# Lecture 11: Time Series Analysis II

MIT 18.S096

Dr. Kempthorne

Fall 2013

# Outline

# Multivariate Time Series

Let $\{\mathbf{X}_t\} = \{\ldots, \mathbf{X}_{t-1}, \mathbf{X}_t, \mathbf{X}_{t+1}, \ldots\}$ be an $m-$dimensional stochastic process consisting of random $m-$vectors

$$\mathbf{X}_t = (X_{1,t}, X_{2,t}, \ldots, X_{m,t})', \text{ a random vector on } \mathcal{R}^m.$$

$\{\mathbf{X}_t\}$ consists of $m$ component time series:

$$\{X_{1,t}\}, \{X_{2,t}\}, \ldots, \{X_{m,t}\}.$$

$\{\mathbf{X}_t\}$ is **Covariance Stationary** if every component time series is covariance stationary.

**Multivariate First and Second-Order Moments:**

$$\boldsymbol{\mu} = E[\mathbf{X}_t] = \begin{bmatrix} E(X_{1,t}) \\ E(X_{2,t}) \\ \vdots \\ E(X_{m,t}) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{bmatrix} \quad (m \times 1)\text{-vector}$$

# Second-Order Moments of Multivariate Time Series

## Variance/Covariance Matrix

$$\mathbf{\Gamma}_0 = Var(\mathbf{X}_t) = E[(X_t - \mu)(X_t - \mu)'],$$

$$= \begin{bmatrix} var(X_{1,t}) & cov(X_{1,t}, X_{2,t}) & \cdots & cov(X_{1,t}X_{m,t}) \\ cov(X_{2,t}, X_{1,t}) & var(X_{2,t}) & \cdots & cov(X_{2,t}, X_{m,t}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_{m,t}, X_{1,t}) & cov(X_{m,t}, X_{2,t}) & \cdots & var(X_{m,t}) \end{bmatrix}$$

## Correlation Matrix

$$\mathbf{R}_0 = corr(\mathbf{X}_t) = \mathbf{D}^{-\frac{1}{2}}\mathbf{\Gamma}_0\mathbf{D}^{-\frac{1}{2}}, \quad \text{where } \mathbf{D} = diag(\mathbf{\Gamma}_0)$$

# Second-Order Cross Moments

**Cross-Covariance Matrix (lag-k)**

$$\mathbf{\Gamma}_k = Cov(\mathbf{X}_t, \mathbf{X}_{t-k}) = E[(X_t - \mu)(X_{t-k} - \mu)'],$$

$$= \begin{bmatrix} cov(X_{1,t}, X_{1,t-k}) & cov(X_{1,t}, X_{2,t-k}) & \cdots & cov(X_{1,t}X_{m,t-k}) \\ cov(X_{2,t}, X_{1,t-k}) & cov(X_{2,t}, X_{2,t-k}) & \cdots & cov(X_{2,t}, X_{m,t-k}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_{m,t}, X_{1,t-k}) & cov(X_{m,t}, X_{2,t-k}) & \cdots & cov(X_{m,t}, X_{m,t-k}) \end{bmatrix}$$

**Cross-Correlation Matrix (lag-k)**

$$\mathbf{R}_k = corr(\mathbf{X}_t) = \mathbf{D}^{-\frac{1}{2}}\mathbf{\Gamma}_k\mathbf{D}^{-\frac{1}{2}}, \quad \text{where } \mathbf{D} = diag(\mathbf{\Gamma}_0)$$

**Properties**

- $\mathbf{\Gamma}_0$ and $\mathbf{R}_0$: $m \times m$ symmetric matrices
- $\mathbf{\Gamma}_k$: and $\mathbf{R}_k$: $m \times m$ matrices, but **not** symmetric
  $$\mathbf{\Gamma}_k = \mathbf{\Gamma}_{-k}^T.$$

# Second-Order Cross Moments (continued)

**Properties**

- If $[\mathbf{\Gamma}_k]_{j,j*} = Cov(X_{t,j}, X_{t-k,j*}) \neq 0$, for some $k > 0$, we say "$\{X_{t,j*}\}$ **leads** $\{X_{t,j}\}$".
- If "$\{X_{t,j*}\}$ **leads** $\{X_{t,j}\}$" and "$\{X_{t,j}\}$ **leads** $\{X_{t,j*}\}$ then there is **feedback**.

