Vector AutoRegression Model

Eduardo Rossi
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:

$$ y_t = c + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \epsilon_t $$

$y_t: (N \times 1)$  $\Phi_i: (N \times N)$  $\forall i$,  $\epsilon_t: (N \times 1)$

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

$\Omega$ positive definite matrix.
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.

\[
\begin{align*}
[I_N - \Phi_1 L - \cdots - \Phi_p L^p] y_t &= c + \epsilon_t \\
\Phi(L)y_t &= c + \epsilon_t
\end{align*}
\]

with

\[
\Phi(L) = [I_N - \Phi_1 L - \cdots - \Phi_p L^p]
\]

\( \Phi(L) \ (N \times N) \) matrix polynomial in \( L \).
The element \((i, j)\) in \(\Phi(L)\) is a scalar polynomial in \(L\)

\[
\delta_{ij} - \phi_{ij}^{(1)} L - \phi_{ij}^{(2)} L^2 - \ldots \phi_{ij}^{(p)} L^p
\]

\[
\delta_{ij} = \begin{cases} 
1 & i = j \\
0 & i \neq j 
\end{cases}
\]

**Stationarity**: A vector process is said *covariance stationary* if its first and second moments, \(E[y_t]\) and \(E[y_t y'_{t-j}]\) respectively, are independent of the date \(t\).
\[ y_t = c + \Phi_1 y_{t-1} + \epsilon_t \]

First equation:

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

\[ y_t = c + \Phi_1 [c + \Phi_1 y_{t-2} + \epsilon_{t-1}] + \epsilon_t \]

\[ = c + \Phi_1 c + \Phi_1^2 y_{t-2} + \epsilon_t + \Phi_1 \epsilon_{t-1} \]

\[ y_t = \ldots \]

\[ y_t = c + \Phi_1 c + \ldots + \Phi_1^{k-1} c + \Phi_1^k y_{t-k} + \epsilon_t + \Phi_1 \epsilon_{t-1} + \ldots + \Phi_1^{k-1} \epsilon_{t-k+1} \]

\[ E[y_t] = \sum_{j=0}^{k-1} \Phi_1^j c + \Phi_1^k E[y_{t-k}] \]

The value of this sum depends on the behavior of \( \Phi_1^j \) as \( j \) increases.
Stability of VAR(1)

Let $\lambda_1, \lambda_2, \ldots, \lambda_N$ be the eigenvalues of $\Phi_1$, the solutions to the characteristic equation

$$|\Phi_1 - \lambda I_N| = 0$$

then, if the eigenvalues are all distinct

$$\Phi_1 = QMQ^{-1}$$

$$\Phi_1^k = QM^kQ^{-1}$$

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

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

$$\Phi^k \to 0, \ k \to \infty$$

If $|\lambda_i| \geq 1, \ \forall i$ then one or more elements of $M^k$ are not vanishing, and may be tending to $\infty$. 
VAR(p): 

\[ y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t \]

as a VAR(1) (Companion Form):

\[ \xi_t = F \xi_{t-1} + v_t \]

\[ \xi_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad (Np \times 1) \quad \xi_{t-1} = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad (Np \times 1) \]
VAR(P) Companion form

\[ F = \begin{bmatrix}
\Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\
I_N & 0 & \cdots & 0 & 0 \\
0 & I_N & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & I_N & 0
\end{bmatrix} \]

\[ (N_p \times N_p) \quad v_t = \begin{bmatrix}
\epsilon_t \\
0 \\
\vdots \\
0
\end{bmatrix} \quad (N_p \times 1) \]

\[ E[v_tv'_\tau] = \begin{cases}
Q & t = \tau \\
0 & t \neq \tau
\end{cases} \quad Q = \begin{bmatrix}
\Omega & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix} \quad (N_p \times N_p) \]
Stability of VAR(p)

\[ y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \epsilon_t \quad t = 1, \ldots, T \]

If the process \( y_t \) has finite variance and an autocovariance sequence that converges to zero at an exponential rate, then \( \xi_t \) must share these properties. This is ensured by having the \( Np \) eigenvalues of \( F \) lie inside the unit circle.

