# Ch8. Multivariate Time Series

# Multivariate Time Series

- Consider the sequence of vector-valued (multi-dimensional) data, namely, $\mathbf{X}_t = (X_{t1}, \ldots, X_{tm})'$.
- For example, closing values of Dow Jones Index and Australian All Ordinaries Index (AAOI) of share prices.

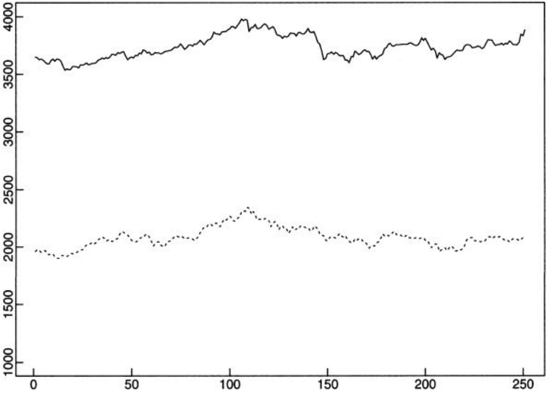



**Figure 7-1**
The Dow Jones Index (top) and Australian All Ordinaries Index (bottom) at closing on 251 trading days ending August 26th, 1994.

# Multivariate Time Series

- In Multivariate TS analysis, we are also interested in the dependence between $m$-component series.
- For example, DJI and AAOI move together or decoupled?
- Denote data vector as

$$\mathbf{X}_t = \begin{pmatrix} X_{t1} \\ X_{t2} \\ \vdots \\ X_{tm} \end{pmatrix}, \quad t = 1, 2, \ldots, n,$$

and observed data is denoted by $\mathbf{x}_t$.

# Moments

- Mean (vector)

$$\boldsymbol{\mu} = \begin{pmatrix} EX_{t1} \\ EX_{t2} \\ \vdots \\ EX_{tm} \end{pmatrix}$$

- Covariance matrix

$$\Gamma(t+h,\, t) := \mathrm{Cov}(\mathbf{X}_{t+h},\, \mathbf{X}_t) = E(\mathbf{X}_{t+h} - \boldsymbol{\mu}_{t+h})(\mathbf{X}_t - \boldsymbol{\mu}_t)'$$

$$= \begin{pmatrix} \mathrm{Cov}(X_{t+h,1},\, X_{t,1}) & \cdots & \mathrm{Cov}(X_{t+h,1},\, X_{t,m}) \\ \mathrm{Cov}(X_{t+h,2},\, X_{t,1}) & \cdots & \mathrm{Cov}(X_{t+h,2},\, X_{t,m}) \\ \vdots & & \vdots \\ \mathrm{Cov}(X_{t+h,m},\, X_{t,1}) & \cdots & \mathrm{Cov}(X_{t+h,m},\, X_{t,m}) \end{pmatrix}$$

# Weakly Stationary

Similar to a univariate case, weakly stationary multivariate time series is defined as follows.

**Definition**

*The $m$-variante series $\{\mathbf{X}_t\}$ is (weakly) stationary if*

i) $\boldsymbol{\mu}_X(t)$ *is independent of* $t$

ii) $\Gamma_X(t+h, t)$ *is independent of* $t$ *for each* $h$.

- If $\{X_t\}$ is stationary

$$\boldsymbol{\mu} = E\mathbf{X}_t = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$$

# Covariance matrix function

- The covariance matrix becomes under stationarity

$$
\begin{aligned}
\Gamma(h) &:= E(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})' = \{\mathrm{Cov}(X_{t+h,i}, X_{t,j})\}_{i,j=1,\ldots,m} \\
&= \{\mathrm{Cov}(X_{h,i},\ X_{0,j})\}_{i,j=1,\ldots,m} \\
&= \begin{pmatrix} \mathrm{Cov}(X_{h,1},\ X_{0,1}) & \cdots & \mathrm{Cov}(X_{h,1},\ X_{0,m}) \\ \vdots & & \vdots \\ \mathrm{Cov}(X_{h,m},\ X_{0,1}) & \cdots & \mathrm{Cov}(X_{h,m},\ X_{0,m}) \end{pmatrix} \\
&:= \begin{pmatrix} \gamma_{11}(h) & \ldots & \gamma_{1m}(h) \\ \vdots & & \vdots \\ \gamma_{m1}(h) & \ldots & \gamma_{mm}(h) \end{pmatrix} = \{\gamma_{ij}(h)\}_{i,j=1,\ldots,m}
\end{aligned}
$$

- If $i = j$, then it is a usual auto-covariance function.
- $\{\gamma_{ij}(h)\} := \mathrm{Cov}(X_{t+h,i}, X_{t,j})$ is called the cross-covariance function.

# Covariance matrix function

- That is, the function $\gamma_{i,j}(h)$, $i \neq j$, measure the covariance between $X_{t,i}$ and $X_{t,j}$ when $h$ lags apart.
- Thus, $\Gamma(h)$ measures auto-covariance from the diagonal entries, and measures cross correlation in the off-diagonal entries.
- If $\gamma_{i,j}(h)$ is normalized

$$\rho_{ij}(h) := \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}},$$

it is called the cross-correlation function and $R(\cdot)$ given by

$$R(h) := \begin{pmatrix} \rho_{11}(h) & \dots & \rho_{1m}(h) \\ \vdots & & \vdots \\ \rho_{m1}(h) & \dots & \rho_{mm}(h) \end{pmatrix},$$

is called the correlation matrix function.

## Basic properties of $\Gamma(\cdot)$

- $|\gamma_{ij}(h)| \le (\gamma_{ii}(0)\,\gamma_{jj}(0))^{\frac{1}{2}}$, $i\ j = 1, \ldots, m$
- $\gamma_{ii}(.)$ is a covariance function of $\{X_{ti}\}$
- Recall that

  $$\boxed{\gamma_{i,j}(h) := \mathrm{Cov}(X_{t+h,i}, X_{t,j})}$$

  Therefore,

  $$\begin{aligned}\gamma_{ij}(h) &= \mathrm{Cov}(X_{t+h,i},\ X_{t,j}) = \mathrm{Cov}(X_{t,j}, X_{t+h,i}) \\ &= \mathrm{Cov}(X_{t'-h,j},\ X_{t',i}) = \gamma_{ji}(-h).\end{aligned}$$

  That is, $\gamma_{ij}(h) = \gamma_{ji}(-h)$. Remark that $\gamma_{ij}(h) \neq \gamma_{ji}(h)$. In a matrix notation we have

  $$\boxed{\Gamma(h) = \Gamma'(-h)}$$

- (non-negative definite) $\sum_{j,k=1}^{n} a_j' \Gamma(j-k) a_k \ge 0$ for all $n \in \{1, 2, \ldots\}$ and $a_1, \ldots, a_n \in \mathbb{R}^m$. Because $E(\sum_{i=1}^{n} a_j' (\mathbf{X}_j - \boldsymbol{\mu}))^2 \ge 0$.