# Outline

# Multivariate Wold Decomposition

**Wold Representation Theorem:** Any multivariate covariance stationary time series $\{\mathbf{X}_t\}$ ($m-$variate) can be decomposed as

$$\begin{aligned} X_t &= \mathbf{V}_t + \boldsymbol{\eta}_t + \boldsymbol{\Psi}_1\boldsymbol{\eta}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\eta}_{t-2} + \cdots \\ &= \mathbf{V}_t + \textstyle\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\eta}_{t-k} \end{aligned}$$

where:

- $\{\mathbf{V}_t\}$ is an $m-$dimensional linearly deterministic process.
- $\{\boldsymbol{\eta}_t\}$ is multivariate white noise process, i.e.,

$$\begin{aligned} &E[\boldsymbol{\eta}_t] = \mathbf{0}_m \quad (m \times 1) \\ &Var[\boldsymbol{\eta}_t] = E[\boldsymbol{\eta}_t\boldsymbol{\eta}_t^T] = \boldsymbol{\Sigma}, \ (m \times m) \text{ positive semi-definite}) \\ &Cov[\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-k}] = E[\boldsymbol{\eta}_t\boldsymbol{\eta}_{t-k}^T] = \mathbf{0}, \ \forall k \neq 0 \ (m \times m) \\ &Cov[\boldsymbol{\eta}_t, \mathbf{V}_{t-k}] = \mathbf{0} \ \forall k \quad (m \times m) \end{aligned}$$

- The terms $\{\boldsymbol{\Psi}_k\}$ are $m \times m$ matrices such that

$$\boldsymbol{\Psi}_0 = \mathbf{I}_m \text{ and } \textstyle\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\Psi}_k^T \text{ converges.}$$

# Outline

# Vector Autoregressive (VAR) Processes

The $m$-dimensional multivariate time series $\{\mathbf{X}_t\}$ follows the $VAR(p)$ model with auto-regressive order $p$ if

$$\mathbf{X}_t = C + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \boldsymbol{\eta}_t$$

where

$C = (c_1, c_2, \ldots, c_m)'$ is an $m$-vector of constants.
$\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_p$ are $(m \times m)$ matrices of coefficients
$\{\boldsymbol{\eta}_t\}$ is multivariate white noise $MVN(\mathbf{0}_m, \mathbf{\Sigma})$

For fixed $j$, the component series $\{X_{t,j}, t \in \mathcal{T}\}$ is a generalization of the $AR(p)$ model for the $j$th component series to include lag-regression terms on all other component series:

$$\begin{aligned} X_{j,t} &= c_j + \sum_{k=1}^{p} [\mathbf{\Phi}_k]_{j,k} X_{j,t-k} + \\ &\quad \sum_{j* \neq j} \left[ \sum_{k=1}^{p} [\mathbf{\Phi}_k]_{j*,k} X_{j*,t-k} \right] \end{aligned}$$

# VAR(1) Representation of VAR(p) Process

A $VAR(p)$ process is equivalent to a $VAR(1)$ process.

