Financial Econometrics I

Lecture 7

Fu Ouyang

December 22, 2018

Outline

Multivariate Time Series

Stationarity and Auto-Correlation Matrices

Vector Autoregressive Models

Estimating VAR Models

Multivariate Time Series Analysis

In practice, time series data on multiple subjects are recorded together.

- e.g., S&P 500 index, S&P 400 (MidCap) index and S&P 600 (SmallCap) index

Data is probably correlated not only over different times but also across different subjects. With multivariate time series, one can

- estimate the cross-subject correlation
- compare effects of a certain shock among different time series
- identify latent stationary properties (e.g., nonstationarity of single time series canceled out in a “system”)

In this lecture, we study d-vector time series $\mathbf{X}_t \equiv (X_{t,1}, ..., X_{t,d})'$, where $X_{t,j}$ is the j-th component series.

Outline

Multivariate Time Series

Stationarity and Auto-Correlation Matrices

Vector Autoregressive Models

Estimating VAR Models

Stationarity of Multivariate Time Series

Definition: Weak Stationarity

A multivariate time series $\{\mathbf{X}_t\}$ is *weakly stationary* (or simply *stationary*) if all its first and second moments are time-invariant, i.e.,

$$E(\mathbf{X}_t) = \mu = (\mu_1, ..., \mu_d)'$$

$$\mathbf{\Gamma}(k) \equiv E[(\mathbf{X}_{t+k} - \mu)(\mathbf{X}_t - \mu)'] = [\gamma_{ij}(k)]_{d \times d}, k = 0, \pm 1, \pm 2, ...$$

are independent of t.

Note that if $\{\mathbf{X}_t\}$ is stationary, then all its component series are stationary univariate time series. However, the converse is *not* necessarily true, which requires the covariances ($\gamma_{ij}(k)$ with $i \neq j$) between different series are also time-invariant.

Stationarity of Multivariate Time Series

Definition: Strict Stationarity

A multivariate time series $\{\mathbf{X}_t\}$ is *strictly stationary* if the joint distribution of $(\mathbf{X}_t, ..., \mathbf{X}_{t+k})$ is independent of t for all $k \geq 1$.

Strict stationarity $\Rightarrow$ weak stationarity provided that $E(\|\mathbf{X}_t\|^2) < \infty$.

In what follows, we assume $\{\mathbf{X}_t\}$ is weakly stationary (stationary) unless otherwise specified.

Auto-Covariance Matrix

The matrix-valued function $\mathbf{\Gamma}(\cdot)$ is called the *auto-covariance matrix function* or *cross variance function*.

The (i, j)-element of $\mathbf{\Gamma}(k)$, $\gamma_{ij}(k)$, is called the *cross covariance* between the i-th and j-th component series at time lag k for $i \neq j$, and *auto-covariance* for the j-th component series at time lag k for $i = j$, i.e.,

$$\gamma_{ij}(k) = \begin{cases} E[(X_{t+k,i} - \mu)(X_{t,j} - \mu)] = Cov(X_{t+k,i}, X_{t,j}) & i \neq j \\ E[(X_{t+k,i} - \mu)(X_{t,i} - \mu)] = Cov(X_{t+k,i}, X_{t,i}) & i = j \end{cases}$$

Auto-Covariance Matrix

Note that

- Auto-covariance is symmetric, i.e., $\gamma_{jj}(-k) = \gamma_{jj}(k)$.
- Cross covariance is *not* necessarily symmetric, i.e., typically, $\gamma_{ij}(k) \neq \gamma_{ij}(-k)$.
- $\mathbf{\Gamma}(k)$ is typically *not* a symmetric matrix for $k \neq 0$, i.e., $\gamma_{ij}(k) \neq \gamma_{ji}(k)$ unless $k = 0$.
- But $\mathbf{\Gamma}(-k) = \mathbf{\Gamma}(k)'$ as

$$
\begin{aligned}
\gamma_{ij}(k) &= E[(X_{t+k,i} - \mu)(X_{t,j} - \mu)] \\
&= E[(X_{\tau,i} - \mu)(X_{\tau-k,j} - \mu)] \; (\tau \equiv t + k\ ) \\
&= E[(X_{t,i} - \mu)(X_{t-k,j} - \mu)] \\
&= \gamma_{ji}(-k)
\end{aligned}
$$