# Example of MTS

Consider the bivariate stationary process $\{\mathbf{X}_t\}$ given by

$$X_{t1} = Z_t$$
$$X_{t2} = Z_t + .75 Z_{t-10},$$

where $Z_t \sim WN(0, 1)$. Then,

$$\boldsymbol{\mu} =$$

$$\Gamma(10) = \qquad \Gamma(0) = \qquad \Gamma(-10) =$$

# Examples of MTS - Multivariate White Noise (MWN)

**Definition**

*The $m$-variate series $\mathbf{Z}_t, t = 0, \pm 1, \pm 2, \ldots\}$ is said to be white noise with mean $\mathbf{0}$ and covariance matrix $\Sigma$*

$$\{\mathbf{Z}_t\} \sim WN(0, \Sigma)$$

*iff $\{\mathbf{Z}_t\}$ is stationary with mean vector $\mathbf{0}$ and covariance function*

$$\Gamma(h) = \begin{cases} \Sigma & \text{if } h = 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

- $\{\mathbf{Z}_t\} \sim \mathrm{IID}(\mathbf{0}, \Sigma)$ to indicate that the random vectors $\mathbf{Z}_t$ are IID with mean $\mathbf{0}$ and covariance matrix $\Sigma$.

# Linear process

**Definition**

*The $m$-variate series $\{\mathbf{X}_t\}$ is a linear process if it has the representation*

$$\mathbf{X}_t = \sum_{k=-\infty}^{\infty} C_k \mathbf{Z}_{t-k}, \quad \mathbf{Z}_t \sim WN(0, \Sigma),$$

*where $C_k$ is a sequence of $m \times m$ matrices with $\sum_{k=-\infty}^{\infty} |C_k(i,j)| < \infty$ for all $i, j = 1, \ldots, m$.*

- Linear process is stationary with mean $\mathbf{0}$ and

$$\Gamma(h) = EX_{t+h}X_t' = \sum_{k=-\infty}^{\infty} C_{k+h}\Sigma C_k', \quad h = 0, \pm 1, \ldots$$

- Will introduce Multivariate ARMA$(p, q)$ model soon.

## 8.2 Estimation of the Mean and Covariance Function

- As in the univariate case, the method of moment estimator will be introduced. However, keep in mind that the covariance function is also non-negative definite

# Estimation of Mean vector $\boldsymbol{\mu}$

- Based on the observations $\mathbf{X}_1, \dots, \mathbf{X}_n$, an unbiased estimate of $\boldsymbol{\mu}$ is given by

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{X}}_n = \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t = \left(\frac{1}{n}\sum_{t=1}^{n} X_{t1}, \dots, \frac{1}{n}\sum_{t=1}^{n} X_{tm}\right)'$$

- Recall for a univariate stationary TS, we have

$$\frac{\overline{X} - \mu}{\sqrt{\nu/n}} \xrightarrow{d} \mathcal{N}(0,1), \quad \nu = \sum_{h=-\infty}^{\infty} \gamma(h).$$

$\nu$ is calle the **long-run variance**. Also (asymptotic) $100(1-\alpha)\%$ CI by

$$\overline{X} \pm z_{\alpha/2}\sqrt{\frac{\hat{\nu}}{n}},$$

where

$$\hat{\nu} = \sum_{|h|<r} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}(h).$$

# Estimation of Mean vector $\boldsymbol{\mu}$

In multivariate setting, we have similar results.

### Theorem

*For the stationary multivariate time series $\{\mathbf{X}_t\}$ written as,*

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{k=-\infty}^{\infty} C_k \mathbf{Z}_{t-k}, \quad \mathbf{Z} \sim \mathrm{IIN}(0, \Sigma),$$

$$\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} MVN\left(0, \widetilde{\Sigma}\right),$$

*where* $\widetilde{\Sigma} = \left(\sum_{k=-\infty}^{\infty} C_k\right) \Sigma \left(\sum_{k=-\infty}^{\infty} C_k'\right)$.

- Thus, the $100(1-\alpha)\%$ confidence region (simultaneous CI) is

$$\{\boldsymbol{\mu} \in \mathbb{R}^m : (\boldsymbol{\mu} - \bar{\mathbf{X}}_n)' n \widetilde{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}_n) \leq \chi^2_{1-\alpha}(m)\}.$$

- However, not practice since $\widetilde{\Sigma}$ is very difficult to estimate.

# Confidence interval for $\boldsymbol{\mu}$

- Instead, we will construct CI for each $\mu_i$ and combine these to form a confidence region for $\boldsymbol{\mu}$.
- For $i$-th componet, univariate analysis says that

$$\mathrm{CI}^{(i)} : \left|\mu_i - \overline{X}_n^{(i)}\right| \leq z_{\alpha/2}\sqrt{\frac{\hat{\nu}^{(i)}}{n}},$$

  where $\hat{\nu}^{(i)}$ is the long-run variance estimator.
- Thus, we might temp to conclude that $100(1-\alpha)\%$ CI for $\boldsymbol{\mu}$ is given by aggregating $m$-components, namely,

$$\left\{\overline{X}_n^{(i)} \pm z_{\alpha/2}\sqrt{\frac{\hat{\nu}^{(i)}}{n}}, i = 1, ..., m\right\}.$$

- However, above confidence region is seriously over-sized!

# Bonferroni correction

- For exmaple, consider the coverage probability

$$P\left(\bigcap_{i=1}^{m} A_i\right) = 1 - P\left(\bigcup_{i=1}^{m} A_i^c\right) \geq 1 - \sum_{i=1}^{m} P(A_i^c) = 1 - m\alpha$$

if $P(A_i) = 1 - \alpha$.

- We want to construct CI with at least $(1-\alpha)$ coverage probability.
- Bonferroni suggested the following simple rule, take

$$P(A_i) = 1 - \alpha/m$$

then $P(\cap_{i=1}^{m} A_i) \geq 1 - \alpha$.

# Confidence region for $\boldsymbol{\mu}$

Therefore, by applying Bonferroni correction, $100(1-\alpha)\%$ CI for $\boldsymbol{\mu}$ is given by

$$\left\{ \overline{X}_n^{(i)} \pm z_{\alpha/2m} \sqrt{\frac{\hat{\nu}^{(i)}}{n}}, i = 1, ..., m \right\},$$

where the Bartlett long-run estimator of $\nu^{(i)}$ is given by

$$\hat{\nu}^{(i)} = \sum_{|h|<r} \left(1 - \frac{|h|}{n}\right) \hat{\gamma}(h))$$

Observe also that in the spectral domain, the long-run variance is equivalent to

$$2\pi f(0) = \sum_{h=-\infty}^{\infty} \gamma_{ii}(h).$$

Hence smoothed spectral density estimator (that is, the Bartlett kernel is replaced by other choices of Kernel) can be used for estimation.

# Estimation of $\Gamma(h)$

- The method of moment estimator of the covariance function

$$\Gamma(h) = E(\mathbf{X}_{t+h} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'$$

is given by

$$\boxed{\widehat{\Gamma}(h) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-h}(\mathbf{X}_{t+h} - \overline{\mathbf{X}}_n)(\mathbf{X}_t - \overline{\mathbf{X}}_n)', & h = 0, ..., n-1 \\ \widehat{\Gamma}'(-h), & \text{if } -n+1 \leq h < 0 \end{cases}}$$

- Hence, the cross-correlation $\rho_{ij}(h) := \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}}$ is estimated by

$$\boxed{\hat{\rho}_{ij}(h) = \frac{\hat{\gamma}_{ij}(h)}{\sqrt{\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)}}}$$

- (Consistancy) Since $\hat{\gamma}_{ij}(h) \xrightarrow{p} \gamma_{ij}(h)$, in turn $\hat{\rho}_{ij}(h) \xrightarrow{p} \rho_{ij}(h)$.