**Define**

$$\begin{aligned} \mathbf{Z}_t &= (\mathbf{X}'_t, \mathbf{X}'_{t-1}, \ldots \mathbf{X}'_{t-p+1})' \\ \mathbf{Z}_{t-1} &= (\mathbf{X}'_{t-1}, \mathbf{X}'_{t-2}, \ldots \mathbf{X}'_{t-p})' \end{aligned}$$

The $(mpx1)$ multivariate time series process $\{\mathbf{Z}_t\}$ satistifes

$$\mathbf{Z}_t = D + A\mathbf{Z}_{t-1} + F$$

where $D$ and $F$ are $(mp \times 1)$ and $A$ is $(mp \times mp)$:

$$D = \begin{bmatrix} C \\ 0_m \\ 0_m \\ \vdots \\ 0_m \\ 0_m \end{bmatrix}, A = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_3 & \cdots & \cdots & \boldsymbol{\Phi}_p \\ \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{I}_m & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_m & \mathbf{0} \end{bmatrix}, F = \begin{bmatrix} \boldsymbol{\eta}_t \\ 0_m \\ 0_m \\ \vdots \\ 0_m \\ 0_m \end{bmatrix}$$

# Stationary VAR(p) Process

A $VAR(p)$ model is **stationary** if either

- All eigen values of the companion matrix $A$ have modulus less than 1, or
- All roots of: $det(\mathbf{I}_m - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots - \mathbf{\Phi}_p z^p) = 0$ as a function of the complex variable $z$, are outside the complex unit circle $|z| \leq 1$.

**Mean of Stationary VAR(p) Process**

For the expression of the $VAR(p)$ model:

$$\mathbf{X}_t = C + \mathbf{\Phi}_1 \mathbf{X}_{t-1} + \mathbf{\Phi}_2 \mathbf{X}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{X}_{t-p} + \eta_t$$

take expectations:

# Mean of Stationary VAR(p) Process

$$\begin{aligned} E[\mathbf{X}_t] &= C + \mathbf{\Phi}_1 E[\mathbf{X}_{t-1}] + \mathbf{\Phi}_2 E[\mathbf{X}_{t-2}] + \cdots + \mathbf{\Phi}_p E[\mathbf{X}_{t-p}] + E[\eta_t] \\ \boldsymbol{\mu} &= C + \sum_{k=1}^{p} [\mathbf{\Phi}_k]\boldsymbol{\mu} + \mathbf{0}_m \end{aligned}$$

$$\implies E[\mathbf{X}_t] = \boldsymbol{\mu} = (I - \mathbf{\Phi}_1 - \cdots - \mathbf{\Phi}_p)^{-1} C.$$

Also

$$\implies C = (I - \mathbf{\Phi}_1 - \cdots - \mathbf{\Phi}_p)\boldsymbol{\mu}$$

$$\begin{aligned} [\mathbf{X}_t - \boldsymbol{\mu}] &= \mathbf{\Phi}_1[\mathbf{X}_{t-1} - \boldsymbol{\mu}] + \mathbf{\Phi}_2[\mathbf{X}_{t-2} - \boldsymbol{\mu}] + \cdots \\ &\quad + \mathbf{\Phi}_p[\mathbf{X}_{t-p} - \boldsymbol{\mu}] + \eta_t \end{aligned}$$

# $VAR(p)$ Model as System of Regression Equations

Consider observations from the $m-$dimensional multivariate time series $\{\mathbf{X}_t\}$ consisting of

- $n$ sample observations:

$$\mathbf{x}_t = (x_{1,t}, x_{2,t}, \dots x_{m,t})', t = 1, 2, \dots, n$$

- $p$ initial conditions expressed as pre-sample observations:

$$\mathbf{x}_{p-1}, \dots, \mathbf{x}_{-1}, \mathbf{x}_0$$

Set up $m$ regression models corresponding to each component $j$ of the $m$-variate time series:

$$\mathbf{y}^{(j)} = \mathbf{Z}\beta^{(j)} + \epsilon^{(j)}, \qquad j = 1, 2, \dots, m$$

where:

- $\mathbf{y}^{(j)} = \begin{bmatrix} x_{j,1} \\ x_{j,2} \\ \vdots \\ x_{j,n} \end{bmatrix}$, $\mathbf{Z} = \begin{bmatrix} 1 & \mathbf{z}_0' \\ 1 & \mathbf{z}_1' \\ \vdots & \vdots \\ 1 & \mathbf{z}_{n-1}' \end{bmatrix}$ with $\mathbf{z}_{t-1} = (\mathbf{x}_{t-1}', \mathbf{x}_{t-2}', \dots, \mathbf{x}_{t-p}')'$.