Auto-Correlation Matrix

The *auto-correlation matrix*, which is also called the *cross correlation matrix*, of $\mathbf{X}_t$ is defined as

$$\mathbf{R}(k) = [\rho_{ij}(k)]_{d\times d} = \mathbf{D}^{-1/2}\mathbf{\Gamma}(k)\mathbf{D}^{-1/2}$$

where $\mathbf{D} = Diag(\gamma_{jj}(0))_{d\times d}$. Then the (i,j)-element of $\mathbf{R}(k)$ is

$$\rho_{ij}(k) = \frac{\gamma_{ij}(k)}{\sqrt{\gamma_{ii}(0)\gamma_{jj}(0)}} = Corr(X_{t+k,i}, X_{t,j})$$

which is called the cross correlation coefficient between the i-th and j-th component series at time lag k.

Auto-Correlation Matrix

Note that

- Auto-correlation $\rho_{jj}(k)$ is symmetric, i.e., $\rho_{jj}(k) = \rho_{jj}(-k)$.
- Cross correlation is not necessarily symmetric, i.e., typically, $\rho_{ij}(k) \neq \rho_{ij}(-k)$.
- In general, $\rho_{ij}(k) \neq \rho_{ji}(k)$ and $\mathbf{R}(k) \neq \mathbf{R}(k)'$.
- However, $\rho_{ij}(k) = \rho_{ji}(-k)$ and hence $\mathbf{R}(-k) = \mathbf{R}(k)'$.

Vector White Noise and Moving Average Processes

A vector series $\{\epsilon_t\}$ is *vector white noise*, denoted by $\epsilon_t \sim \mathrm{WN}(\mathbf{a}, \mathbf{\Sigma}_\epsilon)$, if

$$E(\epsilon_t) = \mathbf{a}, Var(\epsilon_t) = \mathbf{\Sigma}_\epsilon, Cov(\epsilon_t, \epsilon_s) = 0, \forall s \neq t$$

In practice, $\mathbf{a}$ is often normalized to be a vector of all zeros.

A basic stationary multivariate time series is the *vector moving average* process (with order q) defined as

$$\mathbf{X}_t = \mu + \epsilon_t + \mathbf{B}_1 \epsilon_{t-1} + \cdots + \mathbf{B}_q \epsilon_{t-q}$$

where $\mathbf{B}_1, ..., \mathbf{B}_q$ are coefficient matrices and $\epsilon_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$. We denote $\mathbf{X}_t \sim \mathrm{MA}(q)$.

Vector White Noise and Moving Average Processes

It is easy to verify that $E(\mathbf{X}_t) = \mu$ and

$$\mathbf{\Gamma}(k) = (\mathbf{B}_k \mathbf{\Sigma}_\epsilon \mathbf{I}_d + \mathbf{B}_{k+1} \mathbf{\Sigma}_\epsilon \mathbf{B}_1' + \cdots + \mathbf{B}_q \mathbf{\Sigma}_\epsilon \mathbf{B}_{q-k}') \mathbf{1}[0 \leq k \leq q]$$

- Vector moving average process is stationary for all q.
- $\mathbf{\Gamma}(k)$ and $\mathbf{R}(k)$ cut off at $k = q$.
- Overparameterization: the number of unknown coefficients is $O(d^2)$, need regularization or dimension-reduction in practice.

