

# **Vectores Autorregresivos (VAR)**

## Procesos estocásticos multivariados

$$Y_t = [Y_{1t}, Y_{2t}, \dots, Y_{Nt}]', \quad t = 1, 2, \dots, T$$

Estamos interesados en el comportamiento temporal de  $N$  variables simultáneamente.

- $E(Y_t) = \mu$  (vector de esperanzas)
- $E(Y_t - \mu)(Y_{t-\tau} - \mu)' = \Gamma(\tau)$  (matriz de autocovarianzas)

El elemento  $i, j$  es:  $\gamma_{i,j} = E(Y_{it} - \mu_{it})(Y_{j,t-\tau} - \mu_{j,t-\tau})$

Los elementos de la diagonal son las autocovarianzas y los de fuera de la diagonal las *autocovarianzas cruzadas*.

$Y_t$  es *estacionario* si:

1.  $E(Y_t) = \mu < \infty$ .

2.  $E(Y_t - \mu)(Y_{t-\tau} - \mu)' = \Gamma(\tau) < \infty, \quad \tau = 0, 1, 2, \dots$

Estacionariedad de  $Y_t$  implica estacionariedad de  $Y_{it}, i = 1, \dots, N$ . El converso NO es cierto (porque?)

## Ruido Blanco Multivariado

$\varepsilon_t$  es  $N \times 1$ . Es RBM si:

1.  $E(\varepsilon_t) = 0$

2. 
$$E(\varepsilon_t \varepsilon'_s) = \begin{cases} \Omega_{N \times N} & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}$$

**Cuidado:** los elementos de  $\varepsilon_t$  pueden estar correlacionados contemporaneamente pero no en distintos periodos!!

## Vectores Autorregresivos (VAR)

*VAR(1) con media cero*

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim RBM$$

$Y_t$  es  $N \times 1$ ,  $\Phi$  es  $N \times N$ .

*Ejemplo:*

$$\begin{cases} X_t &= \phi_{11}X_{t-1} + \phi_{12}Z_{t-1} + \varepsilon_{1t} \\ Z_t &= \phi_{21}X_{t-1} + \phi_{22}Z_{t-1} + \varepsilon_{2t} \end{cases}$$

Relacion entre el *presente* de una variable y el pasado de *todas* las variables del sistema.

- $\varepsilon_{1t}$  y  $\varepsilon_{2t}$  pueden estar correlacionados.

## Motivacion historica y metodologica del uso de VAR's

- Motivacion: Sims (1980).
- Discusion metodologica: Pagan (1987)

## Algunos resultados utiles de algebra

- Si  $A$  es cuadrada y diagonal con elemento caracteristico  $a_{ii}$ ,  $A^K$  es una matriz diagonal con elemento  $i, i$  igual a  $a_{ii}^K$ .
- $A_{m \times m}$ . Cualquier escalar  $\lambda$  que satisfaga  $Ax = \lambda x$  para un vector  $m \times 1, x \neq 0$  es un *autovalor* de  $A$ .  $x$  es un *autovector* de  $A$  correspondiente al autovalor  $\lambda$ .
- Si  $\lambda$  es un autovalor de  $A$ , entonces  $|A - \lambda I| = 0$ .
- La *ecuacion caracteristica*  $|A - \lambda I| = 0$  es una ecuacion de grado  $m$ . Entonces,  $A_{mm}$  tiene  $m$  autovalores.
- $A_{m \times m}$  con autovalores  $\lambda_1, \lambda_2, \dots, \lambda_m$ , entonces:
  - (a)  $\text{tr}(A) = \sum_{i=1}^m \lambda_i$
  - (b)  $|A| = \prod_{i=1}^m \lambda_i$