# Properties of the $\hat{\rho}_{ij}(h)$

- (Asymptotic Normality) the derivation of the asymptotic distribution of the sample cross-correlation function is quite complicated. For example, even for independent bivariate time series (that is $\rho_{12}(h) = 0, h \neq 0$), we have that

$$\sqrt{n}(\hat{\rho}_{12}(h) - 0) \xrightarrow{d} \mathcal{N}\left(0, \sum_{k=-\infty}^{\infty} \rho_{11}(k)\rho_{22}(k)\right) \quad (1)$$

- From (1), observe that

$$\sqrt{n}\hat{\rho}_{12}(h) \sim \mathcal{N}(0, 1)$$

  if $\rho_{11}(k) = \rho_{22}(k) = 0$ for all except $k = 0$. It means that testing independence of two component series cannot be solely based on the estimation of $\rho_{12}(h)$, and need to estimate $\rho_{11}(\cdot)$, $\rho_{22}(\cdot)$. Hence, using aymptotic (1) is not so practical way of testing independence.

# Improvement: Prewhitening

- Instead, "prewhitening" the two series before computing the cross-correlations $\hat{\rho}_{12}(h)$ circumvent this difficulty.
- Basic idea is to fit ARMA$(p, q)$ model to each series

$$\begin{cases} \Phi(B)X_{t,1} = \Theta(B)Z_{t,1} \\ \Phi^*(B)X_{t,2} = \Theta^*(B)Z_{t,2} \end{cases}$$

to obtain uncorrelated series

$$\begin{cases} Z_{t,1} = \Theta(B)^{-1}\Phi(B)X_{t,1} := \Pi(B)X_{t,1} \\ Z_{t,2} = \Theta^*(B)^{-1}\Phi^*(B) := \Pi^*(B)X_{t,2} \end{cases}$$

- However, $Z_{t1}, Z_{t2}$ are unobservable, we replace them by

$$\widehat{W}_t = \frac{X_t - \widehat{X}_t}{\sqrt{r_{t-1}}} \approx Z_t$$

and calculate the cross-correlation based on $\{\widehat{W}_{t1}\}$ and $\{\widehat{W}_{t2}\}$

# Prewhitening

- Now, reject test $H_0 : \rho_{ij}(h) = 0$ if

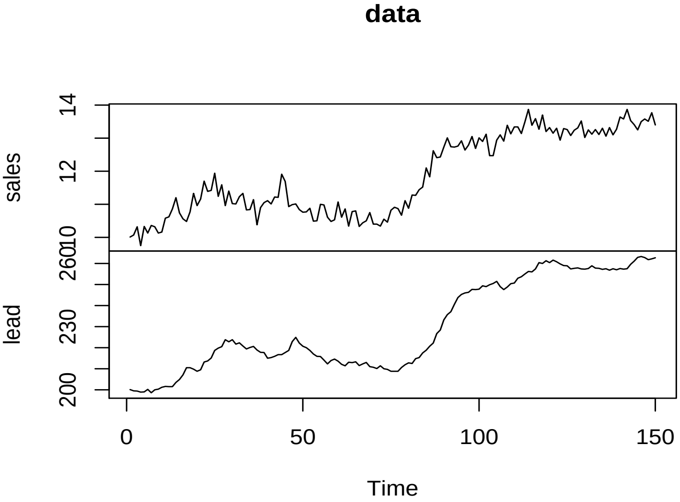
$$\tilde{\rho}_{ij}(h) = \left| \frac{\tilde{\gamma}_{ij}(h)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}} \right| > \frac{z_{\alpha/2}}{\sqrt{n}}$$

  Example 7.3.2 Sales with a leading indicator since $\rho_{11}(j)$ and $\rho_{22}(j)$ are zero in (1) once prewhitened.

- In practice, $2/\sqrt{n}$ rule still can be used to test the significance of the cross-correlations once you prewhiten the data.

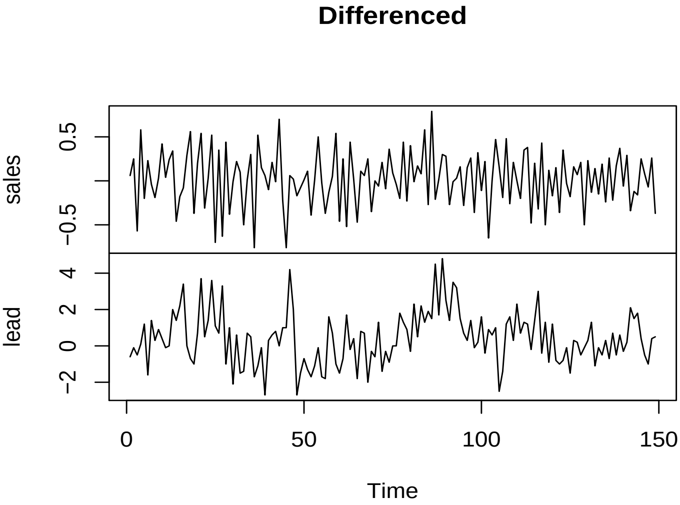
# Example: Sales and Leading indicator

Interested in the total sales and leading indicator



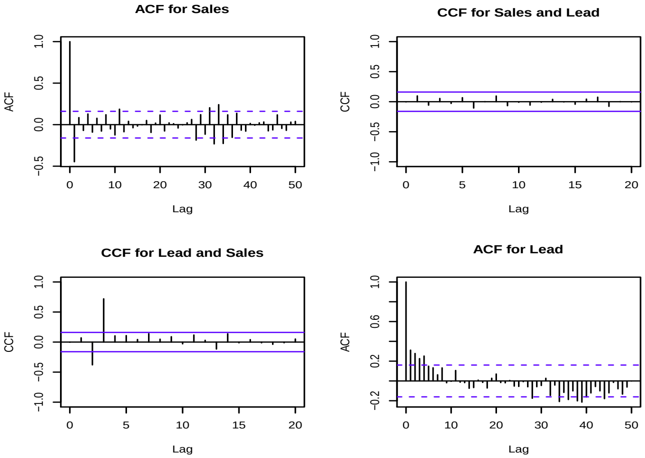
# Example: Sales and Leading indicator

Differenced



# Example: Sales and Leading indicator

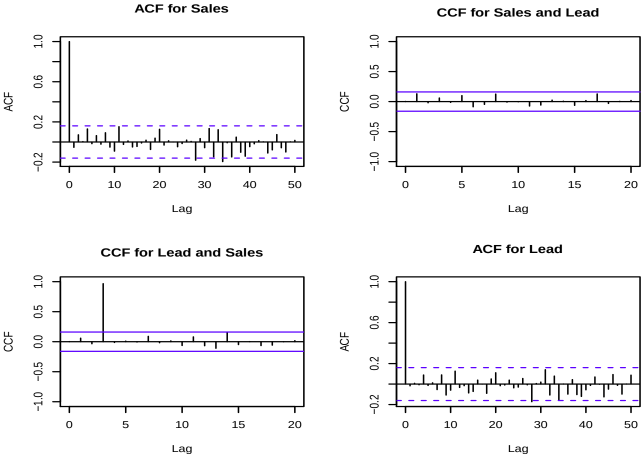
The sample CCF behaves as




However, without taking into account the autocorrelations, it is not possible to draw any conclusions.

# Example: Sales and Leading indicator

By fitting best ARMA model (MA(1) and ARMA(1,1)), we obtain residual series and calculate CCFs based upon them.




Now we conclude that $\rho_{21}(3) = \rho_{12}(-3) \neq 0$!