Sample Cross-Covariance/Correlation Matrices

With observations $\mathbf{X}_1, ..., \mathbf{X}_T$, we can calculate the sample cross-covariance matrix

$$\widehat{\mathbf{\Gamma}}(k) \equiv [\widehat{\gamma}_{ij}(k)]_{d\times d} = \frac{1}{T}\sum_{t=1}^{T-k}(\mathbf{X}_{t+k} - \widehat{\mu})(\mathbf{X}_t - \widehat{\mu})'$$

where

$$\widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t$$

Accordingly, the sample cross-correlation matrix can be obtained as

$$\widehat{\mathbf{R}}(k) \equiv [\widehat{\rho}_{ij}(k)]_{d\times d} = \widehat{\mathbf{D}}^{-1/2}\widehat{\mathbf{\Gamma}}(k)\widehat{\mathbf{D}}^{-1/2}$$

where $\widehat{\mathbf{D}} = Diag(\widehat{\gamma}_{jj}(0))_{d\times d}$.

Illustration with Real Data

Example: $\mathbf{X}_t$ = daily returns of (S&P 500, S&P 400, S&P 600)$'$

See R markdown.

Outline

Multivariate Time Series

Stationarity and Auto-Correlation Matrices

Vector Autoregressive Models

Estimating VAR Models

Vector Autoregressive Models

It is natural to extend univariate ARMA models to multivariate cases.

A d-vector autoregressive model with order p is of the form

$$\mathbf{X}_t = \mathbf{c} + \mathbf{A}_1\mathbf{X}_{t-1} + \cdots + \mathbf{A}_p\mathbf{X}_{t-p} + \epsilon_t$$

where $\epsilon_t \sim \text{WN}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$, $\mathbf{c}$ is a $d \times 1$ vector, $\mathbf{A}_1, \cdots, \mathbf{A}_p$ are $d \times d$ coefficient matrices. Then, we say $\mathbf{X}_t \sim \text{VAR}(p)$.

Example: VAR(1) Model

Example: VAR(1) Model

Consider the following model:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \epsilon_t$$

By recursive substitution, we have

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \epsilon_t = \cdots = \epsilon_t + \sum_{j=1}^{\infty} \mathbf{A}^j \epsilon_{t-j}$$

provided $\mathbf{A}^j \to \mathbf{0}_{d\times d}$ as $j \to \infty$. Then $\mathbf{X}_t$ has a MA(∞) representation and hence weakly stationary (why?).

Let $\mathbf{\Lambda}$ be the Jordan norm form of $\mathbf{A}$, i.e., $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$, where $\mathbf{P}$ is an invertible matrix and $\mathbf{\Lambda}$ is an "almost" diagonal matrix.

Example: VAR(1) Model

Example: VAR(1) Model

Then

$$\mathbf{A}^j = \mathbf{P}\mathbf{\Lambda}^j\mathbf{P}^{-1}$$

Hence, $\mathbf{A}^j$ converges to 0 as $j \to \infty$ if and only if all eigenvalues of $\mathbf{A}$ are within the unit circle.

By multiplying $\mathbf{X}_{t-k}$ on both sides and taking expectations, we have

$$\mathbf{\Gamma}(0) = \mathbf{A}\mathbf{\Gamma}(1)' + \mathbf{\Sigma}_\epsilon$$

$$\mathbf{\Gamma}(k) = \mathbf{A}\mathbf{\Gamma}(k-1), k \geq 1$$

where the second equation is the Yule-Walker equation. Hence, $(\mathbf{A}, \mathbf{\Sigma}_\epsilon)$ and $\{\mathbf{\Gamma}(k), k = 0, 1, 2, ...\}$ uniquely determine each other.

Identification Issue

Consider the following VAR(1) model

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \mathbf{A}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Hence, the VAR(1) model is also an MA(1) model

$$\begin{pmatrix} X_{t,1} \\ X_{t,2} \end{pmatrix} = \begin{pmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{t-1,1} \\ \epsilon_{t-1,2} \end{pmatrix}$$

This won't happen in univariate cases. To avoid this identification issue, we often consider autoregressive models only.

