



Università di Pavia

Vector AutoRegression Model

Eduardo Rossi



VAR

Vector autoregressions (VARs) were introduced into empirical economics by C.Sims (1980), who demonstrated that VARs provide a flexible and tractable framework for analyzing economic time series.

Identification issue: since these models don't dichotomize variables into "endogenous" and "exogenous", the exclusion restrictions used to identify traditional simultaneous equations models make little sense.

A Vector Autoregression model (VAR) of order p is written as:

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t$$

$$\mathbf{y}_t : (N \times 1) \quad \Phi_i : (N \times N) \quad \forall i, \quad \epsilon_t : (N \times 1)$$

$$E(\epsilon_t) = 0 \quad E(\epsilon_t \epsilon_\tau') = \begin{cases} \Omega & t = \tau \\ \mathbf{0} & t \neq \tau \end{cases}$$

Ω positive definite matrix.



VAR(p)

A VAR is a vector generalization of a scalar autoregression. The VAR is a system in which each variable is regressed on a constant and p of its own lags as well as on p lags of each of the other variables in the VAR.

$$\begin{aligned} [\mathbf{I}_N - \Phi_1 L - \dots - \Phi_p L^p] \mathbf{y}_t &= \mathbf{c} + \epsilon_t \\ \Phi(L) \mathbf{y}_t &= \mathbf{c} + \epsilon_t \end{aligned}$$

with

$$\Phi(L) = [\mathbf{I}_N - \Phi_1 L - \dots - \Phi_p L^p]$$

$\Phi(L)$ ($N \times N$) matrix polynomial in L .



VAR(p) - STATIONARITY

The element (i, j) in $\Phi(L)$ is a scalar polynomial in L

$$\delta_{ij} = \phi_{ij}^{(1)}L + \phi_{ij}^{(2)}L^2 + \dots + \phi_{ij}^{(p)}L^p$$
$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Stationarity: A vector process is said *covariance stationary* if its first and second moments, $E[\mathbf{y}_t]$ and $E[\mathbf{y}_t\mathbf{y}'_{t-j}]$ respectively, are independent of the date t .



VAR(1)

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \epsilon_t$$

First equation:

$$y_{1t} = c_1 + \phi_{11}^{(1)} y_{1t-1} + \phi_{12}^{(1)} y_{2t-1} + \dots + \phi_{1N}^{(1)} y_{Nt-1} + \epsilon_{1t}$$

$$\begin{aligned} \mathbf{y}_t &= \mathbf{c} + \Phi_1 [\mathbf{c} + \Phi_1 \mathbf{y}_{t-2} + \epsilon_{t-1}] + \epsilon_t \\ &= \mathbf{c} + \Phi_1 \mathbf{c} + \Phi_1^2 \mathbf{y}_{t-2} + \epsilon_t + \Phi_1 \epsilon_{t-1} \end{aligned}$$

$$\mathbf{y}_t = \dots$$

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{c} + \dots + \Phi_1^{k-1} \mathbf{c} + \Phi_1^k \mathbf{y}_{t-k} + \epsilon_t + \Phi_1 \epsilon_{t-1} + \dots + \Phi_1^{k-1} \epsilon_{t-k+1}$$

$$E[\mathbf{y}_t] = \sum_{j=0}^{k-1} \Phi_1^j \mathbf{c} + \Phi_1^k E[\mathbf{y}_{t-k}]$$

The value of this sum depends on the behavior of Φ_1^j as j increases.



STABILITY OF VAR(1)

Let $\lambda_1, \lambda_2, \dots, \lambda_N$ be the eigenvalues of Φ_1 , the solutions to the characteristic equation

$$|\Phi_1 - \lambda \mathbf{I}_N| = 0$$

then, if the eigenvalues are all distinct

$$\Phi_1 = \mathbf{Q} \mathbf{M} \mathbf{Q}^{-1}$$

$$\Phi_1^k = \mathbf{Q} \mathbf{M}^k \mathbf{Q}^{-1}$$

$$\mathbf{M}^k = \text{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_N^k)$$

If $|\lambda_i| < 1$, $i = 1, \dots, N$

$$\Phi^k \rightarrow 0, \quad k \rightarrow \infty$$

If $|\lambda_i| \geq 1$, $\forall i$ then one or more elements of \mathbf{M}^k are not vanishing, and may be tending to ∞ .