- Si  $A_{m \times m}$  tiene  $m$  autovalores diferentes, entonces los autovectores asociados son todos linealmente independientes.
- Si  $A_{m \times m}$  tiene  $m$  autovalores diferentes y  $\Lambda \equiv \text{diag}(\lambda_1, \dots, \lambda_m)$  y  $X \equiv (x_1, x_2, \dots, x_m)$ , entonces:

$$A = X\Lambda X^{-1}$$

- Bajo las condiciones del resultado anterior:

$$A^K = AA \cdots A = (X\Lambda X^{-1})(X\Lambda X^{-1}) \cdots (X\Lambda^K X^{-1}) = X\Lambda^K X^{-1}$$

$$\text{con } \Lambda^K = \text{diag}(\lambda_1^K, \dots, \lambda_m^K)$$



## Estacionariedad de VAR(1)

Recordar caso AR(1):  $Y_t = \phi Y_{t-1} + \varepsilon_t$ . Estacionario si  $|\phi| < 1$ .

Notar que  $Y_t = \Phi Y_{t-1} + \varepsilon_t$  y  $Y_{t-1} = \Phi Y_{t-2} + \varepsilon_{t-1}$ .

Reemplazando:

$$Y_t = \Phi^2 Y_{t-2} + \varepsilon_t + \Phi \varepsilon_{t-1}$$

Reemplazando ahora  $Y_{t-2}$  y luego  $Y_{t-3}$ , etc.:

$$Y_t = \sum_{j=0}^{k-1} \Phi^j \varepsilon_{t-j} + \Phi^k Y_{t-k}$$

En forma similar a AR(1) requeriremos que  $\Phi^k \rightarrow 0$ , o, equivalentemente,  $|\Phi| < 1$  (en un sentido matricial).

De acuerdo al resultado anterior  $\Phi = X\Lambda X^{-1}$  en donde  $X$  es la matriz de autovectores y  $\Lambda$  la matriz diagonal con los autovalores  $\lambda_1, \dots, \lambda_N$ . Entonces,

$$\Phi^k = X\Lambda^k X^{-1}$$

de modo que  $\Phi^K \rightarrow 0$  si y solo si  $|\lambda_i| < 1, i = 1, \dots, N$ .

*Resumiendo:*  $Y_t = \Phi Y_{t-1} + \varepsilon_t$  es estacionario si y solo si todos los autovalores de  $\Phi$  son  $< 1$  en valor absoluto, o, equivalentemente, si todas las raíces de la ecuación característica  $|\Phi - \lambda I_N|$  son menores a uno en valor absoluto.

Si VAR(1) es estacionario:

$$Y_t = \sum_{j=0}^{\infty} \Phi^j \varepsilon_{t-j}$$

es la representacion  $VMA(\infty)$  estacionaria del VAR(1).

## VAR(p)

$$Y_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim RBM$$

Estacionariedad de VAR(p): todas las raices de

$$|\lambda^p I - \lambda^{p-1} \Phi_1 - \cdots - \Phi_p| = 0$$

son mayores que uno en valor absoluto.

**Resultado importante:** Si VAR(p) es estacionario, tiene una representacion VMA( $\infty$ ) estacionaria. Esto es una consecuencia del Teorema de Descomposicion de Wold.

## Estimacion de VAR

Consideremos VAR(1):

$$\begin{cases} X_t &= \phi_{11}X_{t-1} + \phi_{12}Z_{t-1} + \varepsilon_{1t} \\ Z_t &= \phi_{21}X_{t-1} + \phi_{22}Z_{t-1} + \varepsilon_{2t} \end{cases}$$

- Notar que MCO es consistente, ecuacion por ecuacion.
- Es un caso particular de SUR con las mismas variables explicativas en cada ecuacion: MCG coincide con MCO, no hay ganancia de eficiencia. Explica gran parte de la popularidad de VAR.

## Interpretacion de VAR

- No es una forma estructural. Los coeficientes no tienen una interpretacion clara ya que solo se busca una *representacion* de los datos.
- En la practica es comun basar la interpretacion en tests agregados y ejercicios de simulacion.
- Simulacion: a) Funciones impulso-respuesta, b) Descomposicion de la varianza.
- Tests: a) Relevancia de variables, b) Significatividad de rezagos, c) Relaciones entre ecuaciones, d) Causalidad de Granger.

## Funciones impulso-respuesta

Objetivo: investigar que sucede con el sistema VAR ante variaciones exogenas.