General VAR(p) Models

General VAR(p) Models

A VAR(p) model

$$\mathbf{X}_t = \mathbf{c} + \mathbf{A}_1\mathbf{X}_{t-1} + \cdots + \mathbf{A}_p\mathbf{X}_{t-p} + \epsilon_t$$

is weakly stationary if $|\mathbf{I}_d - \mathbf{A}_1 x - \cdots - \mathbf{A}_p x^p| \neq 0$ for all complex x with $|x| \leq 1$. Then

$$E(\mathbf{X}_t) = (\mathbf{I}_d - \mathbf{A}_1 - \cdots - \mathbf{A}_p)^{-1}\mathbf{c}$$

$$\mathbf{\Gamma}(0) = \mathbf{A}_1\mathbf{\Gamma}(1)' + \cdots + \mathbf{A}_p\mathbf{\Gamma}(p)' + \mathbf{\Sigma}_\epsilon$$

and the Yule-Walker equation is

$$\mathbf{\Gamma}(k) = \mathbf{A}_1\mathbf{\Gamma}(k-1) + \cdots + \mathbf{A}_p\mathbf{\Gamma}(k-p), k = 1, 2, \ldots$$

Outline

Multivariate Time Series

Stationarity and Auto-Correlation Matrices

Vector Autoregressive Models

Estimating VAR Models

Least Squares Estimation (LSE)

Consider a VAR(2) model:

$$\mathbf{X}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2} + \epsilon_t$$

where $\epsilon_t \sim \mathrm{WN}(\mathbf{0}, \mathbf{\Sigma}_\epsilon)$. We want to estimate $\theta \equiv (\mathbf{c}, \mathbf{A}_1, \mathbf{A}_2)$ and $\mathbf{\Sigma}_\epsilon$.

The least squares estimation (LSE) is easy to implement. Note that for each component series $X_{t,j}$, we have

$$X_{t,i} = c_i + \mathbf{X}'_{t-1}\mathbf{a}_i^{(1)} + \mathbf{X}'_{t-2}\mathbf{a}_i^{(2)} + \epsilon_{t,i}$$

where $\mathbf{a}_i^{(j)}$ is the i-th row of $\mathbf{A}_j$.

Least Squares Estimation (LSE)

Let $\theta_i \equiv (c_i, \mathbf{a}_i^{(1)}, \mathbf{a}_i^{(2)})$. Then for each $i \in \{1, ..., d\}$,

$$\widehat{\theta}_i = \arg\min_{\theta_i} \sum_{t=3}^{T} \left[X_{t,i} - c_i - \mathbf{X}'_{t-1}\mathbf{a}_i^{(1)} - \mathbf{X}'_{t-2}\mathbf{a}_i^{(2)} \right]^2$$

Stacking all $\widehat{\theta}_i$ gives $\widehat{\theta} = (\widehat{\mathbf{c}}, \widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2)$. Then the estimator for $\boldsymbol{\Sigma}_\epsilon$ can be defined as

$$\widehat{\boldsymbol{\Sigma}}_\epsilon = \frac{1}{T-2} \sum_{t=3}^{T} \widehat{\epsilon}_t \widehat{\epsilon}'_t$$

where

$$\widehat{\epsilon}_t = \mathbf{X}_t - \widehat{\mathbf{c}} - \widehat{\mathbf{A}}_1 \mathbf{X}_{t-1} - \widehat{\mathbf{A}}_2 \mathbf{X}_{t-2}$$

Yule-Walker Estimation (YWE)

