VAR Estimation

Giovanni Ricco

20th October 2023

Conditional Likelihood

Let us consider the VAR(p)

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \epsilon_t \tag{1}$$

with $\epsilon_t \sim i.i.dN(0, \Omega)$. Suppose we have a sample of $T + p$ observations for such variables

Conditioning on the first p observations we can form the conditional likelihood

$$f(Y_T, Y_{T-1}, ..., Y_1 | Y_0, Y_{-1}, ..., Y_{-p+1}, \theta) \tag{2}$$

where θ is a vector containing all the parameters of the model. We refer to (2) as 'conditional likelihood function'

Conditional Likelihood

The joint density of observations 1 through t conditioned on $Y_0, \ldots, Y_{-p+1}$ satisfies

$$\begin{aligned} f(Y_t, Y_{t-1}, \ldots, Y_1 | Y_0, Y_{-1}, \ldots, Y_{-p+1}, \theta) \\ = f(Y_{t-1}, \ldots, Y_1 | Y_0, Y_{-1}, \ldots, Y_{-p+1}, \theta) \\ \times f(Y_t | Y_{t-1}, \ldots, Y_1, Y_0, Y_{-1}, \ldots, Y_{-p+1}, \theta) \end{aligned}$$

Applying the formula recursively, the likelihood for the full sample is the product of the individual conditional densities

$$f(Y_t, Y_{t-1}, \ldots, Y_1 | Y_0, Y_{-1}, \ldots, Y_{-p+1}, \theta) = \prod_{t=1}^{T} f(Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_{-p+1}, \theta)$$

Conditional Likelihood

At each t, conditional on the values of Y through date $t-1$

$$Y_t | Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p}, \Omega)$$

Recall

$$X_t = \begin{pmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{pmatrix}$$

is an $(np+1) \times 1$ vector and let $\Pi' = [c, A_1, A_2, ..., A_p]$ be an $(n \times np + 1)$ matrix of coefficients

Using this notation we have that

$$Y_t | Y_{t-1}, Y_{t-2}, ..., Y_{-p+1} \sim N(\Pi' X_t, \Omega)$$

Conditional Likelihood

Thus the conditional density of the t-th observation is

$$f(Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_{-p+1}, \theta) = (2\pi)^{-n/2} \left|\Omega^{-1}\right|^{1/2} \exp\left[-\frac{1}{2}(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right] \quad (3)$$

The sample log-likelihood is found by substituting (3) into the likelihood for the full sample and taking logs

$$\begin{aligned} \mathcal{L}(\theta) &= \sum_{t=1}^{T} \log f(Y_t | Y_{t-1}, Y_{t-2}, \ldots, Y_{-p+1}, \theta) \\ &= -\frac{Tn}{2} \log(2\pi) + (T/2) \log |\Omega^{-1}| \\ &\quad -\frac{1}{2} \sum_{t=1}^{T} \left[(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right] \end{aligned}$$

Maximum Likelihood Estimate (MLE) of Π

The MLE estimate of Π are given by

$$\hat{\Pi}'_{MLE} = \left[\sum_{t=1}^{T} Y_t X_t'\right] \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}$$

$\hat{\Pi}'_{MLE}$ is $n \times (np+1)$. The j-th row of $\hat{\Pi}'$ is

$$\hat{\pi}_j' = \left[\sum_{t=1}^{T} Y_{jt} X_t'\right] \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1}$$

which is the estimated coefficient vector from an OLS regression of Y_{jt} on X_t

$\hat{\Pi}'_{MLE} = \hat{\Pi}'_{OLS}$

① Rewrite the last term in the log-likelihood as

$$\sum_{t=1}^{T} \left[(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t)\right]$$

$$\begin{aligned}
&= \sum_{t=1}^{T} [(Y_t - \hat{\Pi}' X_t + \hat{\Pi}' X_t - \Pi' X_t)' \Omega^{-1} \\
&\quad \times (Y_t - \hat{\Pi}' X_t + \hat{\Pi}' X_t - \Pi' X_t)] \\
&= \sum_{t=1}^{T} \left[(\hat{\epsilon}_t + (\hat{\Pi}' - \Pi') X_t)' \Omega^{-1} (\hat{\epsilon}_t + (\hat{\Pi}' - \Pi') X_t)\right] \\
&= \sum_{t=1}^{T} \hat{\epsilon}'_t \Omega^{-1} \hat{\epsilon}_t + 2 \sum_{t=1}^{T} \hat{\epsilon}'_t \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t + \\
&\quad + \sum_{t=1}^{T} X'_t (\hat{\Pi}' - \Pi') \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t
\end{aligned}$$