# Multivariate ARMA processes

Like in a univariate time series, we can define multivariate ARMA($p, q$) processes (or VARMA($p, q$)) as follows:

> **Definition**
>
> $\{\mathbf{X}_t\}$ *is an VARMA(*$p, q$*) process if stationary and if for every* $t$*,*
>
> $$\mathbf{X}_t - \Phi_1 \mathbf{X}_{t-1} - \ldots - \Phi_p \mathbf{X}_{t-p} = \mathbf{Z}_t + \Theta_1 \mathbf{Z}_{t-1} + \ldots + \Theta_q \mathbf{Z}_{t-q},$$
>
> *where* $\{\mathbf{Z}_t\} \sim WN(0, \Sigma)$*.*

- Consider the conditions to be stationary. By setting $p = 1$ and $q = 0$, we have VAR(1) process

$$\begin{aligned} \mathbf{X}_t &= \Phi \mathbf{X}_{t-1} + \mathbf{Z}_t = \Phi(\Phi \mathbf{X}_{t-2} + \mathbf{Z}_{t-1}) + \mathbf{Z}_t = \ldots \\ &= \sum_{j=0}^{\infty} \Phi^j \mathbf{Z}_{t-j} \end{aligned}$$

# VAR(1)

- First, in order to become a linear process, we need absolutely summable condition for $\Phi^j$, that is, each component of the matrix $\sum_{j=0}^{\infty} \Phi^j \mathbf{Z}_{t-j}$ converges.
- Since it is absolutely summable,

$$\Phi^j \to 0, \quad \text{as } j \to \infty.$$

  Furthermore, if $\Phi$ is diagonalizable, then

$$\Phi^j = P\Lambda^j P^{-1} \to 0$$

  implies that all eigenvalues should less than 1 in absolute value.
- Since eigenvalues satisfy $\Phi v = \lambda v$ for non-zero $v$, it is equivalent to

$$\det|\Phi - \lambda I| = 0, \quad |\lambda| < 1$$

# VAR(1)

- Like AR(1) series requires $|\phi| < 1$, we have

$$\boxed{\det(I - z\Phi) \neq 0 \text{ for all } z \in \mathbb{Z} \text{ such that } |z| \leq 1.}$$

  (More rigorous argument need Jordan canonical form)

- Stationarity check:
  i) $E\mathbf{X}_t = 0 \ \forall t$
  ii) Covariance matrix function is calculated as

$$\begin{aligned}
\Gamma(h) &= E(\mathbf{X}_{t+h}\mathbf{X}_t') \\
&= \lim_{n\to\infty} E\left(\sum_{j=0}^{n} \Phi^j \mathbf{Z}_{t-j}\right)\left(\sum_{j=0}^{n} \Phi^j \mathbf{Z}_{t-j}\right)' \\
&= \lim_{n\to\infty} \sum_{k=0}^{n} \Phi^{h+k} \Sigma \Phi^k
\end{aligned}$$

  Thus, does not depend on $t$.

# Example: VAR(1)

Consider VAR(1) process with coefficients

$$\mathbf{X}_t = \begin{pmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t$$

Then, the determinant of characteristic polynomial becomes

$$\det\left( \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - z \begin{pmatrix} .5 & 0 & 0 \\ .1 & .1 & .3 \\ 0 & .2 & .3 \end{pmatrix} \right)$$

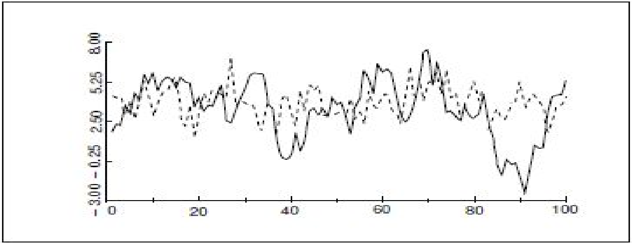
$$= \begin{pmatrix} 1-.5z & 0 & 0 \\ -.1z & 1-.1z & -.3z \\ 0 & -.2z & 1-.3z \end{pmatrix} = (1-.5z)(1-.4z-.03z^2)$$

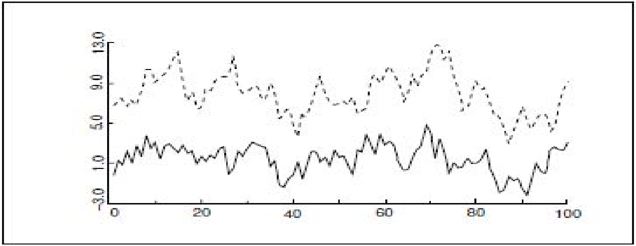
Hence, the roots are

$$z_1 = 2, \quad z_2 = 2.1525, \quad z_3 = -15.4858$$

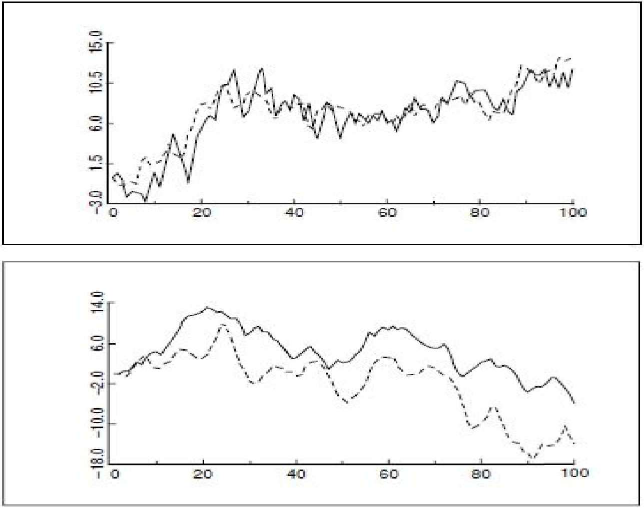
so that the given process is stationary.

# Stationary (stable) VAR



# Nonstationary (unstable) VAR



# Causality

## Definition

*An ARMA*($p, q$) *process* $\{\mathbf{X}_t\}$ *is causal, or a* **causal function of** $\{\mathbf{Z}_t\}$, *if there exist matrices* $\{\Psi_j\}$ *with absolutely summable condition such that*

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{Z}_{t-j} \quad \text{for all } t.$$

*Causality is equivalent to the condition*

$$\det\Phi(z) \neq 0 \text{ for all } z \in \mathbb{C} \text{ such that } |z| \leq 1.$$

# Invertibility

**Definition**

*An ARMA($p, q$) process $\{\mathbf{X}_t\}$ is invertible, if there exist matrices $\{\Pi_j\}$ with absolutely summable condition such that*

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \Pi_j \mathbf{X}_{t-j} \quad \textit{for all } t.$$

*Invertibility is equivalent to the condition*

$$\det \Pi(z) \neq 0 \textit{ for all } z \in \mathbb{C} \textit{ such that } |z| \leq 1.$$

# Example: VARMA(1,1)

Consider VARMA(1,1) given by

$$\mathbf{X}_t - \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0.5 \end{pmatrix} \mathbf{X}_{t-1} = \mathbf{Z}_t + \begin{pmatrix} 0.5 & 0 \\ 0.5 & 0.5 \end{pmatrix} \mathbf{Z}_{t+1}$$

Want to find the causal representation.