# $VAR(p)$ Model as a Multivariate Regression Model

- $\boldsymbol{\beta}^{(j)}$ is the $(mp+1)$-vector of regression parameters for the $j$th component time series.
- $\boldsymbol{\epsilon}^{(j)}$ is the $n$-vector of innovation errors which are $WN(0, \sigma_j^2)$ with variance depending on the variate $j$.

There are $m$ Linear Regression Models:

$$\begin{aligned}
\mathbf{y}^{(1)} &= \mathbf{Z}\boldsymbol{\beta}^{(1)} + \boldsymbol{\epsilon}^{(1)} \\
\mathbf{y}^{(2)} &= \mathbf{Z}\boldsymbol{\beta}^{(2)} + \boldsymbol{\epsilon}^{(2)} \\
&\vdots \\
\mathbf{y}^{(m)} &= \mathbf{Z}\boldsymbol{\beta}^{(m)} + \boldsymbol{\epsilon}^{(m)};
\end{aligned}$$

these can be expressed together as one **Multivariate Regression Model**

$$\begin{aligned}
\left[\mathbf{y}^{(1)}\mathbf{y}^{(2)}\cdots\mathbf{y}^{(m)}\right] &= \mathbf{Z}\left[\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(2)}\cdots\boldsymbol{\beta}^{(m)}\right] + \left[\boldsymbol{\epsilon}^{(1)}\boldsymbol{\epsilon}^{(2)}\cdots\boldsymbol{\epsilon}^{(m)}\right] \\
\mathcal{Y} &= \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}
\end{aligned}$$

Form of model: *Seemingly Unrelated Regressions (SUR)*.

# Outline

## Component-Wise OLS Estimation of VAR(p) Model

- The parameters $\hat{\beta}^{(j)}$ are easily estimated by OLS, applying the same algorithm
$$\hat{\beta}^{(j)} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}^{(j)},\ j = 1, 2, \ldots, m$$
- The residuals $\hat{\epsilon}^{(j)}$ have the usual formula
$$\hat{\epsilon}^{(j)} = \mathbf{Z}\hat{\beta}^{(j)}$$
- Identify estimates of the $VAR(p)$ innovations $\{\eta_t\}$ ($m$-variate time series) as
$$\begin{bmatrix} \hat{\eta}_1' \\ \hat{\eta}_2' \\ \vdots \\ \hat{\eta}_n' \end{bmatrix} = \begin{bmatrix} \hat{\eta}_{1,1} & \hat{\eta}_{2,1} & \cdots & \hat{\eta}_{m,1} \\ \hat{\eta}_{1,2} & \hat{\eta}_{2,2} & \cdots & \hat{\eta}_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\eta}_{1,n} & \hat{\eta}_{2,n} & \cdots & \hat{\eta}_{m,n} \end{bmatrix} = \left[\hat{\epsilon}^{(1)}\hat{\epsilon}^{(2)}\cdots\hat{\epsilon}^{(m)}\right]$$
and define the unbiased estimate of the $(m \times m)$ innovation covariance matrix $\mathbf{\Sigma} = E[\eta_t\eta_t']$
$$\hat{\mathbf{\Sigma}} = \frac{1}{n-pm}\sum_{t=1}^{n}\hat{\eta}_t\hat{\eta}_t' = \frac{1}{n-pm}\mathcal{Y}^T(I_n - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T)\mathcal{Y}$$

# Outline

# Optimality of OLS Estimates

**Theorem:** For the $VAR(p)$ model where there are no restrictions on the coefficient matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$:

- The component-wise OLS estimates are equal to the GLS (generalized least squares) estimates accounting for the general case of innovation covariance matrix $\mathbf{\Sigma}$ $(m \times m)$ with possibly unequal component variances and non-zero correlations.
- Under the assumption that $\{\eta_t\}$ are i.i.d. multivariate Gaussian distribution $MN(\mathbf{0}_m, \mathbf{\Sigma})$, the component-wise OLS estimates are also the **maximum likelihood** estimates.