Consideremos la representacion VMA( $\infty$ ) de VAR(p):

$$Y_t = \sum_{s=0}^{\infty} \Psi_s \varepsilon_{t-s}, \quad \Psi_0 = I_m, \quad \varepsilon_t \sim RBM$$

$$\frac{\partial Y_t}{\partial \varepsilon_{t-j}} = \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \Psi_s$$

$\Psi_{ij}(s)$  : Efecto sobre  $Y_{i,t+s}$  de cambiar  $\varepsilon_{jt}$  en una unidad.

$\Psi_{ij}(s)/\sqrt{V(\varepsilon_{jt})}$ : Efecto sobre  $Y_{i,t+s}$  de cambiar  $\varepsilon_{jt}$  en un desvio estandar: *funcion de impulso-respuesta simple*.

*Ejemplo:* VAR(1) con  $N = 2$ , hay 4 funciones de impulso respuesta simple.

Problema:  $\varepsilon_{it}$  no son independientes.  $N = 2$

$$V(\varepsilon_t) = \Omega = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

No es intuitivo hablar de cambios en  $\varepsilon_{1t}$  dejando  $\varepsilon_{2t}$ .



## Mas resultados de algebra

1. Si  $\Omega$  es simetrica y positiva definida, existe  $H$  no singular, tal que  $\Omega = HH'$ .  $H$  es una 'raiz cuadrada' de  $\Omega$ .  $H$  no es unica.
2. Si  $\varepsilon_t$  es un vector aleatorio con  $E(\varepsilon_t) = 0$  y  $V(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \Omega$ , existe  $H$  tal que  $\varepsilon_t = H v_t$ , en donde  $v_t$  es un vector aleatorio con  $E(v_t) = 0$  y  $V(v_t) = I$ .

$v_t$  son las *innovaciones ortogonales* de  $\varepsilon_t$ . La idea es:

$$v_t \longrightarrow H v_t \longrightarrow \varepsilon_t \longrightarrow Y_t$$

Entonces, lo que realmente nos interesa es  $\partial Y_{t+s} / \partial v_t$ .

- Problema:  $H$  no es unica. Si  $\Omega$  es simetrica y positiva definida, existe una *unica* matrix  $P$  no-singular y triangular inferior tal que  $PP' = \Omega$ .  $P$  es la *matriz de Choleski* de  $\Omega$ .

La idea, entonces, es utilizar:  $\varepsilon_t = P v_t$ . Reemplazando:

$$Y_t = \sum_{i=1}^{\infty} \Psi_s P v_{t-s} = \sum_{i=1}^{\infty} \Phi_s v_{t-s}, \quad \Phi \equiv \Psi_s P$$

$$\frac{\partial Y_t}{\partial v_{t-s}} = \frac{\partial Y_{t+s}}{\partial v_t} = \Phi_s$$

Elemento característico:  $\phi_{ij}^{(s)}$  = cambio en  $Y_{i,t+s}$  resultante de cambios en  $v_{j,t}$  en una unidad.

$$\phi_{ij}^{(s)} = \text{funcion de impulso-respuesta}$$

con respecto a las innovaciones ortogonales.

Notar que como  $V(v_{j,t} = 1)$  tambien mide el efecto de cambios en 1 desvio estandar.

**Cuestion importante:** el uso de  $P$  (la descomposicion de Choleski) implica cierta arbitrariedad.

Ejemplo:  $N = 2$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

$\varepsilon_{1t}$  depende solo de  $v_{1t}$

$\varepsilon_{2t}$  depende de  $v_{1t}$  y  $v_{2t}$

La dependencia entre  $\varepsilon_{1t}$  y  $\varepsilon_{2t}$  es generada por esta estructura recursiva.

Que implicaciones tiene, por ejemplo, si  $Y_1$  es la tasa de crecimiento de la cantidad nominal de dinero y  $Y_2$  es la tasa de inflacion?

Entonces, el *orden* de las variables en la matriz de Choleski es muy importante.