- Method 1: Use $\Psi(z) = \Phi^{-1}(z)\Theta(z)$
  $= \begin{pmatrix} 1-0.5z & -0.5z \\ 0 & 1-0.5z \end{pmatrix}^{-1} \begin{pmatrix} 1+0.5z & 0 \\ 0.5z & 1+0.5z \end{pmatrix}$
  $= (1-0.5z)^{-2} \begin{pmatrix} 1 & 0.5z(1+0.5z) \\ 0.5z(1-0.5z) & 1-0.25z^2 \end{pmatrix}$
  Then,
  $$\mathbf{X}_t = \Psi(B)\mathbf{Z}_t$$

# Example: VARMA(1,1)

- Method2: Recursively, we have

$$\begin{aligned}\mathbf{X}_t &= \Phi_1 \mathbf{X}_{t-1} + \mathbf{Z}_t + \Theta_1 \mathbf{Z}_{t-1} \\ &= \mathbf{Z}_t + \Theta_1 \mathbf{Z}_{t-1} + \Phi_1(\Phi_1 \mathbf{X}_{t-2} + \mathbf{Z}_{t-1} + \Theta_1 \mathbf{Z}_{t-2}) \\ &= \mathbf{Z}_t + (\Theta_1 + \Phi_1)\mathbf{Z}_{t-1} + \Phi_1 \Theta_1 \mathbf{Z}_{t-2} + \Phi_1^2(\Phi_1 \mathbf{X}_{t-1} + \cdots)\end{aligned}$$

- In general

$$\boxed{\Psi_j = \Theta_j + \sum_{k=1}^{\infty} \Phi_k \Psi_{j-k}.}$$

- From this we can also deduce that the covariance matrix function of VARMA$(p, q)$ process becomes

$$\boxed{\Gamma(h) = \sum_{j=0}^{\infty} \Psi_{j+h} \Sigma \Psi_j', \quad h = 0, \pm 1, \ldots}$$

# Non-uniqueness of VARMA representation

- This is quite different perspective in VARMA representation. Consider VAR(1) given by

$$\mathbf{X}_t = \Phi \mathbf{X}_{t-1} + \mathbf{Z}_t, \quad \Phi = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}.$$

  Then,

$$\begin{aligned} \mathbf{X}_t &= \Phi(\Phi \mathbf{X}_{t-2} + \mathbf{Z}_{t-1}) + \mathbf{Z}_t \\ &= \Phi^2 \mathbf{X}_{t-2} + \Phi \mathbf{Z}_{t-1} + \mathbf{Z}_t \\ &= \mathbf{Z}_t + \Phi \mathbf{Z}_{t-1} \end{aligned}$$

  since $\Phi^2 = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
- It means that the VAR(1) process can also be written as VMA(0,1).
- For the uniqueness of representation such as Echelon form or standard form, some conditions are required. See *Lütkepohl* (§ 7.1) for details.

# Estimation of VAR - LSE

- Now, we consider the estimation of VAR($p$) process. Two methods - LSE and MLE will be introduced.
- Consider the VAR($p$) process written as

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \ldots + \Phi_p \mathbf{X}_{t-p} + \mathbf{Z}_t, \quad t = 1, \ldots, n. \tag{2}$$

We can write (2) in a huge block matrix. Denote $x_1, \ldots, x_n$ be the vector of data observations. Then,

$$(x_{p+1}, \ldots, x_n) = (\Phi_1 x_p + \ldots + \Phi_1 x_1, \Phi_1 x_{p+1} + \Phi_p x_2$$

$$, \ldots, \Phi_1 X_{n-1} + \cdots + \Phi_p X_{n-p}) + (z_{p+1}, , ..., z_n)$$

# Estimation of VAR - LSE

$$= (\Phi_1 \; \Phi_2 \; \cdots \; \Phi_p) \begin{pmatrix} x_p & x_{p+1} & x_{p+2} & \cdots & x_{n-1} \\ x_{p-1} & x_p & x_{p+1} & \cdots & x_{n-2} \\ \vdots & \vdots & x_0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1 & x_2 & x_3 & \ldots & x_{n-p} \end{pmatrix} + (z_{p+1}, ..., z_n)$$

$$\boxed{Y^0 = AX + Z,} \tag{3}$$

where $Y^0$ is $m \times (n-p)$, $A$ is a $m \times mp$ parameter matrix, $X$ is $mp \times (n-p)$ design matrix and $Z$ is $m \times (n-p)$ error matrix.

# Vectorize and Kronecker product

- To apply OLS regression formula, it would be nice if we can represent (3) in a column vector form.
- Such operation is possible and called vectorisation.
- In R, you can simple apply `as.vector()`.

Let $A = (a_1, \ldots, a_n)$ be an $m \times n$ matrix with $m \times 1$ columns $a_i$. The *vec operator* transforms $A$ into an $mn \times 1$ vector by stacking the columns, that is,

$$\mathrm{vec}(A) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

# $\mathrm{vec}()$ operator

- For example,

$$\mathrm{vec}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}.$$

- If

$$A = \begin{pmatrix} 3 & 0 \\ 2 & -1 \\ 4 & 0 \end{pmatrix},$$

then

$$\mathrm{vec}(A) =$$

# Kronecker Product $\otimes$

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ and $p \times q$ matrices. Then, the *Kronecker product* or *direct product* of $A$ and $B$ is the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \vdots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix},$$

where

$$a_{ij}B = \begin{pmatrix} a_{ij}b_{11} & \dots & a_{ij}b_{1q} \\ a_{ij}b_{21} & \dots & a_{ij}b_{2q} \\ \vdots & \vdots & \vdots \\ a_{ij}b_{p1} & \dots & a_{ij}b_{pq} \end{pmatrix}$$

# Kronecker Product $\otimes$

$$A = \begin{pmatrix} 3 & 4 & -1 \\ 2 & 0 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 5 & -1 \\ 3 & 3 \end{pmatrix}$$

$A \otimes B =$

$B \otimes A =$

# Properties of Kronecker product

For $A_{m\times n}$ and $B_{p\times q}$ otherwise specified

1. $A \otimes B \neq B \otimes A$ in general
2. $(A \otimes B)' = A' \otimes B'$
3. $A \otimes (B + C) = A \otimes B + A \otimes C$
4. $(A \otimes B)(C \otimes D) = AC \otimes BD$
5. If $A$ and $B$ are invertible, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
6. If $A$ and $B$ are $m \times m$ and $n \times n$ square matrices, respectively, then $|A \otimes B| = |A|^n |B|^m$
7. If $A$ and $B$ are square matrices

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$$

# Properties of $\mathrm{vec}()$ operator

Let $A, B, C$ be matrices with appropriate dimensions.

1. $\mathrm{vec}(A + B) = \mathrm{vec}(A) + \mathrm{vec}(B)$
2. $\mathrm{vec}(ABC) = (C' \otimes A)\ \mathrm{vec}(B)$
3. $\mathrm{vec}(AB) = \mathrm{vec}(ABI) = (I \otimes A)\ \mathrm{vec}(B) = (B' \otimes I)\ \mathrm{vec}(A)$
4. $\mathrm{vec}(ABC) = (I \otimes AB)\ \mathrm{vec}(C) = (C'B' \otimes I)\ \mathrm{vec}(A)$
5. $(\mathrm{vec}(B'))'\ \mathrm{vec}(A) = \mathrm{tr}(BA) = \mathrm{tr}(AB) = (\mathrm{vec}(A'))'\ \mathrm{vec}(B)$

- The $\mathrm{vech}()$ operator is closely related to $\mathrm{vec}$. It only stacks the diagonal entries of a square matrix.