VAR(P) COMPANION FORM

VAR(p):

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t$$

as a VAR(1) (Companion Form):

$$\boldsymbol{\xi}_t = \mathbf{F} \boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{y}_t \\ \vdots \\ \mathbf{y}_{t-p+1} \end{bmatrix} \quad (Np \times 1) \quad \boldsymbol{\xi}_{t-1} = \begin{bmatrix} \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \quad (Np \times 1)$$



VAR(P) COMPANION FORM

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_N & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I}_N & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{I}_N & 0 \end{bmatrix} \quad (Np \times Np) \quad \mathbf{v}_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (Np \times 1)$$

$$E[\mathbf{v}_t \mathbf{v}'_{\tau}] = \begin{cases} \mathbf{Q} & t = \tau \\ \mathbf{0} & t \neq \tau \end{cases} \quad \mathbf{Q} = \begin{bmatrix} \Omega & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \quad (Np \times Np)$$



STABILITY OF VAR(p)

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t \quad t = 1, \dots, T$$

If the process \mathbf{y}_t has finite variance and an autocovariance sequence that converges to zero at an exponential rate, then ξ_t must share these properties. This is ensured by having the Np eigenvalues of \mathbf{F} lie *inside the unit circle*.

The determinant defining the characteristic equation is

$$|\mathbf{F} - \lambda \mathbf{I}_{Np}| = (-1)^{Np} |\lambda^p \mathbf{I}_N - \lambda^{p-1} \Phi_1 - \dots - \Phi_p| = 0$$



STABILITY OF VAR(p)

The required condition is that the roots of the equation

$$|\lambda^p \mathbf{I}_N - \lambda^{p-1} \Phi_1 - \dots - \Phi_p| = 0$$

a polynomial of order Np must lie inside the unit circle.

Stability condition can also be expressed as the roots of

$$|\Phi(z)| = 0$$

lie outside the unit circle, where $\Phi(z)$ is a $(N \times N)$ matrix polynomial in the lag operator of order p .



STABILITY OF VAR(p)

When $p = 1$, the roots of

$$|\mathbf{I}_N - \Phi_1 z| = 0$$

outside the unit circle, i.e. $|z| > 1$, implies that the eigenvalues of Φ_1 be inside the unit circle. Note that the eigenvalues, roots of $|\Phi_1 - \lambda \mathbf{I}_N| = 0$, are the reciprocal of the roots of $|\mathbf{I}_N - \Phi_1 z| = 0$.



STABILITY OF VAR(p)

Three conditions are necessary for stationarity of the VAR(p) model:

- Absence of mean shifts;
- The vectors $\{\epsilon_t\}$ are identically distributed, $\forall t$;
- Stability condition on \mathbf{F} .

If the process is covariance stationary we can take the expectations of both sides of

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t$$

$$\boldsymbol{\mu} = \mathbf{c} + \Phi_1 \boldsymbol{\mu} + \dots + \Phi_p \boldsymbol{\mu}$$

$$\boldsymbol{\mu} = [\mathbf{I}_N - \Phi_1 \dots - \Phi_p]^{-1} \mathbf{c}$$

$$= \Phi(1)^{-1} \mathbf{c}$$



VECTOR MA(∞)

If the VAR(p) is stationary then it has a VMA(∞) representation:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\epsilon}_{t-2} + \dots \equiv \boldsymbol{\mu} + \boldsymbol{\Psi}(L) \boldsymbol{\epsilon}_t$$

\mathbf{y}_{t-j} is a linear function of $\boldsymbol{\epsilon}_{t-j}, \boldsymbol{\epsilon}_{t-j-1}, \dots$ each of which is uncorrelated with $\boldsymbol{\epsilon}_{t+1}$ for $j = 0, 1, \dots$