Notar que si

$$\Omega = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \Rightarrow P = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix}$$

- Si  $\Omega$  es diagonal el orden de la descomposicion de Choleski no importa.
- Trivialmente, las innovaciones originales son tambien las innovaciones ortogonales.
- En la practica, entonces, es importante evaluar empiricamente la diagonalidad de  $\Omega$ .

## Seguimos con el algebra...

- Si  $\Omega$  es simetrica y pd, existe una unica matriz  $H$  triangular inferior con 1's en la diagonal principal, y una unica matriz diagonal  $D$  con todos sus elementos positivos, tal que  $\Omega = HDH'$ .
- Si  $\varepsilon_t$  es un vector aleatorio con  $E(\varepsilon_t) = 0$  y  $V(\varepsilon_t) = \Omega$ , entonces  $\varepsilon_t$  puede escribirse com  $\varepsilon_t = Hu_t$ , con  $E(u_t) = 0$  y  $V(u_t) = D$ , en donde  $H$  y  $D$  satisfacen  $\Omega = HDH'$ .

Notar que  $V(u_{jt}) = d_j$ , el  $j$ -esimo elemento de la diagonal principal de  $D$ . La idea, en este caso, es permitir que cada una de las innovaciones ortogonales tenga su propia varianza.

## Disgresion notacional

$$X = \begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \equiv \begin{bmatrix} A & B \end{bmatrix}$$

Chequear que:

$$XX' = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \begin{bmatrix} A_1 & A_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \end{bmatrix}$$

En general, si  $x_i$  es la  $i$ -esima columna de  $X$ ,  $XX' = \sum x_i x_i'$ .

Es facil verificar que si  $D$  es diagonal con elemento caracteristico  $d_i$ ,  $XD X' = \sum d_i x_i x_i'$

## Descomposicion de la varianza

- Supongamos un VAR con  $N$  variables. Cuanto de la variabilidad total de una variable es atribuible a movimientos en cada una de las otras variables?
- $R^2$  en el modelo lineal: proporcion de la varianza de  $Y$  que es explicada por movimientos en  $X$ . La idea es proporcionar una idea similar para descomponer la variabilidad total de una variable del VAR en terminos de las variabilidades exogenas de cada una de las variables del sistema.

Partimos de la representacion  $MA(\infty)$  de VAR:

$$Y_t = \sum_{s=0}^{\infty} \Psi_s \varepsilon_{t-s}, \quad \Psi_0 = I_m, \quad \varepsilon_t \sim RBM$$

que implica que:

$$\begin{aligned} V(Y_t) &= \sum_{s=0}^{\infty} \Psi_s \Omega \Psi_s' \\ &= \sum_{s=0}^{\infty} \Psi_s H D H' \Psi_s' \\ &= \sum_{s=0}^{\infty} \Psi_s \left\{ \sum_{i=1}^N d_i h_i h_i' \right\} \Psi_s' \\ &= \sum_{i=1}^N d_i \left\{ \sum_{s=0}^{\infty} \Psi_s h_i h_i' \Psi_s' \right\} \end{aligned}$$



$$\begin{aligned}
V(Y_t) &= \sum_{i=1}^N V(u_i) \left\{ \sum_{s=0}^{\infty} \Psi_s h_i h_i' \Psi_s' \right\} \\
&= \sum_{i=1}^N V(u_i) B^{(i)}
\end{aligned}$$

En particular, la  $j$ -ésima varianza:

$$\begin{aligned}
V(Y_{tj}) &= \sum_{i=1}^N V(u_i) B_{jj}^{(i)} \\
1 &= \sum_{i=1}^N \frac{V(u_i) B_{jj}^{(i)}}{V(Y_{tj})} = \sum_{i=1}^N z_{ji}
\end{aligned}$$

$z_{ji}$  = proporción de la varianza de la  $j$ —ésima variable que es explicada por el componente exógeno de cada una de las  $i = 1, 2, \dots, N$  variables del sistema.

*En la practica:* no podemos computar  $\infty$  rezagos!