# Estimation of VAR - LSE

Now, consider

$$\boxed{Y^0 = AX + Z,}$$

Then, applying $\mathrm{vec}$ operation gives

$$\begin{aligned}
\mathrm{vec}(Y^0) &= \mathrm{vec}(AX + Z) \\
&= \mathrm{vec}(AX) + \mathrm{vec}(Z) \\
&= \mathrm{vec}(I_m AX) + \mathrm{vec}(Z) \\
&= (X' \otimes I_m)\mathrm{vec}(A) + \mathrm{vec}(Z)
\end{aligned}$$

Denote parameter vectors

$$\alpha := \mathrm{vec}(A) = \begin{pmatrix} \mathrm{vec}(\Phi_1) \\ \mathrm{vec}(\Phi_2) \\ \vdots \\ \mathrm{vec}(\Phi_p) \end{pmatrix},$$

# Estimation of VAR - LSE

Then, the least squares estimator of $\alpha$ is given by

$$\begin{aligned}
\widehat{\alpha}^{LSE} &= \left((X' \otimes I_m)'(X' \otimes I_m)\right)^{-1}(X' \otimes I_m)'\mathrm{vec}(Y^0) \\
&= \left((X \otimes I_m)(X' \otimes I_m)\right)^{-1}(X \otimes I_m)\mathrm{vec}(Y^0) \\
&= \left(XX' \otimes I_m\right)^{-1}(X \otimes I_m)\mathrm{vec}(Y^0) \\
&= \left((XX')^{-1} \otimes I_m\right)(X \otimes I_m)\mathrm{vec}(Y^0) \\
&= \left((XX')^{-1}X \otimes I_m\right)\mathrm{vec}(Y^0) \\
&= \mathrm{vec}\left(Y^0 X'(XX')^{-1}\right)
\end{aligned}$$

Therefore

$$\boxed{\widehat{A}^{LSE} = Y^0 X'(XX')^{-1}}$$

# Asymptotics of LSE estimator

- For LSE estimator $\hat{\alpha}$, we have the following asymptotics

$$\boxed{\sqrt{n}(\hat{\alpha} - \alpha) = \sqrt{n}(\text{vec}(\hat{A} - A)) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma),}$$

  where $n^{-1}XX' \xrightarrow{p} \Gamma$. (Note that $X$ is a random matrix)

- Furtermore $\Sigma = E(\mathbf{Z}_t\mathbf{Z}_t{}')$ implies that

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{n}\sum_{t=1}^{n} \hat{\mathbf{Z}}_t\hat{\mathbf{Z}}_t{}' = \frac{1}{n}ZZ' = \frac{1}{n}(Y^0 - \hat{A}X)(Y^0 - \hat{A}X)' \\
&= \frac{1}{n}(Y^0 - Y^0X'((XX')^{-1}X)(Y^0 - Y^0X'((XX')^{-1}X)' \\
&= \frac{1}{n}Y^0(I_n - X'(XX')^{-1}X)(I_n - X'(XX')^{-1}X)Y^{0\prime} \\
&= \frac{1}{n}Y^0(I_n - X'(XX')^{-1}X)Y^{0\prime}
\end{aligned}$$

# Yule-Walker equation for $\Gamma(h)$

In VAR($p$) model,

$$\mathbf{X}_{t+h} = \Phi_1 \mathbf{X}_{t+h-1} + \ldots + \Phi_p \mathbf{X}_{t+h-p} + \mathbf{Z}_{t+h},$$

postmultiply $\mathbf{X}_t{}'$ and take expectation gives

$$\Gamma(h) = \Phi_1 \Gamma(h-1) + \ldots + \Phi_p \Gamma(h-p) + E(\mathbf{Z}_{t+h}\mathbf{X}_t{}')$$

$$\left\{ \begin{array}{lcl} \Gamma(0) & = & \Phi_1 \Gamma(1)' + \cdots + \Phi_p \Gamma(p)' + \Sigma \\ \Gamma(h) & = & \Phi_1 \Gamma(h-1) + \cdots + \Phi_p \Gamma(h-p), \ h \geq 1 \end{array} \right. \tag{4}$$

In a matrix form

$$(\Gamma(1), \ldots, \Gamma(p)) = (\Phi(1), \ldots, \Phi(p)) \begin{pmatrix} \Gamma(0) & \ldots & \Gamma(p-1) \\ \vdots & \vdots & \vdots \\ \Gamma(1-p) & \cdots & \Gamma(0) \end{pmatrix}$$

# Yule-Walker equation

Therefore, we have that

$$(\Phi(1),\ldots,\Phi(p)) = (\Gamma(1),\ldots,\Gamma(p)) \begin{pmatrix} \Gamma(0) & \cdots & \Gamma(p-1) \\ \vdots & \vdots & \vdots \\ \Gamma(1-p) & \cdots & \Gamma(0) \end{pmatrix}^{-1}$$

- Estimating

$$\widehat{\Gamma}(0) = \frac{1}{n}XX', \quad (\widehat{\Gamma}(1),\ldots,\widehat{\Gamma}(p)) = \frac{1}{n}YX'$$

  gives LSE estimator as expected.
- Plug-in $(\widehat{\Phi}_1,\ldots,\widehat{\Phi}_p)$ to (4) gives $\widehat{\Sigma}^{YW}$.

# MLE for VAR($p$)

Assume that $\mathbf{Z}_t \sim \mathcal{N}(0, \Sigma)$ Then, $Y^0 = AX + Z$ is a Gaussian also. From a vectored version

$$\boxed{\widetilde{Y} = \text{vec}(Y^0) = (X' \otimes I_m)\alpha + \text{vec}(Z)}$$

$\text{vec}(Z) \sim MVN(0, I_n \otimes \Sigma)$ implies that the likelihood of $\widetilde{Y} = \text{vec}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ becomes

$$L(\alpha; \Sigma) = |2\pi(I_n \otimes \Sigma)|^{-1/2} \times$$

$$\exp\left(-\frac{1}{2}\big(\tilde{Y} - (X' \otimes I_m)\alpha\big)'(I_n \otimes \Sigma)^{-1}\big(\tilde{Y} - (X' \otimes I_m)\alpha\big)\right)$$

# MLE for VAR($p$)

Then, the log-likelihood $\ell(\alpha;\Sigma) := \log L(\alpha;\Sigma)$ becomes

$$-\frac{1}{2}\log|2\pi(I_n\otimes\Sigma)| - \frac{1}{2}(\tilde{Y}-(X'\otimes I_m)\alpha)'(I_n\otimes\Sigma)^{-1}(\tilde{Y}-(X'\otimes I_m)\alpha)$$

$$= -\frac{nm}{2}\log 2\pi - \frac{n}{2}\log|\Sigma|$$

$$-\frac{1}{2}(\tilde{Y}-(X'\otimes I_m)\alpha)'(I_n\otimes\Sigma)^{-1}(\tilde{Y}-(X'\otimes I_m)\alpha).$$

- Now, we will fine MLE by solving 'derivative = 0'.
- Useful formula is

$$\begin{aligned} \frac{\partial(y-X\beta)'\Omega(y-X\beta)}{\partial\beta} &= -2X'\Omega(y-X\beta); \quad \Omega \text{ is symm} \\ \frac{\partial\log|A|}{\partial A} &= (A')^{-1} \end{aligned}$$

# MLE for VAR($p$)

- Note that

$$\frac{\partial \ell}{\partial \alpha} = (X' \otimes I_n)'(I_n \otimes \Sigma)^{-1}(\tilde{Y} - (X' \otimes I_m)\alpha)$$
$$= (X \otimes I_m)(I_n \otimes \Sigma^{-1})(\tilde{Y} - (X' \otimes I_m)\alpha)$$
$$= (X \otimes \Sigma^{-1})\tilde{Y} - (XX' \otimes \Sigma^{-1})\alpha$$
$$\frac{\partial \ell}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}(Y^0 - AX)(Y^0 - AX)'\Sigma^{-1}$$

- Equating to zero gives

$$\widehat{\Sigma}^M = \frac{1}{n}(Y^0 - \widehat{A}X)(Y^0 - \widehat{A}X)'$$
$$\widehat{\alpha}^M = (XX' \otimes \Sigma^{-1})^{-1}(X \otimes \Sigma^{-1})\widetilde{Y}$$
$$= ((XX')^{-1} \otimes \Sigma)(X \otimes \Sigma^{-1})\widetilde{Y} = \left((XX')^{-1}X \otimes I_m\right)\widetilde{Y}$$

- Under the Gaussian assumption, LSE and MLE are the same as expected.