$(\mathbf{A}_1, \mathbf{A}_2)$ can be estimation via Yule-Walker equations:

$$\mathbf{\Gamma}(1) = \mathbf{A}_1\mathbf{\Gamma}(0) + \mathbf{A}_2\mathbf{\Gamma}(-1) = \mathbf{A}_1\mathbf{\Gamma}(0) + \mathbf{A}_2\mathbf{\Gamma}(1)'$$

$$\mathbf{\Gamma}(2) = \mathbf{A}_1\mathbf{\Gamma}(1) + \mathbf{A}_2\mathbf{\Gamma}(0)$$

or equivalently

$$\begin{pmatrix} \mathbf{A}_1, & \mathbf{A}_2 \end{pmatrix} \begin{pmatrix} \mathbf{\Gamma}(0) & \mathbf{\Gamma}(1) \\ \mathbf{\Gamma}(1)' & \mathbf{\Gamma}(0) \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}(1), & \mathbf{\Gamma}(2) \end{pmatrix}$$

Hence, applying the analog principle, we obtain

$$\begin{pmatrix} \widehat{\mathbf{A}}_1 & \widehat{\mathbf{A}}_2 \end{pmatrix} = \begin{pmatrix} \widehat{\mathbf{\Gamma}}(1) & \widehat{\mathbf{\Gamma}}(2) \end{pmatrix} \begin{pmatrix} \widehat{\mathbf{\Gamma}}(0) & \widehat{\mathbf{\Gamma}}(1) \\ \widehat{\mathbf{\Gamma}}(1)' & \widehat{\mathbf{\Gamma}}(0) \end{pmatrix}^{-1}$$

Yule-Walker Estimation (YWE)

Furthermore,

$$\widehat{\mathbf{c}} = \bar{\mathbf{X}}_{3,T} - \widehat{\mathbf{A}}_1 \bar{\mathbf{X}}_{2,T-1} - \widehat{\mathbf{A}}_2 \bar{\mathbf{X}}_{1,T-2}$$

where $\bar{\mathbf{X}}_{i,j} = \sum_{i \leq t \leq j} \mathbf{X}_t / (j - i + 1)$.

In practice, most statistical packages simply use $\bar{\mathbf{X}}_{i,j} = \bar{\mathbf{X}} = \sum_{t=1}^{T} \mathbf{X}_t / T$.

By doing so, the YWE and LSE then give exactly the same estimator.

Maximum Likelihood Estimation (MLE)

Assuming $\epsilon_t \sim \text{MVN}(0, \mathbf{\Sigma}_\epsilon)$ (MVN means "multivariate normal"),

$$\mathbf{X}_t | \mathbf{X}_{t-1}, \mathbf{X}_{t-2} \sim \text{MVN}(\mathbf{c} + \mathbf{A}_1 \mathbf{X}_{t-1} + \mathbf{A}_2 \mathbf{X}_{t-2}, \mathbf{\Sigma}_\epsilon)$$

By the chain rule

$$\begin{aligned} f(\mathbf{X}_1, ..., \mathbf{X}_T) &= f(\mathbf{X}_T | \mathbf{X}_1, ..., \mathbf{X}_{T-1}) f(\mathbf{X}_1, ..., \mathbf{X}_{T-1}) \\ &= \cdots = f(\mathbf{X}_1, \mathbf{X}_2) \prod_{t=3}^{T} f(\mathbf{X}_t | \mathbf{X}_{t-1}, \mathbf{X}_{t-2}) \end{aligned}$$

Denote

$$\mathbf{m}_t = \mathbf{X}_t - \mathbf{c} - \mathbf{A}_1 \mathbf{X}_{t-1} - \mathbf{A}_2 \mathbf{X}_{t-2}$$

Maximum Likelihood Estimation (MLE)

Focusing on the conditional density gives the likelihood function

$$
\begin{aligned}
L(\theta, \boldsymbol{\Sigma}_\epsilon) &= \left(\frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_\epsilon|^{1/2}} \right)^{T-2} \exp\left(-\frac{1}{2} \sum_{t=3}^{T} \mathbf{m}_t' \boldsymbol{\Sigma}_\epsilon^{-1} \mathbf{m}_t \right) \\
&= \left\{ \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_\epsilon|^{1/2}} \right\}^{T-2} \exp\left\{ -\frac{1}{2} \mathbf{tr} \left(\boldsymbol{\Sigma}_\epsilon^{-1} \sum_{t=3}^{T} \mathbf{m}_t \mathbf{m}_t' \right) \right\}
\end{aligned}
$$