It follows that

- $\boldsymbol{\epsilon}_{t+1}$ is uncorrelated with \mathbf{y}_{t-j} for any $j \geq 0$.
- Linear forecast of \mathbf{y}_{t+1} based on $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$ is given by

$$\hat{\mathbf{y}}_{t+1|t} = \boldsymbol{\mu} + \boldsymbol{\Phi}_1(\mathbf{y}_t - \boldsymbol{\mu}) + \boldsymbol{\Phi}_2(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \dots + (\mathbf{y}_{t-p+1} - \boldsymbol{\mu})$$



ϵ_{t+1} can be interpreted as the fundamental innovation in \mathbf{y}_{t+1} , that is the error in forecasting \mathbf{y}_{t+1} on the basis of a linear function of a constant and $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$

A forecast of \mathbf{y}_{t+s} on the basis of $\mathbf{y}_t, \mathbf{y}_{t-1}, \dots$ will take the form

$$\hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \dots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu})$$

The moving average matrices $\boldsymbol{\Psi}_j$ can be calculated as:

$$\boldsymbol{\Psi}(L) = [\boldsymbol{\Phi}(L)]^{-1}$$

$$\boldsymbol{\Psi}(L)\boldsymbol{\Phi}(L) = \mathbf{I}_N$$



$$[\mathbf{I}_N + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \dots][\mathbf{I}_N - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 + \dots - \boldsymbol{\Phi}_p L^p] = \mathbf{I}_N$$

Setting the coefficient on L^1 equal to the zero matrix,

$$\boldsymbol{\Psi}_1 - \boldsymbol{\Phi}_1 = 0$$

on L^2

$$\boldsymbol{\Psi}_2 = \boldsymbol{\Psi}_1 \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2$$

In general for L^s

$$\boldsymbol{\Psi}_s = \boldsymbol{\Phi}_1 \boldsymbol{\Psi}_{s-1} + \dots + \boldsymbol{\Phi}_p \boldsymbol{\Psi}_{s-p} \quad s = 1, 2, \dots$$

$$\boldsymbol{\Psi}_0 = \mathbf{I}_N, \quad \boldsymbol{\Psi}_s = 0 \quad s < 0$$



The innovation in the MA(∞) representation is $\boldsymbol{\epsilon}_t$, *the fundamental innovation* for \mathbf{y} .

There are alternative MA representation based on VWN processes other than $\boldsymbol{\epsilon}_t$.

Let \mathbf{H} be a nonsingular ($N \times N$) matrix

$$\mathbf{u}_t \equiv \mathbf{H}\boldsymbol{\epsilon}_t$$

$\mathbf{u}_t \sim VWN$.

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\mu} + \mathbf{H}^{-1}\mathbf{H}\boldsymbol{\epsilon}_t + \boldsymbol{\Psi}_1\mathbf{H}^{-1}\mathbf{H}\boldsymbol{\epsilon}_{t-1} + \dots \\ &= \boldsymbol{\mu} + \mathbf{J}_0\mathbf{u}_t + \mathbf{J}_1\mathbf{u}_{t-1} + \mathbf{J}_2\mathbf{u}_{t-2} + \dots\end{aligned}$$

$$\mathbf{J}_s \equiv \boldsymbol{\Psi}_s\mathbf{H}^{-1}$$



For example \mathbf{H} can be any matrix that diagonalizes $\mathbf{\Omega}$, the var-cov of $\boldsymbol{\epsilon}_t$,

$$\mathbf{H}\mathbf{\Omega}\mathbf{H}' = \mathbf{D}$$

the elements of \mathbf{u}_t are uncorrelated with one another.

It is always possible to write a stationary VAR(p) process as a convergent infinite MA of a VWN whose elements are mutually uncorrelated.

To obtain the MA representation for the fundamental innovations, we must impose $\boldsymbol{\Psi}_0 = \mathbf{I}_N$ (while \mathbf{J}_0 is not the identity matrix).