# Forecasting - BLP (best linear predictor)

- The BLP for $\mathbf{X}_{n+h}$ based on the observations $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ is obtained by minimizing the mean squared error

$$E(X_{n+h} - \widehat{\mathbf{X}}_{n+h})^2$$

amongst the linear estimator given by

$$\widehat{\mathbf{X}}_{n+h} := P_n \mathbf{X}_{n+h} = A_1 \mathbf{X}_n + A_2 \mathbf{X}_{n-1} + \ldots + A_n \mathbf{X}_1$$

- Hence, the solution is to solve

$$(\mathbf{X}_{n+h} - \widehat{\mathbf{X}}_{n+h}) \perp \mathbf{X}_i, \quad i = 1, \dots, n$$

or equivalently,

$$E(\mathbf{X}_{n+h} - \widehat{\mathbf{X}}_{n+h})\mathbf{X}_i' = \mathbf{0}, \quad i = 1, \dots, n$$

# Forecasting - BLP (best linear predictor)

- As a special case for VAR($p$), using a relationship

$$\mathbf{X}_{t+h} = \Phi_1 \mathbf{X}_{t+h-1} + \ldots + \Phi_p \mathbf{X}_{t+h-p} + \mathbf{Z}_{t+h},$$

  BLP is recursively calculated as follows.

$$\begin{aligned} \widehat{\mathbf{X}}_{n+1} &= \Phi_1 \mathbf{X}_n + \ldots + \Phi_p \mathbf{X}_1 \\ \widehat{\mathbf{X}}_{n+2} &= \Phi_1 \widehat{\mathbf{X}}_{n+1} + \ldots \Phi_p \mathbf{X}_2 \\ &\vdots \\ \widehat{\mathbf{X}}_{n+h} &= \Phi_1 \widehat{\mathbf{X}}_{n+h-1} + \ldots + \Phi_p X_{n+h-p} \end{aligned}$$

- It can be deduced, for example $h = 1$, $\mathbf{Z}_{n+1}$ is orthogonal to $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

# Forecasting - BLP (best linear predictor)

- MSPE can be calculated from linear process representation

$$\mathbf{X}_n = \sum_{j=0}^{\infty} \Psi_j \mathbf{Z}_{n-j},$$

which implies that

$$\mathbf{X}_{n+h} - \widehat{\mathbf{X}}_{n+h} = \sum_{j=0}^{h-1} \Psi_j \mathbf{Z}_{n+h-j}.$$

- Hence,

$$\mathrm{MSPE} = E(\mathbf{X}_{n+h} - \widehat{\mathbf{X}}_{n+h})(\mathbf{X}_{n+h} - \widehat{\mathbf{X}}_{n+h})' = \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j{}'.$$

- Recall for a univariate case, MSPE is given by $\sigma^2 \sum_{j=0}^{h-1} \psi_j^2$.

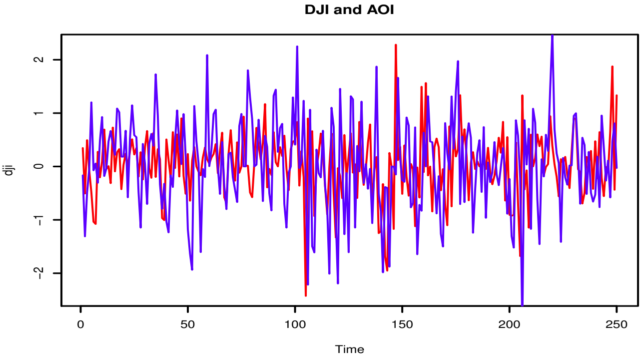
# VAR($p$) order selection by Information Criteria

- Similar to univariate case, order selection for multivariate VARMA models can be made by minimizing information criteria:

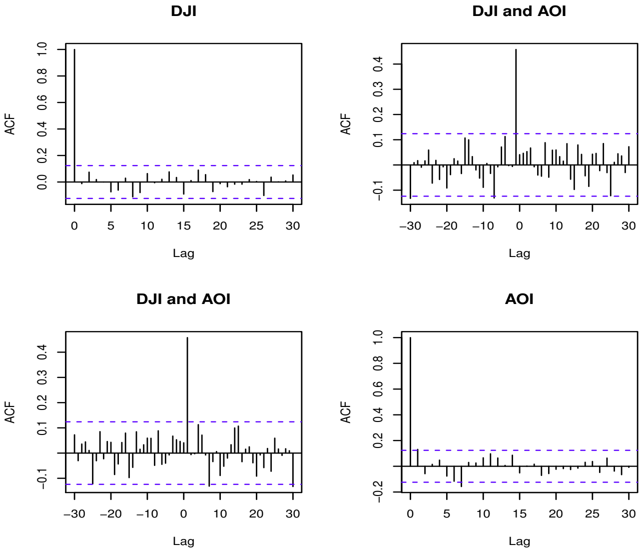
$$
\begin{aligned}
\mathrm{AIC}(p) &= \log|\widehat{\Sigma}^{MLE}(p)| + \frac{2(pm^2)}{n} \\
\mathrm{AICC}(p) &= \log|\widehat{\Sigma}^{MLE}(p)| + \frac{2(pm^2+1)nm}{nm - pm^2 - 2} \\
\mathrm{BIC}(p) &= \log|\widehat{\Sigma}^{MLE}(p)| + \frac{\log n(pm^2)}{n} \\
\mathrm{HQ}(p) &= \log|\widehat{\Sigma}^{MLE}(p)| + \frac{\log n(pm^2)}{n} \\
\mathrm{FPE}(p) &= \left(\frac{n+pm+1}{n-pm-1}\right)^m \left|\widehat{\Sigma}^{MLE}(p)\right| \approx e^{(\mathrm{AIC}(p)+2m/n)}
\end{aligned}
$$

# R: `vars` package

- For the estimation of VAR($p$) use `vars` package in R
- DJI an AOI example



# Example: DJI and AOI using `vars` package



# Example: DJI and AOI using vars package

- ```
  # for fixed order
  out.p2 = VAR(data, p=2, type="const")
  summary(out.p2)
  #If  you want to extract coefficients in a matrix form
  ```

  ```
  > Bcoef(out.p2)
           dji.l1      aoi.l1     dji.l2      aoi.l2      const
  dji -0.01288041 0.005136636  0.0704363  0.03423397 0.02733510
  aoi  0.67745517 0.155525997 -0.1043473 -0.09366776 0.03126549
  ```