The determinant defining the characteristic equation is

\[ |F - \lambda I_{Np}| = (-1)^{Np}|\lambda^p I_N - \lambda^{p-1} \Phi_1 - \ldots - \Phi_p| = 0 \]
The required condition is that the roots of the equation

$$|\lambda^p I_N - \lambda^{p-1} \Phi_1 - \ldots - \Phi_p| = 0$$

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 $p$. 
When $p = 1$, the roots of

$$|I_N - \Phi_1 z| = 0$$

outside the unit circle, i.e. $|z| > 1$, implies that the eigenvalues of $\Phi_1$ be inside the unit circle. Note that the eigenvalues, roots of $$|\Phi_1 - \lambda I_N| = 0,$$
are the reciprocal of the roots of $$|I_N - \Phi_1 z| = 0.$$
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 \( F \).

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

\[
y_t = c + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \epsilon_t
\]

\[
\mu = c + \Phi_1 \mu + \ldots + \Phi_p \mu
\]

\[
\mu = [I_N - \Phi_1 \ldots - \Phi_p]^{-1} c
\]

\[
= \Phi(1)^{-1} c
\]
If the VAR(p) is stationary then it has a VMA(\infty) representation:

\[ y_t = \mu + \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \ldots \equiv \mu + \Psi(L) \epsilon_t \]

\( y_{t-j} \) is a linear function of \( \epsilon_{t-j}, \epsilon_{t-j-1}, \ldots \) each of which is uncorrelated with \( \epsilon_{t+1} \) for \( j = 0, 1, \ldots \).

It follows that

- \( \epsilon_{t+1} \) is uncorrelated with \( y_{t-j} \) for any \( j \geq 0 \).
- Linear forecast of \( y_{t+1} \) based on \( y_t, y_{t-1}, \ldots \) is given by

\[
\hat{y}_{t+1|t} = \mu + \Phi_1(y_t - \mu) + \Phi_2(y_{t-1} - \mu) + \ldots + (y_{t-p+1} - \mu)
\]
Forecasting with VAR

$\epsilon_{t+1}$ can be interpreted as the fundamental innovation in $y_{t+1}$, that is the error in forecasting $y_{t+1}$ on the basis of a linear function of a constant and $y_t, y_{t-1}, \ldots$.

A forecast of $y_{t+s}$ on the basis of $y_t, y_{t-1}, \ldots$ will take the form

$$\hat{y}_{t+s|t} = \mu + F_{11}^{(s)}(y_t - \mu) + F_{12}^{(s)}(y_{t-1} - \mu) + \ldots + F_{1p}^{(s)}(y_{t-p+1} - \mu)$$

The moving average matrices $\Psi_j$ can be calculated as:

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

$$\Psi(L)\Phi(L) = I_N$$
VMA coefficient matrices

\[
\begin{bmatrix}
I_N + \Psi_1 L + \Psi_2 L^2 + \ldots \end{bmatrix}
\begin{bmatrix}
I_N - \Phi_1 L - \Phi_2 L^2 + \ldots - \Phi_p L^p \end{bmatrix} = I_N
\]

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

\[
\Psi_1 - \Phi_1 = 0
\]
on \( L^2 \)

\[
\Psi_2 = \Psi_1 \Phi_1 + \Phi_2
\]

In general for \( L^s \)

\[
\Psi_s = \Phi_1 \Psi_{s-1} + \ldots + \Phi_p \Psi_{s-p} \quad s = 1, 2, \ldots
\]

\[
\Psi_0 = I_N, \quad \Psi_s = 0 \quad s < 0
\]
The innovation in the MA(∞) representation is $\epsilon_t$, the fundamental innovation for $y$.
There are alternative MA representation based on VWN processes other than $\epsilon_t$.
Let $H$ be a nonsingular $(N \times N)$ matrix

$$u_t \equiv H \epsilon_t$$

$u_t \sim VWN$.

$$y_t = \mu + H^{-1}H \epsilon_t + \Psi_1 H^{-1}H \epsilon_{t-1} + \ldots$$
$$= \mu + J_0 u_t + J_1 u_{t-1} + J_2 u_{t-2} + \ldots$$

$$J_s \equiv \Psi_s H^{-1}$$
For example $H$ can be any matrix that diagonalizes $\Omega$, the var-cov of $\epsilon_t$,

$$H\Omega H' = D$$

the elements of $u_t$ 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 $\Psi_0 = I_N$ (while $J_0$ is not the identity matrix).
Assumptions implicit in a VAR

- For a covariance stationary process, the parameters \( c, \Phi_1, \ldots, \Phi_p \) could be defined as the coefficients of the projections of \( y_t \) on \( 1, y_{t-1}, y_{t-2}, \ldots, y_{t-p} \).