# Kronecker Products and the vec Operator

**Definition:** The **Kronecker Product** of the $(m \times n)$ matrix $A$ and the $(p \times q)$ matrix $B$ is the $(mp \times qn)$ matrix $C$, given by:

$$C = A \bigotimes B = \begin{bmatrix} a_{1,1}B & a_{1,2}B & \cdots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \cdots & a_{2,n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}B & a_{m,2}B & \cdots & a_{m,n}B \end{bmatrix}$$

**Properties:**

- $(A \bigotimes B)' = (A') \bigotimes (B')$
- $(A \bigotimes B)(D \bigotimes F) = (AD) \bigotimes (BF)$,
  (matrix $D$ has $n$ rows and $F$ has $q$ rows)

# The vec Operator

**Definition** The *vec* operator converts a rectangular matrix to a column vector by stacking the columns. For an $(n \times m)$ matrix $A$:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix}, \; vec(A) = \begin{bmatrix} \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \end{bmatrix} \\ \begin{bmatrix} a_{1,2} \\ \vdots \\ a_{n,2} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} a_{1,m} \\ \vdots \\ a_{n,m} \end{bmatrix} \end{bmatrix}$$

# Vectorizing the Multivariate Regression Model

Recall the **Multivariate Regression Model**

$$\begin{aligned}\left[\mathbf{y}^{(1)}\mathbf{y}^{(2)}\dots\mathbf{y}^{(m)}\right] &= \mathbf{Z}\left[\boldsymbol{\beta}^{(1)}\boldsymbol{\beta}^{(2)}\dots\boldsymbol{\beta}^{(m)}\right] + \left[\boldsymbol{\epsilon}^{(1)}\boldsymbol{\epsilon}^{(2)}\dots\boldsymbol{\epsilon}^{(m)}\right] \\ \mathcal{Y} &= \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}\end{aligned}$$

**Define**

$$\begin{aligned} y_* &= vec(\mathcal{Y}) & (nm \times 1) \\ X_* &= \mathbf{I}_m \bigotimes \mathbf{Z} & (nm \times (1+pm^2)) \\ \beta_* &= vec(\boldsymbol{\beta}) & ((1+pm^2)\times 1) \\ \epsilon_* &= vec(\boldsymbol{\epsilon}) & (nm \times 1)\end{aligned}$$

The model is given by:

$$y_* = X_*\beta_* + \epsilon_*,$$

where $\epsilon_* \sim WN(\mathbf{0}_{nm}, \mathbf{\Sigma}_*)$ with $\mathbf{\Sigma}_* = \mathbf{I}_n \bigotimes \Sigma$

# GLS Estimates of $\beta^*$

By the **Generalized Least Squares (GLS)** case of the Gauss-Markov Theorem, the following estimator is *BLUE*:

$$\hat{\beta}_* = [X_*^T \Sigma_*^{-1} X_*]^{-1}[X_*^T \Sigma_*^{-1} y_*]$$

- $$\begin{aligned} X_*^T \Sigma_*^{-1} X_* &= (I_m \otimes \mathbf{Z})^T (\mathbf{\Sigma}^{-1} \otimes I_n)(I_m \otimes \mathbf{Z}) \\ &= (I_m \otimes \mathbf{Z}^T)(\mathbf{\Sigma}^{-1} \otimes \mathbf{Z}) \\ &= \mathbf{\Sigma}^{-1} \otimes (\mathbf{Z}^T \mathbf{Z}) \end{aligned}$$

$$\Longrightarrow \quad [X_*^T \Sigma_*^{-1} X_*]^{-1} = [\mathbf{\Sigma}^{-1} \otimes (\mathbf{Z}^T \mathbf{Z})]^{-1} = [\mathbf{\Sigma} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1}]$$