- Model selection

  ```
  > VARselect(data, lag.max = 5, type="const")

  $selection
  AIC(n)  HQ(n)  SC(n) FPE(n)
       1      1      1      1

  $criteria
                  1          2          3          4          5
  AIC(n) -1.4744721 -1.4664093 -1.4388379 -1.4257781 -1.4164037
  HQ(n)  -1.4399426 -1.4088601 -1.3582691 -1.3221897 -1.2897956
  SC(n)  -1.3887270 -1.3235008 -1.2387660 -1.1685428 -1.1020050
  FPE(n)  0.2289001  0.2307552  0.2372106  0.2403373  0.2426142
  ```
  ```

# Example: DJI and AOI using vars package

```
> pred1 = predict(out.p1, n.ahead = 5, ci = 0.95)
> pred1
$dji
            fcst     lower    upper       CI
[1,] 0.006355922 -1.187387 1.200099 1.193743
[2,] 0.056813215 -1.138057 1.251683 1.194870
[3,] 0.030549489 -1.164624 1.225723 1.195173
[4,] 0.029179791 -1.165996 1.224356 1.195176
[5,] 0.028430939 -1.166746 1.223608 1.195177

$aoi
           fcst      lower    upper       CI
[1,] 0.91227178 -0.6131683 2.437712 1.525440
[2,] 0.12730445 -1.6072596 1.861869 1.734564
[3,] 0.07307632 -1.6641303 1.810283 1.737207
[4,] 0.04940296 -1.6880332 1.786839 1.737436
[5,] 0.04583077 -1.6916124 1.783274 1.737443
```

# Example: DJI and AOI using `vars` package

- A natural question is why we need to consider VAR instead of univariate model to each series.
- We can confirm that joint modelling gives smaller MSPE. (Details are on Example 7.6.3). For example, for AOI series, the best univariate model is AR(1) and MSPE for 1-step ahead prediction is .7572

```
Coefficients:
         ar1  intercept
      0.1309     0.0406
s.e.  0.0626     0.0633

sigma^2 estimated as 0.7572:  log likelihood = -319.98,  aic = 645.96
```

- However, if you use VAR(1) model, it reduces to .60575.

```
> summary(out.p1)

Covariance matrix of residuals:
        dji     aoi
dji 0.37096 0.02272
aoi 0.02272 0.60575
```

# Extension to VARMA$(p, q)$ model

- Now we consider estimating VARMA$(p, q)$ model give by

$$\mathbf{X}_t - \Phi_1 \mathbf{X}_{t-1} - \ldots - \Phi_p \mathbf{X}_{t-p} = \mathbf{Z}_t + \Theta_1 \mathbf{Z}_{t-1} + \ldots + \Theta_q \mathbf{Z}_{t-q}.$$

- As in the univariate ARMA$(p, q)$ case, we will consider orthogonalization (GS orthogonalization) to write down likelihood. For example,

$$\begin{aligned} u_1 &= \mathbf{X}_1, \\ u_2 &= \mathbf{X}_2 - \widehat{\mathbf{X}}_2 = \mathbf{X}_2 - P_1(\mathbf{X}_2), \\ u_3 &= \mathbf{X}_3 - \widehat{\mathbf{X}}_3 = \mathbf{X}_3 - P_{1,2}(\mathbf{X}_3) \\ &\vdots \end{aligned}$$

- Then, $\{u_j := \mathbf{X}_j - \widehat{\mathbf{X}}_j,\ j = 1, .., n\}$ has the same likelihood with $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$, but uncorrelated with variance

$$v_{j-1} := \mathrm{Var}(u_j) = E(u_j u_j').$$

# Extension to VARMA$(p, q)$ model

- Therefore, the likelihood of $\{u_j\}$ is given by

$$(2\pi)^{-\frac{nm}{2}} \sum_{j=1}^{n} |v_{j-1}|^{-\frac{1}{2}} \exp\Bigg( -\frac{1}{2} \sum_{j=1}^{n} u_j{}' v_{j-1}{}^{-1} u_j \Bigg)$$

- As $j \to \infty$

$$v_{j-1} \to \Psi_0 \Sigma \Psi_0{}' = \Sigma$$

since $v_{j-1}$ is one-step ahead MSPE and as $j \to \infty$, it becomes the prediction based on the infinite past.

- Thus, in practice we maximize

$$\ell(\theta, \Sigma) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{j=1}^{n} (\mathbf{X}_j - \widehat{\mathbf{X}}_j)' \Sigma^{-1} (\mathbf{X}_j - \widehat{\mathbf{X}}_j)$$

# MLE of VARMA$(p, q)$

- First, note that

$$\frac{\partial \ell}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{X}_j - \widehat{\mathbf{X}}_j)(\mathbf{X}_j - \widehat{\mathbf{X}}_j)'\right)\Sigma^{-1}.$$

- Hence, solving '$\partial\ell/\partial\Sigma = 0$' gives

$$\tilde{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \widehat{\mathbf{X}}_j)(\mathbf{X}_j - \widehat{\mathbf{X}}_j)'$$

- By plug-in $\tilde{\Sigma}$ to likelihood function, profile likelihood becomes

$$\begin{aligned}\ell(\theta, \tilde{\Sigma}) &= -\frac{n}{2}\log|\tilde{\Sigma}| - \frac{1}{2}\sum_{j=1}^{n}(\mathbf{X}_j - \widehat{\mathbf{X}}_j)'\tilde{\Sigma}^{-1}(\mathbf{X}_j - \widehat{\mathbf{X}}_j)\\ &= -\frac{n}{2}\log|\tilde{\Sigma}| - \frac{1}{2}\mathrm{tr}\left(\widetilde{\Sigma}^{-1}\sum_{j=1}^{n}(\mathbf{X}_j - \widehat{\mathbf{X}}_j)(\mathbf{X}_j - \widehat{\mathbf{X}}_j)'\right)\end{aligned}$$

because $a'a = \mathrm{tr}(aa')$.

# MLE of VARMA$(p, q)$

- Thus, it further reduces to

$$-\frac{n}{2}\log|\tilde{\Sigma}| - \frac{Tm}{2}.$$

  Therefore, it is equivalent to minimize

$$\log|\tilde{\Sigma}|$$

- Numerical optimisation is carried over by iterating
  i) $\Sigma^{(i+1)} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \widehat{\mathbf{X}}_j^{(i)})(\mathbf{X}_j - \widehat{\mathbf{X}}_j^{(i)})'$
  ii) $\widehat{\theta}^{(i+1)} = \underset{\theta}{\operatorname{argmin}} \log\left|\widetilde{\Sigma}^{(i+1)}\right|$

  with some initial estimates such as LSE.

# VARIMA$(p, d, q)$ model

- Recall that nonstationary univariate time series can frequently be made stationary by applying the differencing operator $\nabla = 1 - B$.
- In a multivariate setting, we can define component-wise differencing, that is

$$\nabla \mathbf{X}_t = \left(X_{t1} - X_{(t-1)1}, \ldots, X_{tm} - X_{(t-1)m}\right)'$$

- VARIMA model is defined as

$$\Psi(B)D(B)\mathbf{X}_t = \Theta(B)\mathbf{Z}_t,$$

where $D(z) = 1 - d_1 z - \ldots - d_r z^d$ is the $d$-th order differencing polynomial.

- That is,

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Psi_j{}^* \mathbf{Z}_{t-j}, \quad \Psi_j{}^*(B) = D(B)^{-1}\Psi^{-1}(B)\Theta(B)$$