- \( \epsilon_t \) is uncorrelated with \( y_{t-1}, \ldots, y_{t-p} \) by the definition of \( \Phi_1, \ldots, \Phi_p \).

- The parameters of a VAR can be estimated consistently with \( n \) OLS regressions.

- \( \epsilon_t \) defined by this projection is uncorrelated with \( y_{t-p-1}, y_{t-p-2}, \ldots \).

- The assumption of \( y_t \sim VAR(p) \) is basically the assumption that \( p \) lags are sufficient to summarize the dynamic correlations between elements of \( y \).
Vector MA(q) Process

\[ y_t = \mu + \epsilon_t + \Theta_1 \epsilon_{t-1} + \ldots + \Theta_q \epsilon_{t-q} \]

\( \epsilon_t \sim VWN, \Theta_j (N \times N) \) matrix of MA coefficients \( j = 1, 2, \ldots, q \).

\[ E(y_t) = \mu \]

\[ \Gamma_0 = E[(y_t - \mu)(y_t - \mu)'] \]
\[ = E[\epsilon_t \epsilon_t'] + \Theta_1 E[\epsilon_{t-1} \epsilon_{t-1}'] \Theta_1' + \ldots + \Theta_q E[\epsilon_{t-q} \epsilon_{t-q}'] \Theta_q' \]
\[ = \Omega + \Theta_1 \Omega \Theta_1' + \ldots + \Theta_q \Omega \Theta_q' \]
Vector MA(q) Process

\[
\Gamma_j = E[(\epsilon_t + \Theta_1 \epsilon_{t-1} + \ldots + \Theta_q \epsilon_{t-q})(\epsilon_{t-j} + \Theta_1 \epsilon_{t-j-1} + \ldots + \Theta_q \epsilon_{t-j-q})']
\]

\[
\Gamma_j = \begin{cases} 
\Theta_j \Omega + \Theta_{j+1} \Omega \Theta_1' + \Theta_{j+2} \Omega \Theta_2' + \ldots + \Theta_q \Omega \Theta_{q-1}' & j = 1, \ldots, q \\
\Omega \Theta_{-j}' + \Theta_1 \Omega \Theta_{-j+1}' + \ldots + \Theta_{q+j} \Omega \Theta_q' & j = -1, \ldots, -q \\
0 & |j| > q 
\end{cases}
\]

\[\Theta_0 = I_N. \text{ Any VMA}(\infty) \text{ is covariance stationary.}\]
VAR(p) Autocovariances

Given:
\[ \Gamma_j = E[(y_t - \mu)(y_{t-j} - \mu)'] \]
\[ \Gamma_{-j} = E[(y_t - \mu)(y_{t+j} - \mu)'] \]

then
\[ \Gamma_j \neq \Gamma_{-j} \]
\[ \Gamma'_j = \Gamma_{-j} \]

\[ \{\Gamma_j\}_{1,2} = cov(y_{1t}, y_{2t-j}) \]
\[ \{\Gamma_j\}_{2,1} = cov(y_{2t}, y_{1t-j}) \]
\[ \{\Gamma_{-j}\}_{1,2} = cov(y_{1t}, y_{2t+j}) \]

\[ \Gamma_j = E[(y_{t+j} - \mu)(y_{(t+j)-j} - \mu)'] \]
\[ = E[(y_{t+j} - \mu)(y_t - \mu)'] \]
\[ \Gamma'_j = E[(y_t - \mu)(y_{t+j} - \mu)'] = \Gamma_{-j} \]
VAR(p) Autocovariances