$\hat{\Pi}'_{MLE} = \hat{\Pi}'_{OLS}$

② The term $2\sum_{t=1}^{T} \hat{\epsilon}'_t \Omega^{-1}(\hat{\Pi}' - \Pi')' X_t$ is a scalar so that

$$\begin{aligned} \sum_{t=1}^{T} \hat{\epsilon}'_t \Omega^{-1}(\hat{\Pi}' - \Pi')' X_t &= tr\left[\sum_{t=1}^{T} \hat{\epsilon}'_t \Omega^{-1}(\hat{\Pi}' - \Pi')' X_t\right] \\ &= tr\left[\sum_{t=1}^{T} \Omega^{-1}(\hat{\Pi}' - \Pi')' X_t \hat{\epsilon}'_t\right] \\ &= tr\left[\Omega^{-1}(\hat{\Pi}' - \Pi')' \sum_{t=1}^{T} X_t \hat{\epsilon}'_t\right] \end{aligned}$$

But $\sum_{t=1}^{T} X_t \hat{\epsilon}'_t = 0$ by construction since regressors are orthogonal to the residuals, hence the term is zero

$\hat{\Pi}'_{MLE} = \hat{\Pi}'_{OLS}$

3. We have

$$\sum_{t=1}^{T} \left[(Y_t - \Pi' X_t)' \Omega^{-1} (Y_t - \Pi' X_t) \right] =$$

$$\sum_{t=1}^{T} \hat{\epsilon}_t' \Omega^{-1} \hat{\epsilon}_t + \sum_{t=1}^{T} X_t' (\hat{\Pi}' - \Pi') \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t$$

4. Given that Ω is positive definite, so is Ω^{-1}, thus the smallest values of

$$\sum_{t=1}^{T} X_t' (\hat{\Pi}' - \Pi') \Omega^{-1} (\hat{\Pi}' - \Pi')' X_t$$

is achieved by setting $\Pi = \hat{\Pi}$, i.e. the log-likelihood is maximised when $\Pi = \hat{\Pi}$

$\hat{\Pi}'_{MLE} = \hat{\Pi}'_{OLS}$

Recall the SUR representation

$$\mathbf{Y} = \mathbf{XA} + \mathbf{u}$$

where $\mathbf{X} = [X_1, ..., X_T]'$, $X_t = [Y'_{t-1}, Y'_{t-2} ..., Y'_{t-p}]'$ $\mathbf{Y} = [Y_1, ..., Y_T]'$, $\mathbf{u} = [\epsilon_1, ..., \epsilon_T]'$ and $\mathbf{A} = [A_1, ..., A_p]'$.The MLE estimator is given by

$$\hat{\mathbf{A}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

(notice that $\hat{\mathbf{A}} = \hat{\Pi}'_{MLE}$, different notation same estimator)

MLE of Ω

Let X be an $n \times 1$ vector and let A be a non-symmetric and unrestricted matrix. Consider the quadratic form $X'AX$.

- **Result 1**:
$$\frac{\partial X'AX}{\partial A} = XX'$$
- **Result 2**:
$$\frac{\partial \log |A|}{\partial A} = (A')^{-1}$$

MLE of Ω

We now find the MLE of Ω. When evaluated at $\hat{\Pi}$ the log likelihood is

$$\mathcal{L}(\theta) = -(Tn/2)\log(2\pi) + (T/2)\log|\Omega^{-1}| - (1/2)\sum_{t=1}^{T} \hat{\epsilon}_t'\Omega^{-1}\hat{\epsilon}_t$$

Taking derivatives and using results for matrix derivatives we have:

$$\begin{aligned} \frac{\partial \mathcal{L}(\Omega, \hat{\Pi})}{\partial \Omega^{-1}} &= (T/2)\frac{\partial \log|\Omega^{-1}|}{\partial \Omega^{-1}} - (1/2)\frac{\sum_{t=1}^{T} \partial \hat{\epsilon}_t'\Omega^{-1}\hat{\epsilon}_t}{\partial \Omega^{-1}} \\ &= (T/2)\Omega' - (1/2)\sum_{t=1}^{T} \hat{\epsilon}_t\hat{\epsilon}_t' \end{aligned}$$

MLE of Ω

The likelihood is maximised when the derivative is set to zero, or when

$$\Omega' = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t'$$

Remark: The ML estimator for Ω coincide with the average of squared residuals from OLS regressions

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t\hat{\epsilon}_t'$$

Asymptotic Distribution of $\hat{\Pi}$

Proposition. Let

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \varepsilon_t$$

where

1. ε_t is $i.i.d.$ with mean zero and variance Ω
2. $E(\varepsilon_{it}\varepsilon_{jt}\varepsilon_{lt}\varepsilon_{mt}) < \infty$ for all i, j, l, m
3. the roots of
$$|I - A_1 z + A_2 z^2 + ... + A_p z^p| = 0$$
lie outside the unit circle.

