma_w = \mathrm{E}[w_t w_t^T],$$

which is generally not diagonal.

Contents

Parameter Estimation: Likelihood for VARMA(p, q)

If we have a VARMA(p, q) process:

$$x_t = \sum_{j=1}^{p} \Phi_j \, x_{t-j} + \sum_{i=1}^{q} \Theta_i w_{t-i} + w_t,$$

where $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$, Φ_j and Θ_i are $k \times k$ matrices, and w_t is the vector Gaussian white noise.

Then similarly to the univariate ARMA(p) case, one can write the likelihood function explicitly.

Contents

VAR with Exogenous Inputs (VARX)

Def.

Assume there is a vector $u_t = (u_{t,1}, u_{t,2}, \ldots, u_{t,r})^T$ of r pre-determined inputs.

We say that a vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ follows *Vector Autoregressive* model of order p with *exogenous inputs*, abbreviated VARX(p), if

$$x_t = A u_t + \Phi_1 \, x_{t-1} + \Phi_2 \, x_{t-2} + \ldots + \Phi_p \, x_{t-p} + w_t,$$

where for each $j \in \{1, 2, \ldots, p\}$

$$\Phi_j = \begin{bmatrix} \Phi_{j,11} & \Phi_{j,12} & \ldots & \Phi_{j,1k} \\ \Phi_{j,21} & \Phi_{j,22} & \ldots & \Phi_{j,2k} \\ \vdots & & \ddots & \\ \Phi_{j,k1} & \Phi_{j,k2} & \ldots & \Phi_{j,kk} \end{bmatrix}, \; A = \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1r} \\ A_{21} & A_{22} & \ldots & A_{2r} \\ \vdots & & \ddots & \\ A_{k1} & A_{k2} & \ldots & A_{kr} \end{bmatrix},$$

and w_t is the *vector Gaussian white noise*, that is, multivariate normal with zero mean, $\mathrm{E}[w_s w_t^T] = 0$ for all $s \neq t$, and some covariance matrix,

$$\Sigma_w = \mathrm{E}[w_t w_t^T],$$

which is generally not diagonal.

VARMA with Exogenous Inputs (VARMAX)

Def.

Assume there is a vector $u_t = (u_{t,1}, u_{t,2}, \ldots, u_{t,r})$ of r pre-determined inputs. We say that a vector-valued time series $x_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ follows *Vector Autoregressive Moving Average* model of order p, q with *exogenous inputs*, abbreviated VARMAX(p, q), if

$$x_t = Au_t + \sum_{j=1}^{p} \Phi_j \, x_{t-j} + \sum_{i=1}^{q} \Theta_i w_{t-i} + w_t,$$

where for each $j \in \{1, 2, \ldots, p\}$ and $i \in \{1, 2, \ldots, q\}$ we have

$$\Phi_j = \begin{bmatrix} \Phi_{j,11} & \Phi_{j,12} & \ldots & \Phi_{j,1k} \\ \Phi_{j,21} & \Phi_{j,22} & \ldots & \Phi_{j,2k} \\ \vdots & & \ddots & \\ \Phi_{j,k1} & \Phi_{j,k2} & \ldots & \Phi_{j,kk} \end{bmatrix}, \quad \Theta_i = \begin{bmatrix} \Theta_{i,11} & \Theta_{i,12} & \ldots & \Theta_{i,1k} \\ \Theta_{i,21} & \Theta_{i,22} & \ldots & \Theta_{i,2k} \\ \vdots & & \ddots & \\ \Theta_{i,k1} & \Theta_{i,k2} & \ldots & \Theta_{i,kk} \end{bmatrix},$$

respectively, A is an $k \times r$ matrix, and w_t is the *vector white noise*, that is, multivariate normal with zero mean, $\mathrm{E}[w_s w_t^T] = 0$ for all $s \neq t$, and some covariance matrix,

$$\Sigma_w = \mathrm{E}[w_t w_t^T],$$

which is generally not diagonal.