Financial Econometrics

Lecture 9

Fu Ouyang

May 9, 2018

Outline

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

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

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 $\mathbf{\Sigma}_\epsilon$ can be defined as

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

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

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 $\boldsymbol{\Sigma}_\epsilon$ (so is $\boldsymbol{\Sigma}_\epsilon^{-1}$), we have

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

Comparing LS, YW and ML Estimators

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

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

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

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

Furthermore,

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

Model Selection using Information Criteria

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

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

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

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

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,

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

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 code.

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

- See R code.