Let $k = np + 1$ and let X_t be the $1 \times k$ vector

$$X_t' = [1, Y_{t-1}', Y_{t-2}', ..., Y_{t-p}']$$

Asymptotic Distribution of $\hat{\Pi}$

Let $\hat{\pi}_T = vec(\hat{\Pi}_T)$ denote the $nk \times 1$ vector of coefficients resulting from the OLS regressions of each of the element of Y_t on X_t for a sample of size T

$$\hat{\pi}_T = \begin{pmatrix} \hat{\pi}_{1T} \\ \hat{\pi}_{2T} \\ \vdots \\ \hat{\pi}_{nT} \end{pmatrix}$$

where

$$\hat{\pi}_{iT} = \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1} \left[\sum_{t=1}^{T} X_t Y_{it}\right]$$

and let π denote the vector of corresponding population coefficients

Asymptotic Distribution of $\hat{\Pi}$

Finally let

$$\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

where $\hat{\varepsilon}_t = [\hat{\varepsilon}_{1t} \; \hat{\varepsilon}_{2t} \; \ldots \; \hat{\varepsilon}_{nt}]$, and $\hat{\varepsilon}_{it} = Y_{it} - X_t' \hat{\pi}_{iT}$.

Then:

1. $\frac{1}{T} \sum_{t=1} X_t X_t' \xrightarrow{p} Q$ where $Q = E(X_t X_t')$
2. $\hat{\pi}_T \xrightarrow{p} \pi$
3. $\hat{\Omega} \xrightarrow{p} \Omega$
4. $\sqrt{T}(\hat{\pi}_T - \pi) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega \otimes Q^{-1})$

Asymptotic Distribution of $\hat{\Pi}$

Remark: Result (4) implies that

$$\sqrt{T}(\hat{\pi}_{iT} - \pi_i) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_i^2 Q^{-1})$$

where σ_i^2 is the variance of the error term of the ith equation. σ_i^2 is consistently estimated by $\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{it}^2$ and that Q^{-1} is consistently estimated by

$$\left(\frac{1}{T}\sum_{t=1} X_t X_t'\right)^{-1}$$

Therefore we can treat $\hat{\pi}_i$ approximately as

$$\hat{\pi}_i \approx \mathcal{N}\left(\pi, \hat{\sigma}_i^2 \left[\sum_{t=1} X_t X_t'\right]^{-1}\right)$$

Asymptotic Distribution of $\hat{\Pi}$

Remark: MLEs are consistent even if the true innovations are non-Gaussian

Addendum: Lag Order Selection

As in the univariate case, care must be taken to account for all systematic dynamics in multivariate models. In VAR models, this is usually done by choosing a sufficient number of lags to ensure that the residuals in each of the equations are white noise

AIC: Akaike information criterion – p that minimises

$$AIC(p) = \ln|\hat{\Omega}| + 2\frac{n^2 p}{T}$$

BIC: Bayesian information criterion – p that minimises

$$BIC(p) = \ln|\hat{\Omega}| + \frac{n^2 p}{T}\ln T$$

HQ: Hannan-Quinn information criterion – p that minimises

$$HQ(p) = \ln|\hat{\Omega}| + 2\frac{n^2 p}{T}\ln\ln T$$

Addendum: Lag Order Selection

Remarks:

- $\hat{p}$ obtained using BIC and HQ are consistent
- $\hat{p}$ obtained using AIC it is not consistent
- AIC overestimate the true order with positive probability and underestimate the true order with zero probability
- Suppose a VAR(p) is fitted to $Y_1, \ldots, Y_T$ (Y_t not necessarily stationary). In small sample the following relations hold:

$$\hat{p}_{BIC} \leq \hat{p}_{AIC} \text{ if } T \geq 8$$
$$\hat{p}_{BIC} \leq \hat{p}_{HQ} \text{ for all } T$$
$$\hat{p}_{HQ} \leq \hat{p}_{AIC} \text{ if } T \geq 16$$