Review Session 2: VAR, MLE

Jiahui Shui
jishui@ucsd.edu

Rady School of Management, UCSD

November 15, 2024

Stationary: Recall

Let $\{x_t\}_{t=0}^{\infty}$ be a time series.

Definition (Covariance Stationary)

If $\{x_t\}$ satisfies

- $\mathbb{E}[x_t] = \mu$, does not depend on t.
- $\forall j$, $\mathrm{Cov}(x_t, x_{t-j}) = \gamma(j), \forall t > j$, does not depend on t.

Then we call $\{x_t\}$ is covariance stationary, or weakly stationary.

Definition (Ergodic)

A covariance stationary process is said to be ergodic for the mean if

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t = \mu \tag{1}$$

Stationarity For AR(p) Process

- Consider the following AR(p) process:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t \qquad (2)$$

- If the root of the following polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0 \qquad (3)$$

lie outside the unit circle, i.e. for each (complex) root z_i, it satisfies $|z_i| > 1$, then x_t is weak stationary.

VAR(1)

Consider the following regression

$$Z_t = \alpha + \Phi Z_{t-1} + \varepsilon_t \tag{4}$$

where Z_t is a vector time series. A 2-dimensional case is

$$\begin{aligned} z_{1t} &= \alpha_1 + \Phi_{11} z_{1,t-1} + \Phi_{12} z_{2,t-1} + \varepsilon_{1t} \\ z_{2t} &= \alpha_2 + \Phi_{21} z_{1,t-1} + \Phi_{22} z_{2,t-1} + \varepsilon_{2t} \end{aligned} \tag{5}$$

Stationary condition:

$$\max \quad |\lambda(\Phi)| < 1 \tag{6}$$

Why?

VAR(p)

- How about a more general form

$$Z_t = \alpha + \Phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + \varepsilon_t \tag{7}$$

- Stationary condition: the roots of the following equation:

$$\det(I_n - \Phi_1 z - \cdots - \Phi_p z^p) = 0 \tag{8}$$

  lie outside the unit circle, i.e. $\|z\| > 1$.
- When $p = 1$, the above condition becomes

$$\det(I_n - \Phi_1 z) = |z| \det(1/z \cdot I_n - \Phi_1) = 0 \Rightarrow \det(1/z \cdot I_n - \Phi_1) = 0 \tag{9}$$

- $1/z$ is eigenvalue of Φ_1
- $\|z\| > 1 \Leftrightarrow \|1/z\| < 1 \Leftrightarrow \max \quad |\lambda(\Phi_1)| < 1$

Maximum Likelihood Estimation

- Suppose that $f(y|\theta)$ is the conditional density function of random variable Y. That is, the distribution of Y depends on parameter θ, where $\theta \in \Theta$.
- Then we observe a series $\{y_t\}_{t=1}^{T}$. We can consider the joint density function of this 'observation' as a function of θ.
- Likelihood function:

$$L(\theta) = \prod_{t=1}^{T} f(y_t|\theta) \tag{10}$$

- Log-likelihood function:

$$\Lambda(\theta) = \log L(\theta) = \sum_{t=1}^{T} \log f(y_t|\theta) \tag{11}$$

MLE

- The MLE estimator is the value $\hat{\theta}$ that maximizes $\Lambda(\theta)$, i.e.

$$\hat{\theta}^{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \Lambda(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta) \tag{12}$$

- Suppose that there exists a unique interior solution.
- The FOC is

$$\frac{\partial \Lambda(\theta)}{\partial \theta} = 0 \tag{13}$$

- Here θ might be a vector.
- Hessian matrix: $H(\theta) = \frac{\partial^2 \Lambda(\theta)}{\partial \theta \partial \theta'}$

A Simple Example: MLE for linear Regression

Consider the following regression:

$$y_t = \beta x_t + \varepsilon_t \tag{14}$$

Suppose that $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d. Find the MLE of β and σ^2

Solution: Note that $y_t|\beta \sim N(\beta x_t, \sigma^2)$, then

$$f(y_t|\beta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t - \beta x_t)^2}{2\sigma^2}}$$

Then the likelihood function is

$$L(\beta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_t - \beta x_t)^2}{2\sigma^2}}$$

The log-likelihood function is

$$\Lambda(\beta, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \beta x_t)^2$$

A Simple Example: MLE for linear Regression

Take derivative with respect to β, we get FOC

$$\frac{\partial \Lambda}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=1}^{T} (y_t - \beta x_t) x_t = 0$$

Hence

$$\sum_{t=1}^{T} y_t x_t = \beta \sum_{t=1}^{T} x_t^2 \Rightarrow \hat{\beta}^{MLE} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2}$$

Under i.i.d and normal assumptions, MLE estimator and OLS estimator for linear regressions are the same.

For σ^2, we have the FOC

$$\frac{\partial \Lambda}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} (y_t - \beta x_t)^2 = 0$$

Then

$$\hat{\sigma}^2_{MLE} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \beta x_t)^2 = \frac{e'e}{T}$$

Recall the unbiased estimator for σ^2 is $s^2 = \frac{e'e}{T-1}$.

Exact MLE and Conditional MLE

- In the lecture, we considered MLE for AR(1) process:

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{15}$$

Suppose that $\varepsilon_t \sim N(0, \sigma^2)$ and i.i.d.