- $$\begin{aligned} [X_*^T \Sigma_*^{-1} y_*] &= (I_m \otimes \mathbf{Z})^T (\mathbf{\Sigma}^{-1} \otimes I_n) y_* \\ &= (I_m \otimes \mathbf{Z}^T)(\mathbf{\Sigma}^{-1} \otimes I_n) y_* \\ &= (\mathbf{\Sigma}^{-1} \otimes \mathbf{Z}^T) y_* \end{aligned}$$

- $$\begin{aligned} \hat{\beta}_* &= [X_*^T \Sigma_*^{-1} X_*]^{-1}[X_*^T \Sigma_*^{-1} y_*] \\ &= [\mathbf{\Sigma} \otimes (\mathbf{Z}^T \mathbf{Z})^{-1}](\mathbf{\Sigma}^{-1} \otimes \mathbf{Z}^T) y_* \\ &= [I_m \otimes [(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] y_* = vec([(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathcal{Y}) \end{aligned}$$

# Outline

# Maximum-Likelihood Estimation of VAR(p) Models

For the Multivariate Regression Model representation of the VAR(p) model assume that the innovations are Gaussian:

$$y_* = X_* \beta_* + \epsilon_*, \text{ where } \epsilon_* \sim N_{nm}(\mathbf{0}_{nm}, \Sigma_*)$$
$$\text{where } \Sigma_* = I_n \bigotimes \Sigma.$$

The likelihood function is the conditional pdf $p(y_* \mid X_*, \beta_*, \Sigma_*)$ evaluated as a function of $(\beta_*, \Sigma)$ for given data $y_*$, (and $X_*$):

$$L(\beta_*, \Sigma) = \frac{1}{(2\pi)^{nm/2}} |\Sigma_*|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_* - X_*\beta_*)^T \Sigma_*^{-1} (y_* - X_*\beta_*)}$$

The log-likelihood function is

$$\begin{aligned}
\log\text{-}L(\beta_*, \Sigma) &= -\frac{nm}{2} log(2\pi) - \frac{1}{2} log(|\Sigma_*|) - \frac{1}{2}(y_* - X_*\beta_*)^T \Sigma_*^{-1}(y_* - X_*\beta_*) \\
&= -\frac{nm}{2} log(2\pi) - \frac{1}{2} log(|I_n \bigotimes \Sigma|) \\
&\qquad -\frac{1}{2}(y_* - X_*\beta_*)^T (I_n \bigotimes \Sigma^{-1})(y_* - X_*\beta_*) \\
&\propto -\frac{n}{2} log(|\Sigma|) - \frac{1}{2} trace[(\mathcal{Y} - Z\beta)\Sigma^{-1}(\mathcal{Y} - Z\beta)^T] \\
&\propto -\frac{n}{2} log(|\Sigma|) - \frac{1}{2} Q(\beta, \Sigma)
\end{aligned}$$

# Maximum-Likelihood Estimation of VAR(p) Models

The expression $Q(\beta, \Sigma)$ is the Generalized Least Squares criterion which is minimized by the component-by-component OLS estimates of $\beta$, for any non-singular covariance matrix $\Sigma$.