Partimos de la representacion  $VMA(\infty)$  de VAR en  $t + 1$ :

$$Y_{t+1} = \varepsilon_{t+1} + \Psi_1 \varepsilon_t + \Psi_2 \varepsilon_{t-1} + \dots$$

La prediccion de  $Y_{t+1}$  con la informacion disponible en  $t$  es la esperanza de  $Y_{t+1}$  condicional en toda la informacion disponible en  $t$ :

$$E_t(Y_{t+1}) = \Psi_1 \varepsilon_t + \Psi_2 \varepsilon_{t-1} + \dots$$

El error de pronostico de predecir  $Y_{t+1}$  en  $t$  sera:

$$Y_{t+1} - E_t(Y_{t+1}) = \varepsilon_{t+1}$$

Analogamente, el error de pronóstico de predecir  $Y_{t+s}$  en  $t$  será:

$$e_{t+s} \equiv Y_{t+s} - E_t(Y_{t+s}) = \varepsilon_{t+s} + \Psi_1 \varepsilon_{t+s-1} + \Psi_2 \varepsilon_{t+s-2} + \cdots + \Psi_{s-1} \varepsilon_{t+1}$$

El valor esperado de  $e_{t+s}$  es cero y su varianza es:

$$V(e_{t+s}) = \Omega + \Psi'_1 \Omega \Psi_1 + \Psi'_2 \Omega \Psi_2 + \cdots + \Psi'_{s-1} \Omega \Psi_{s-1}$$

- Da una idea de la magnitud del error de predicción (es el error cuadrático medio ya que la esperanza es cero!).
- Notar que cuando  $s \rightarrow \infty$ ,  $V(e_{t+s}) \rightarrow V(Y_t)$ .

En la práctica: computar la descomposición para  $s = 1, 2, 3, \dots$

## Tests de hipotesis basicas en VAR

Supongamos que nos interesa evaluar si cierta restriccion sobre los parametros del VAR es correcta,  $H_0 : h(\theta) = 0$  versus  $h(\theta) \neq 0$  en donde  $\theta$  son todos los parametros del VAR y  $h$  es cualquier funcion continua.

Si  $\varepsilon_t \sim RBN(0, \Omega)$

$$(T - c) \left( \ln |\hat{\Omega}_r| - \ln |\hat{\Omega}_u| \right) \sim \chi^2(m)$$

asintoticamente bajo  $H_0$ , en donde  $\hat{\Omega}_r$  y  $\hat{\Omega}_u$  son las matrices de varianzas estimadas bajo el modelo restringido y sin restringir, respectivamente,  $m$  es el numero de restricciones,  $c$  es el numero de parametros estimados en cada ecuacion en el modelo sin restringir, y  $T$  es el numero de observaciones utilizables.

## Ejemplos

1. *Relevancia de rezagos:* Ejemplo, si en con series trimestrales los rezagos del 9 al 12 son relevantes. Modelo sin restringir: incluye los 12 rezagos, sin restringir: solo los primeros 8.
2. *Relevancia de variables:* todos los rezagos de una variable son irrelevantes en *todas* las ecuaciones.

## Causalidad de Granger

Consideremos la primera ecuacion de la representacion VAR(p) para  $x_t$  y  $y_t$ :

$$x_t = c + \sum_{i=1}^p \alpha_i x_{t-i} + \sum_{i=1}^p \beta_i y_{t-i}$$

$y_t$  no causa a  $x_t$  en el sentido de Granger si

$$\beta_1 = \beta_2 = \dots = \beta_p = 0$$

La hipotesis puede ser facilmente evaluada con un test  $F$  de significatividad conjunta (existen varias alternativas). En forma similar se puede definir ausencia de causalidad de Granger de  $x_t$ .

- Es uno de los resultados mas abusados en econometria.
- En realidad evalua si el pasado de  $y_t$  contribuye a predecir  $x_t$ . Es mas un test que tiene que ver con precedencia temporal y no con causalidad.

*Ejemplo 1: Forward-looking behavior y precios de acciones.*

*Ejemplo 2: Shocks en los precios de petroleo y recesiones.*

*Ejemplo 3: Efectos reales del dinero (Stock y Watson (1989), Christiano y Ljungqvist (1988)).*