- Likelihood function

$$L(\phi) = f_{y_1}(y_1|\phi) \prod_{t=2}^{T} f_{y_t|y_{t-1}}(y_t|y_{t-1}; \phi) \tag{16}$$

- For $t \geq 2$, we consider the **conditional** distribution of y_t given y_{t-1}. But for y_1, we can only consider the unconditional distribution. $y_1 \sim N(0, \sigma^2/(1-\phi^2))$
- The log likelihood function is

$$\begin{aligned}\Lambda(\phi) = &-\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) \\ &- \frac{y_1^2}{2\sigma^2/(1-\phi^2)} - \sum_{t=2}^{T} \frac{(y_t - \phi y_{t-1})^2}{2\sigma^2}\end{aligned} \tag{17}$$

Exact MLE and Conditional MLE

- Exact MLE: Optimization with respect to $\Lambda(\phi)$
- Conditional MLE: Take y_1 as **known**. Then optimize the following function:

$$\tilde{\Lambda}(\phi) = -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - -\sum_{t=2}^{T}\frac{(y_t - \phi y_{t-1})^2}{2\sigma^2} \quad (18)$$

- When T is large enough, the contribution of y_1 to likelihood can be ignored.
- If $|\phi| < 1$, then exact MLE and conditional MLE has same asymptotic distribution.
- If $|\phi| > 1$, then conditional MLE is still **consistent** but exact MLE is not.

Cramér-Rao lower bound (Optional)

- Score function: $s(\theta) := \frac{\partial \Lambda(\theta)}{\partial \theta}$
- Fisher Information: $I(\theta) = \mathbb{E}[\frac{\partial \Lambda(\theta)}{\partial \theta}]^2$
- Under certain condition, $I(\theta) = -\mathbb{E}[H(\theta)]$

Theorem (Cramér-Rao Inequality)

Let $X_1, \cdots, X_n$ be a (i.i.d) sample with pdf $f(x|\theta)$ and Let $W = W(X_1, \cdots, X_n)$ be any estimator for $g(\theta)$ satisfying

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[W(X)] = \int_{\mathcal{X}} \frac{\partial}{\partial \theta}[W(x_1, \cdots, x_n) f(x_1, \cdots, x_n|\theta)] \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{19}$$

Then

$$\mathrm{Var}(W) \geq \frac{(\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}[W(X)])^2}{I(\theta)} \tag{20}$$

In particular, if $W(\theta) = \theta$, then $\mathrm{Var}(\hat{\theta}) \geq [I(\theta)]^{-1}$

Exercise (Optional)

Suppose that $y_t = \beta x_t + \varepsilon_t$, with $\varepsilon_t \sim N(0, \sigma^2)$ i.i.d. Now find the Cramér-Rao lower bound for $\theta = (\beta, \sigma^2)'$.

Recall from previous question:

$$\Lambda(\beta, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \beta x_t)^2$$

We have

$$\frac{\partial \Lambda}{\partial \beta} = \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - \beta x_t)x_t, \quad \frac{\partial \Lambda}{\partial(\sigma^2)} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t - \beta x_t)^2$$

and

$$\frac{\partial^2 \Lambda}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}x_t^2, \quad \frac{\partial^2 \Lambda}{\partial(\sigma^2)^2} = \frac{T}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{t=1}^{T}(y_t - \beta x_t)^2$$

Also

$$\frac{\partial^2 \Lambda}{\partial \beta \partial(\sigma^2)} = \frac{\partial^2 \Lambda}{\partial \sigma^2 \partial \beta} = -\frac{1}{(\sigma^2)^2}\sum_{t=1}^{T}(y_t - \beta x_t)x_t$$

Exercise (Optional)

Note that

$$\mathbb{E}\left[\frac{\partial^2 \Lambda}{\partial \beta \partial(\sigma^2)}\right] = -\frac{1}{(\sigma^2)^2}\sum_{t=1}^{T} x_t(\mathbb{E}[y_t - \beta x_t]) = 0$$

And

$$\mathbb{E}\left[\frac{\partial^2 \Lambda}{\partial \beta^2}\right] = \frac{1}{\sigma^2}\sum_{t=1}^{T} x_t^2, \quad \mathbb{E}\left[\frac{\partial^2 \Lambda}{\partial (\sigma^2)^2}\right] = -\frac{T}{2(\sigma^2)^2}$$

Finally,

$$I(\theta) = -\mathbb{E}[H(\theta)] = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T} x_t^2 & 0 \\ 0 & \frac{T}{2(\sigma^2)^2} \end{pmatrix}$$

The Cramér-Rao lower bound in this case is

$$[I(\theta)]^{-1} = \begin{pmatrix} \sigma^2(\sum_{t=1}^{T} x_t^2)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{pmatrix}$$

Remark: Note that $\mathrm{Var}(\hat{\beta}^{OLS}|X) = \sigma^2(\sum_{t=1}^{T} x_t^2)^{-1}$, OLS estimator for β achieves Cramér-Rao lower bounds. But unbiased estimator for σ^2, i.e. s^2 has variance $2\sigma^4/(T-1)$.

Properties of MLE

- Consistency: $\text{plim}_{T\to\infty}\hat{\theta}^{MLE} = \theta$
- Asymptotic variance: $\mathrm{Avar}(\hat{\theta}^{MLE}) = T[I(\theta)]^{-1}$.
- Asymptotic normal: $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, T[I(\theta)]^{-1})$.
  Intuitively, it can be thought as

$$\hat{\theta}^{MLE} \sim N(\theta, [I(\theta)]^{-1}) \tag{21}$$