ASSUMPTIONS IMPLICIT IN A VAR

- For a covariance stationary process, the parameters $\mathbf{c}, \Phi_1, \dots, \Phi_p$ could be defined as the coefficients of the projections of \mathbf{y}_t on $1, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_{t-p}$.
- ϵ_t is uncorrelated with $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$ by the definition of Φ_1, \dots, Φ_p .
- The parameters of a VAR can be estimated consistently with n OLS regressions.
- ϵ_t defined by this projection is uncorrelated with $\mathbf{y}_{t-p-1}, \mathbf{y}_{t-p-2}, \dots$
- The assumption of $\mathbf{y}_t \sim VAR(p)$ is basically the assumption that p lags are sufficient to summarize the dynamic correlations between elements of \mathbf{y} .



VECTOR MA(Q) PROCESS

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Theta}_1 \boldsymbol{\epsilon}_{t-1} + \dots + \boldsymbol{\Theta}_q \boldsymbol{\epsilon}_{t-q}$$

$\boldsymbol{\epsilon}_t \sim VWN$, $\boldsymbol{\Theta}_j$ ($N \times N$) matrix of MA coefficients $j = 1, 2, \dots, q$.

$$E(\mathbf{y}_t) = \boldsymbol{\mu}$$

$$\begin{aligned} \boldsymbol{\Gamma}_0 &= E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})'] \\ &= E[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t'] + \boldsymbol{\Theta}_1 E[\boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}'] \boldsymbol{\Theta}_1' + \dots + \boldsymbol{\Theta}_q E[\boldsymbol{\epsilon}_{t-q} \boldsymbol{\epsilon}_{t-q}'] \boldsymbol{\Theta}_q' \\ &= \boldsymbol{\Omega} + \boldsymbol{\Theta}_1 \boldsymbol{\Omega} \boldsymbol{\Theta}_1' + \dots + \boldsymbol{\Theta}_q \boldsymbol{\Omega} \boldsymbol{\Theta}_q' \end{aligned}$$



VECTOR MA(Q) PROCESS

$$\begin{aligned}\Gamma_j &= E[(\epsilon_t + \Theta_1 \epsilon_{t-1} + \dots + \Theta_q \epsilon_{t-q})(\epsilon_{t-j} + \Theta_1 \epsilon_{t-j-1} + \dots + \Theta_q \epsilon_{t-j-q})'] \\ \Gamma_j &= \begin{cases} \Theta_j \Omega + \Theta_{j+1} \Omega \Theta_1' + \Theta_{j+2} \Omega \Theta_2' + \dots + \Theta_q \Omega \Theta_{q-1}' & j = 1, \dots, q \\ \Omega \Theta_{-j}' + \Theta_1 \Omega \Theta_{-j+1}' + \dots + \Theta_{q+j} \Omega \Theta_q' & j = -1, \dots, -q \\ 0 & |j| > q \end{cases}\end{aligned}$$

$\Theta_0 = \mathbf{I}_N$. Any VMA(∞) is covariance stationary.



VAR(p) AUTOCOVARIANCES

Given:

$$\mathbf{\Gamma}_j = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-j} - \boldsymbol{\mu})']$$

$$\mathbf{\Gamma}_{-j} = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+j} - \boldsymbol{\mu})']$$

then

$$\mathbf{\Gamma}_j \neq \mathbf{\Gamma}_{-j}$$

$$\mathbf{\Gamma}'_j = \mathbf{\Gamma}_{-j}$$

$$\{\mathbf{\Gamma}_j\}_{1,2} = \text{cov}(y_{1t}, y_{2t-j})$$

$$\{\mathbf{\Gamma}_j\}_{2,1} = \text{cov}(y_{2t}, y_{1t-j})$$

$$\{\mathbf{\Gamma}_{-j}\}_{1,2} = \text{cov}(y_{1t}, y_{2t+j})$$

$$\mathbf{\Gamma}_j = E[(\mathbf{y}_{t+j} - \boldsymbol{\mu})(\mathbf{y}_{(t+j)-j} - \boldsymbol{\mu})']$$

$$= E[(\mathbf{y}_{t+j} - \boldsymbol{\mu})(\mathbf{y}_t - \boldsymbol{\mu})']$$

$$\mathbf{\Gamma}'_j = E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t+j} - \boldsymbol{\mu})'] = \mathbf{\Gamma}_{-j}$$



VAR(p) AUTOCOVARIANCES

Companion form:

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

$$\boldsymbol{\Sigma} = E[\boldsymbol{\xi}_t \boldsymbol{\xi}_t']$$

$$\begin{aligned} \boldsymbol{\Sigma} &= E \left\{ \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix} [(\mathbf{y}_t - \boldsymbol{\mu})' \dots (\mathbf{y}_{t-p+1} - \boldsymbol{\mu})'] \right\} \\ &= \begin{bmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \dots & \boldsymbol{\Gamma}_{p-1} \\ \boldsymbol{\Gamma}'_1 & \boldsymbol{\Gamma}_0 & \dots & \boldsymbol{\Gamma}_{p-2} \\ \vdots & & & \vdots \\ \boldsymbol{\Gamma}'_{p-1} & \boldsymbol{\Gamma}'_{p-2} & \dots & \boldsymbol{\Gamma}_0 \end{bmatrix} \quad (Np \times Np) \end{aligned}$$



VAR(p) AUTOCOVARIANCES

$$\begin{aligned} E[\boldsymbol{\xi}_t \boldsymbol{\xi}_t'] &= E[(\mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)(\mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t)'] \\ &= \mathbf{F}E(\boldsymbol{\xi}_{t-1} \boldsymbol{\xi}_{t-1}')\mathbf{F}' + E(\mathbf{v}_t \mathbf{v}_t') \end{aligned}$$

where $\mathbf{F}E(\boldsymbol{\xi}_{t-1} \mathbf{v}_t') = 0$.

$$\boldsymbol{\Sigma} = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' + \mathbf{Q}$$

$$\text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\mathbf{F}\boldsymbol{\Sigma}\mathbf{F}') + \text{vec}(\mathbf{Q})$$

$$\text{vec}(\boldsymbol{\Sigma}) = \text{vec}(\mathbf{F}\boldsymbol{\Sigma}\mathbf{F}') + \text{vec}(\mathbf{Q})$$

$$\begin{aligned} \text{vec}(\boldsymbol{\Sigma}) &= (\mathbf{F} \otimes \mathbf{F})\text{vec}(\boldsymbol{\Sigma}) + \text{vec}(\mathbf{Q}) \\ &= [\mathbf{I}_{(Np)^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1} \text{vec}(\mathbf{Q}) \end{aligned}$$

The eigenvalues of $(\mathbf{F} \otimes \mathbf{F})$ are all of the form $\lambda_j \lambda_j$ where λ_i and λ_j are eigenvalues of \mathbf{F} . Since $|\lambda_i| < 1, \forall i$, it follows that all eigenvalues of $(\mathbf{F} \otimes \mathbf{F})$ are inside the unit circle $|\mathbf{I}_{(Np)^2} - (\mathbf{F} \otimes \mathbf{F})| \neq 0$.



VAR(p) AUTOCOVARIANCES

The j -th autocovariance of $\boldsymbol{\xi}_t$ is

$$E[\boldsymbol{\xi}_t \boldsymbol{\xi}'_{t-j}] = \mathbf{F} E[\boldsymbol{\xi}_{t-1} \boldsymbol{\xi}'_{t-j}] + E[\mathbf{v}_t \boldsymbol{\xi}'_{t-j}]$$

$$\boldsymbol{\Sigma}_j = \mathbf{F} \boldsymbol{\Sigma}_{j-1}$$

$$\boldsymbol{\Sigma}_j = \mathbf{F}^j \boldsymbol{\Sigma} \quad j = 1, 2, \dots$$

The j -th autocovariance of \mathbf{y}_t is given by the first rows and n columns of $\boldsymbol{\Sigma}_j$

$$\boldsymbol{\Gamma}_j = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}_{j-1} + \dots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}_{j-p}$$



MAXIMUM LIKELIHOOD ESTIMATION

$$\mathbf{y}_t = \mathbf{c} + \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t \quad (1)$$

$$\epsilon_t \sim i.i.d.N(\mathbf{0}, \Omega)$$

$(T + p)$ observations. Conditioning on the first p observations we estimate using the last T observations.