The maximum likelihood estimator maximizes the log-likelihood function

$$
\begin{aligned}
(\widehat{\theta}, \widehat{\boldsymbol{\Sigma}}_\epsilon) &= \arg\max_{\theta, \boldsymbol{\Sigma}_\epsilon} \log L(\theta, \boldsymbol{\Sigma}_\epsilon) \\
&= \arg\max_{\theta, \boldsymbol{\Sigma}_\epsilon} \left\{ -\frac{T-2}{2} \log |\boldsymbol{\Sigma}_\epsilon| - \frac{1}{2} \mathbf{tr} \left(\boldsymbol{\Sigma}_\epsilon^{-1} \sum_{t=3}^{T} \mathbf{m}_t \mathbf{m}_t' \right) \right\}
\end{aligned}
$$

Comparing LS, YW and ML Estimators

Let $\widehat{\theta}^{LS}$, $\widehat{\theta}^{YW}$ and $\widehat{\theta}^{ML}$ denote the LS, YW and ML estimators for the VAR model, respectively. We have known that $\widehat{\theta}^{LS} = \widehat{\theta}^{YW}$ if choosing $\bar{\mathbf{X}}_{i,j} = \bar{\mathbf{X}}$ for YWE.

Let $\mathbf{m}_t(\widehat{\theta}^{LS}) \equiv \mathbf{X}_t - \widehat{\mathbf{c}}^{LS} - \widehat{\mathbf{A}}_1^{LS}\mathbf{X}_{t-1} - \widehat{\mathbf{A}}_2^{LS}\mathbf{X}_{t-2}$. It can be shown (with some linear algebra, e.g., residual making matrix, etc.) that

$$\mathbf{Z}_T \equiv \sum_{t=3}^{T} \mathbf{m}_t \mathbf{m}_t' - \sum_{t=3}^{T} \mathbf{m}_t(\widehat{\theta}^{LS})\mathbf{m}_t(\widehat{\theta}^{LS})' \geq \mathbf{0}_{d\times d}$$

Then for all positive-definite $\mathbf{\Sigma}_\epsilon$ (so is $\mathbf{\Sigma}_\epsilon^{-1}$), we have

$$\begin{aligned}
\operatorname{tr}\left(\mathbf{\Sigma}_\epsilon^{-1}\mathbf{Z}_T\right) &= \operatorname{tr}\left(\mathbf{\Sigma}_\epsilon^{-1/2}\mathbf{\Sigma}_\epsilon^{-1/2}\mathbf{Z}_T^{1/2}\mathbf{Z}_T^{1/2}\right) = \operatorname{tr}\left(\mathbf{\Sigma}_\epsilon^{-1/2}\mathbf{Z}_T^{1/2}\mathbf{Z}_T^{1/2}\mathbf{\Sigma}_\epsilon^{-1/2}\right) \\
&= \operatorname{tr}\left(\mathbf{\Sigma}_\epsilon^{-1/2}\mathbf{Z}_T^{1/2}\mathbf{Z}_T^{1/2}\mathbf{\Sigma}_\epsilon^{-1/2}\right) = \operatorname{tr}\left(\mathbf{\Omega}\mathbf{\Omega}'\right) \geq 0
\end{aligned}$$

Comparing LS, YW and ML Estimators

This means for any given positive-definite $\bar{\mathbf{\Sigma}}_\epsilon$, $\widehat{\theta}^{LS} = \arg\max_\theta \log L(\theta, \bar{\mathbf{\Sigma}}_\epsilon)$.