With $\hat{\beta}_* = vec(\hat{\beta})$, the MLE for $\Sigma$ minimizes the concentrated log likelihood: $l^*(\Sigma) =$log-$L(\hat{\beta}_*, \Sigma)$.

$$
\begin{aligned}
log\text{-}L(\hat{\beta}_*, \Sigma) &= -\tfrac{n}{2}log(|\Sigma|) - \tfrac{1}{2}Q(\hat{\beta}, \Sigma) \\
&= -\tfrac{n}{2}log(|\Sigma|) - \tfrac{1}{2}trace[(\mathcal{Y} - Z\hat{\beta})\Sigma^{-1}(\mathcal{Y} - Z\hat{\beta})^T] \\
&= -\tfrac{n}{2}log(|\Sigma|) - \tfrac{1}{2}trace[\Sigma^{-1}(\mathcal{Y} - Z\hat{\beta})^T(\mathcal{Y} - Z\hat{\beta})] \\
&= -\tfrac{n}{2}log(|\Sigma|) - \tfrac{n}{2}trace[\Sigma^{-1}\hat{\Sigma}]
\end{aligned}
$$

where $\hat{\Sigma} = \frac{1}{n}(\mathcal{Y} - Z\hat{\beta})^T(\mathcal{Y} - Z\hat{\beta})$.

Thereom: $\hat{\Sigma}$ is the mle for $\Sigma$; Anderson and Olkin (1979).

# Model Selection

Statistical model selection critera are used to select the order of the $VAR(p)$ process:

- Fit all $VAR(p)$ models with $0 \leq p \leq p_{max}$, for a chosen maximal order.
- Let $\tilde{\Sigma}(p)$ be the MLE of $\Sigma = E(\eta_t \eta_t')$, the covariance matrix of Gaussian $VAR(p)$ innovations.
- Choose $p$ to minimize one of:

  Akaike Information Criterion
  $$AIC(p) = -log(|\tilde{\Sigma}(p)|) + 2\frac{pm^2}{n}$$
  Bayes Information Criterion
  $$BIC(p) = -log(|\tilde{\Sigma}(p)|) + log(n)\frac{pm^2}{n}$$
  Hannan-Quinn Criterion
  $$HQ(p) = -log(|\tilde{\Sigma}(p)|) + 2log(log(n))\frac{pm^2}{n}$$

# Outline

# Asymptotic Distribution of Least-Squares Estimates

For a covariance-stationary $VAR(p)$ model, the least-squares estimates of the model coefficients are the least-squares coefficients of a covariance stationary linear model:

$$y_* = X_* \beta_* + \epsilon_*,$$

where $\epsilon_* \sim WN(\mathbf{0}_{nm}, \mathbf{\Sigma}_*)$ with $\mathbf{\Sigma}_* = \mathbf{I}_n \bigotimes \Sigma$

which arises from the vectorization of

$$\mathcal{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad (\mathcal{Y} \text{ and } \boldsymbol{\epsilon} \text{ are } (n \times m); \text{ and } \mathbf{Z} \text{ is } (n \times (mp+1)) \ )$$

If the white noise process $\{\eta_t\}$ underlying $\epsilon_*$ has finite and bounded 4-th order moments, and are independent over $t$, then:

- The $(mp+1) \times (mp+1)$ matrix
  $$\Gamma := plim \ \frac{\mathbf{Z}^T\mathbf{Z}}{n} \text{ exists and is non-singular.}$$
- The $(m(mp+1) \times 1)$ vector $\hat{\beta}_*$ is asymptotically jointly normally distributed:
  $$\sqrt{n}\left(\hat{\beta}_* - \beta_*\right) \xrightarrow{d} N(\mathbf{0}, \Sigma \bigotimes \Gamma^{-1})$$

If $n >> 0$ the following estimates are applied

- $\hat{\Gamma} = (\frac{1}{n})\mathbf{Z}^T\mathbf{Z}$
- $\hat{\Sigma} = (\frac{1}{n})\mathcal{Y}^T[I_n - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T]\mathcal{Y}$

Asymptotically, the least-squares estimates are distributed identically to the maximum-likelihood estimates for the model assuming Gaussian innovations.

18.S096 Topics in Mathematics with Applications in Finance
Fall 2013