Companion form:

$$\xi_t = F\xi_{t-1} + v_t$$

$$\Sigma = E[\xi_t \xi_t']$$

$$\Sigma = E \begin{pmatrix}
    y_t - \mu \\
    y_{t-1} - \mu \\
    \vdots \\
    y_{t-p+1} - \mu
\end{pmatrix}
\begin{pmatrix}
    (y_t - \mu)' \\
    \cdots \\
    (y_{t-p+1} - \mu)'
\end{pmatrix}$$

$$= \begin{bmatrix}
    \Gamma_0 & \Gamma_1 & \ldots & \Gamma_{p-1} \\
    \Gamma'_1 & \Gamma_0 & \ldots & \Gamma_{p-2} \\
    \vdots & \vdots & \ddots & \vdots \\
    \Gamma'_{p-1} & \Gamma'_{p-2} & \ldots & \Gamma_0
\end{bmatrix}
(Np \times Np)$$
VAR(p) Autocovariances

\[
E[\xi_t \xi_t'] = E[(F\xi_{t-1} + v_t)(F\xi_{t-1} + v_t)'] \\
= FE(\xi_{t-1} \xi_{t-1}')F' + E(v_tv_t')
\]

where \(FE(\xi_{t-1}v_t') = 0\).

\[
\Sigma = F\Sigma F' + Q
\]

\[
vec(\Sigma) = vec(F\Sigma F') + vec(Q)
\]

\[
vec(\Sigma) = vec(F\Sigma F') + vec(Q)
\]

\[
vec(\Sigma) = (F \otimes F)vec(\Sigma) + vec(Q)
\]

\[
= [I_{(Np)^2} - (F \otimes F)]^{-1}vec(Q)
\]

The eigenvalues of \((F \otimes F)\) are all of the form \(\lambda_j \lambda_j\) where \(\lambda_i\) and \(\lambda_j\) are eigenvalues of \(F\). Since \(|\lambda_i| < 1, \forall i\), it follows that all eigenvalues of \((F \otimes F)\) are inside the unit circle \(|I_{(Np)^2} - (F \otimes F)| \neq 0\).
The $j$-th autocovariance of $\xi_t$ is

$$E[\xi_t \xi'_{t-j}] = FE[\xi_{t-1} \xi'_{t-j}] + E[v_t \xi'_{t-j}]$$

$$\Sigma_j = F \Sigma_{j-1}$$

$$\Sigma_j = F^j \Sigma \quad j = 1, 2, \ldots$$

The $j$-th autocovariance of $y_t$ is given by the first rows and $n$ columns of $\Sigma_j$

$$\Gamma_j = \Phi_1 \Gamma_{j-1} + \ldots + \Phi_p \Gamma_{j-p}$$
Maximum Likelihood Estimation

\[ y_t = c + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \epsilon_t \]  

\[ \epsilon_t \sim i.i.d.N(0, \Omega) \]

\((T + p)\) observations. Conditioning on the first \(p\) observations we estimate using the last \(T\) observations. Conditional likelihood:

\[ f_{Y_T, Y_{T-1}, \ldots, Y_1 | Y_0, \ldots, Y_{1-p}}(y_T, \ldots, y_1 | y_0, \ldots, y_{1-p}; \theta) \]
Maximum Likelihood Estimation

\[ \theta = (c', vec(\Phi_1)', vec(\Phi_2)', \ldots, vec(\Phi_p)', vech(\Omega)')' \]

\[ y_t | y_{t-1}, \ldots, y_{-p+1} \sim N(c + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p}, \Omega) \]

\[ x_t = \begin{bmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \quad (Np + 1) \times 1 \]

\[ \Pi' \equiv \begin{bmatrix} c & \Phi_1 & \Phi_2 & \ldots & \Phi_p \end{bmatrix} \quad (N \times (Np + 1)) \]

\[ E(y_t | y_{t-1}, \ldots, y_{-p+1}) = \Pi' x_t \]
The joint density, conditional on the $y_0, \ldots, y_{1-p}$

\[
f_{Y_t, \ldots, Y_1 | Y_0, \ldots, Y_{1-p}}(y_t, \ldots, y_1 | y_0, \ldots, y_{1-p}; \theta) = f_{Y_t | Y_{t-1}, \ldots, Y_{1-p}}(\theta) \times f_{Y_{t-1}, \ldots, Y_1 | Y_0, \ldots, Y_{1-p}}(\theta)
\]
Maximum Likelihood Estimation