Conditional likelihood:

$$f_{Y_T, Y_{T-1}, \dots, Y_1 | Y_0, \dots, Y_{1-p}}(\mathbf{y}_T, \dots, \mathbf{y}_1 | \mathbf{y}_0, \dots, \mathbf{y}_{1-p}; \theta)$$



MAXIMUM LIKELIHOOD ESTIMATION

$$\boldsymbol{\theta} = (\mathbf{c}', \text{vec}(\boldsymbol{\Phi}_1)', \text{vec}(\boldsymbol{\Phi}_2)', \dots, \text{vec}(\boldsymbol{\Phi}_p)', \text{vech}(\boldsymbol{\Omega})')'$$

$$\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_{-p+1} \sim N(\mathbf{c} + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{y}_{t-p}, \boldsymbol{\Omega})$$

$$\mathbf{x}_t = \begin{bmatrix} 1 \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-p} \end{bmatrix} \quad (Np + 1) \times 1$$

$$\boldsymbol{\Pi}' \equiv \begin{bmatrix} \mathbf{c} & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_p \end{bmatrix} \quad (N \times (Np + 1))$$

$$E(\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_{-p+1}) = \boldsymbol{\Pi}' \mathbf{x}_t$$



MAXIMUM LIKELIHOOD ESTIMATION

$$f_{Y_t|Y_{t-1}, \dots, Y_{1-p}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_{1-p}; \boldsymbol{\theta}) = \\ (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Omega}^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)' \boldsymbol{\Omega}^{-1}(\mathbf{y}_t - \boldsymbol{\Pi}'\mathbf{x}_t)\right)$$

The joint density, conditional on the $\mathbf{y}_0, \dots, \mathbf{y}_{1-p}$

$$f_{Y_t, \dots, Y_1 | Y_0, \dots, Y_{1-p}}(\mathbf{y}_t, \dots, \mathbf{y}_1 | \mathbf{y}_0, \dots, \mathbf{y}_{1-p}; \boldsymbol{\theta}) = f_{Y_t | Y_{t-1}, \dots, Y_{1-p}}(\boldsymbol{\theta}) \times \\ f_{Y_{t-1}, \dots, Y_1 | Y_0, \dots, Y_{1-p}}(\boldsymbol{\theta})$$



MAXIMUM LIKELIHOOD ESTIMATION

Recursively, the likelihood function for the full sample, conditioning on $\mathbf{y}_0, \dots, \mathbf{y}_{1-p}$, is the product of the single conditional densities:

$$f_{Y_T, \dots, Y_1 | Y_0, \dots, Y_{1-p}} = \prod_{t=1}^T f_{Y_t | Y_{t-1}, \dots, Y_{1-p}}(\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_{1-p}; \boldsymbol{\theta})$$

The log likelihood function:

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{TN}{2} \log(2\pi) + \frac{T}{2} \log |\boldsymbol{\Omega}^{-1}| - \frac{1}{2} \sum_{t=1}^T [(\mathbf{y}_t - \boldsymbol{\Pi}' \mathbf{x}_t)' \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \boldsymbol{\Pi}' \mathbf{x}_t)]$$



MAXIMUM LIKELIHOOD ESTIMATION

The ML estimate of $\mathbf{\Pi}$:

$$\hat{\mathbf{\Pi}}' = \left[\sum_{t=1}^T y_t \mathbf{x}'_t \right] \left[\sum_t \mathbf{x}_t \mathbf{x}'_{t=1} \right]^{-1}$$

The j -th row of $\hat{\mathbf{\Pi}}'$ is:

$$\mathbf{u}'_j \hat{\mathbf{\Pi}}' = \mathbf{u}'_j \left[\sum_{t=1}^T y_t \mathbf{x}'_t \right] \left[\sum_t \mathbf{x}_t \mathbf{x}'_t \right]^{-1}$$

This is the estimated coefficient vector from an OLS regression of y_{jt} on \mathbf{x}_t . ML estimates are found by an OLS regression of y_{jt} on a constant and p lags of all the variables in the system.