Anderson (2003) shows that $\widehat{\mathbf{\Sigma}}_\epsilon^{ML} = \widehat{\mathbf{\Sigma}}_\epsilon^{LS}$ if $\epsilon_t \sim \mathrm{MVN}(0, \mathbf{\Sigma}_\epsilon)$ is assumed.

As a result, if choosing $\bar{\mathbf{X}}_{i,j} = \bar{\mathbf{X}}$ for YWE, then

$$\widehat{\theta}^{LS} = \widehat{\theta}^{YW} = \widehat{\theta}^{ML}, \widehat{\mathbf{\Sigma}}_\epsilon^{LS} = \widehat{\mathbf{\Sigma}}_\epsilon^{YW} = \widehat{\mathbf{\Sigma}}_\epsilon^{ML}$$

Furthermore,

$$\max_{\theta, \mathbf{\Sigma}_\epsilon} \log L(\theta, \mathbf{\Sigma}_\epsilon) = \log L(\widehat{\theta}^{ML}, \widehat{\mathbf{\Sigma}}_\epsilon^{ML}) = -\frac{T-2}{2} \log |\widehat{\mathbf{\Sigma}}_\epsilon^{ML}| - \frac{1}{2} d$$

Model Selection using Information Criteria

In practice, we can use the following information criterion to select the order p of a VAR(p) model:

$$\mathrm{AIC}(p) = \log(|\widehat{\boldsymbol{\Sigma}}_\epsilon(p)|) + 2d^2p/T$$

$$\mathrm{BIC}(p) = \log(|\widehat{\boldsymbol{\Sigma}}_\epsilon(p)|) + d^2p\log(T)/T$$

$$\mathrm{HQIC}(p) = \log(|\widehat{\boldsymbol{\Sigma}}_\epsilon(p)|) + 2d^2p\log(\log T)/T$$

We may choose p such that one of AIC, BIC and/or HQIC is minimized, but the selection should also make practical sense.

In general,

$$\widehat{p}(\text{AIC}) \geq \widehat{p}(\text{HQIC}) \geq \widehat{p}(\text{BIC})$$

where $\widehat{p}(\cdot)$ denotes the p selected by according information criterion.

Model Diagnostics using Portmanteau Tests

As in univariate cases, we can do model diagnostic via examining if the residuals behave like (vector) white noise.

Let $\Upsilon(k)$ denote the cross-covariance matrix of the residuals at lag k. The null hypothesis

$$H_0 : \Upsilon(1) = \cdots = \Upsilon(k) = \mathbf{0}_{d\times d}$$

can be tested by a general purpose (portmanteau) Q-statistic. First, we compute residuals

$$\widehat{\epsilon}_t = \mathbf{X}_t - \widehat{\mathbf{c}} - \widehat{\mathbf{A}}_1\mathbf{X}_{t-1} - \cdots - \widehat{\mathbf{A}}_p\mathbf{X}_{t-p}$$

Let $\widehat{\Upsilon}(k)$ denote the sample analogue of $\Upsilon(k)$.

Model Diagnostics using Portmanteau Tests

Then the portmanteau test statistic is defined as

$$Q_m = \frac{1}{T^2} \sum_{k=1}^{m} \sum \frac{1}{T-k} \mathrm{tr} \left\{ \widehat{\Upsilon}(k)' \widehat{\Upsilon}(0)^{-1} \widehat{\Upsilon}(k) \widehat{\Upsilon}(0)^{-1} \right\} \rightsquigarrow \chi^2_{d^2(m-p)}$$

where $m > p$ is an integer. In practice, this test is sensitive to the choice of m, and so it is good to do the test with different values of m.

Portmanteau tests may not be powerful when d is large.

Illustration with Real Data

Wind financial terminal (WFT)

- Wind data feed services (WDFS).
- Wind economic database (WEDB).
- WindR package for R.
- See handouts.

Example: S&P 500/400/600 Data

- See R markdown.

Example: S&P500, SH index and SZ index Data

- See R markdown.