Recursively, the likelihood function for the full sample, conditioning on $y_0, \ldots, y_{1-p}$, is the product of the single conditional densities:

$$f_{Y_T, \ldots, Y_1|Y_0, \ldots, Y_{1-p}} = \prod_{t=1}^{T} f_{Y_t|Y_{t-1}, \ldots, Y_{1-p}}(y_t|y_{t-1}, \ldots, y_{1-p}; \theta)$$

The log likelihood function:

$$\mathcal{L}(\theta) = -\frac{TN}{2} \log (2\pi) + \frac{T}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{t=1}^{T} [(y_t - \Pi' x_t)' \Omega^{-1} (y_t - \Pi' x_t)]$$
Maximum Likelihood Estimation

The ML estimate of \( \Pi \):

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

The \( j \)-th row of \( \hat{\Pi}' \) is:

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

This is the estimated coefficient vector from an OLS regression of \( y_{jt} \) on \( x_t \). 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 \( \Omega \) is given by:

\[
\hat{\Omega} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}'_t
\]
Maximum Likelihood Estimation

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:

$$\Phi(L)y_t = \epsilon_t$$
$$\epsilon_t \sim i.i.d.(0, \Omega)$$
$$E[\epsilon_{it}\epsilon_{jt}\epsilon_{lt}\epsilon_{mt}] < \infty \quad \forall i, j, l, m$$

the roots of

$$|I_N - \Phi_1 z - \ldots - \Phi_p z^p| = 0$$

Let $K \equiv N \cdot p + 1$, $x_t'$ ($1 \times K$):

$$x_t' = [1, y_t', y_{t-1}', \ldots, y_{t-p}']$$

$$\hat{\pi}_T = vec(\hat{\Pi}_T)$$
Maximum Likelihood Estimation

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

\[ \hat{\pi}_{i,t} = \left( \sum_t x_t x_t' \right)^{-1} \left( \sum_t x_t y_{it} \right) \]

\[ \hat{\Omega}_T = \frac{1}{T} \sum_t \hat{\epsilon}_t \hat{\epsilon}_t' \]

\[ \hat{\epsilon}_{it} = y_{it} - x_t' \hat{\pi}_{i,T} \]
Then,

\[
\frac{1}{T} \sum_{t} x_t x_t' \xrightarrow{p} Q = E(x_t x_t')
\]

\[\hat{\pi}_T \xrightarrow{p} \pi\]

\[\hat{\Omega}_T \xrightarrow{p} \Omega\]

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

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

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

\[\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t} \hat{\epsilon}_{it}^2 \xrightarrow{p} \sigma_i^2\]
Maximum Likelihood Estimation

\[ \hat{\pi}_i \approx N \left( \pi_i, \hat{\sigma}_i^2 \left( \sum_t x_t x_t' \right)^{-1} \right) \]

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

\[ R\pi = d \]

can be tested using a generalization of the Wald form of the OLS \( \chi^2 \) test

\[ \sqrt{T}(R\hat{\pi}_T - d) \xrightarrow{d} N(0, R (\Omega \otimes Q^{-1}) R') \]
\[ \xi^2(m) = T (R\hat{\pi}_T - d)' \left[ R \left( \hat{\Omega}_T \otimes \hat{Q}_T^{-1} \right) R' \right]^{-1} (R\hat{\pi}_T - d) \]

\[ = (R\hat{\pi}_T - d)' \left[ R \left( \hat{\Omega}_T \otimes (T\hat{Q}_T)^{-1} \right) R' \right]^{-1} (R\hat{\pi}_T - d) \]

\[ = (R\hat{\pi}_T - d)' \left[ R \left( \hat{\Omega}_T \otimes \left( \sum_t x_t x_t' \right)^{-1} \right) R' \right]^{-1} (R\hat{\pi}_T - d) \]