The ML estimate of $\mathbf{\Omega}$ is given by:

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



Asymptotic distribution of $\hat{\Pi}$

The ML estimates $\hat{\Pi}$ and $\hat{\Omega}$ will give consistent estimates of the population parameters even if the true innovations are non-gaussian:

$$\Phi(L)\mathbf{y}_t = \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\epsilon}_t \sim i.i.d.(\mathbf{0}, \boldsymbol{\Omega})$$

$$E[\epsilon_{it}\epsilon_{jt}\epsilon_{lt}\epsilon_{mt}] < \infty \quad \forall i, j, l, m$$

the roots of

$$|\mathbf{I}_N - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_p z^p| = 0$$

Let $K \equiv N \cdot p + 1$, \mathbf{x}'_t ($1 \times K$):

$$\mathbf{x}'_t = [1, \mathbf{y}'_t, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}]$$

$$\hat{\boldsymbol{\pi}}_T = \text{vec}(\hat{\boldsymbol{\Pi}}_T)$$



MAXIMUM LIKELIHOOD ESTIMATION

$$\hat{\boldsymbol{\pi}}_T = \begin{bmatrix} \hat{\boldsymbol{\pi}}_{1,T} \\ \vdots \\ \hat{\boldsymbol{\pi}}_{N,T} \end{bmatrix}$$

$$\hat{\boldsymbol{\pi}}_{i,t} = \left(\sum_t \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_t \mathbf{x}_t y_{it} \right)$$

$$\hat{\boldsymbol{\Omega}}_T = \frac{1}{T} \sum_t \hat{\boldsymbol{\epsilon}}_t \hat{\boldsymbol{\epsilon}}_t'$$

$$\hat{\boldsymbol{\epsilon}}_{it} = y_{it} - \mathbf{x}_t' \hat{\boldsymbol{\pi}}_{i,T}$$



MAXIMUM LIKELIHOOD ESTIMATION

Then,

$$\frac{1}{T} \sum_t \mathbf{x}_t \mathbf{x}_t' \xrightarrow{p} \mathbf{Q} = E(\mathbf{x}_t \mathbf{x}_t')$$

$$\hat{\boldsymbol{\pi}}_T \xrightarrow{p} \boldsymbol{\pi}$$

$$\hat{\boldsymbol{\Omega}}_T \xrightarrow{p} \boldsymbol{\Omega}$$

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_T - \boldsymbol{\pi}) \xrightarrow{d} N(0, (\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1}))$$

$$\sqrt{T}(\hat{\boldsymbol{\pi}}_{i,T} - \boldsymbol{\pi}) \xrightarrow{d} N(0, (\sigma_i^2 \mathbf{Q}^{-1}))$$

$$\sigma_i^2 = E(\epsilon_{it}^2)$$

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_t \hat{\epsilon}_{it}^2 \xrightarrow{p} \sigma_i^2$$



MAXIMUM LIKELIHOOD ESTIMATION

$$\hat{\pi}_i \approx N \left(\pi_i, \hat{\sigma}_i^2 \left(\sum_t \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \right)$$

OLS t and F statistics applied to the coefficients of any single equation in the VAR are asymptotically valid. A more general hypothesis

$$\mathbf{R}\pi = \mathbf{d}$$

can be tested using a generalization of the Wald form of the OLS χ^2 test

$$\sqrt{T}(\mathbf{R}\hat{\pi}_T - \mathbf{d}) \xrightarrow{d} N(\mathbf{0}, \mathbf{R}(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1})\mathbf{R}')$$



MAXIMUM LIKELIHOOD ESTIMATION

$$\begin{aligned}\xi^2(m) &= T(\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d})' \left[\mathbf{R} \left(\hat{\boldsymbol{\Omega}}_T \otimes \hat{\mathbf{Q}}_T^{-1} \right) \mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d}) \\ &= (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d})' \left[\mathbf{R} \left(\hat{\boldsymbol{\Omega}}_T \otimes (T\hat{\mathbf{Q}}_T)^{-1} \right) \mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d}) \\ &= (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d})' \left[\mathbf{R} \left(\hat{\boldsymbol{\Omega}}_T \otimes \left(\sum_t \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \right) \mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\pi}}_T - \mathbf{d})\end{aligned}$$