

Vectores Auto-Regresivos (VAR)

Randall Romero Aguilar, PhD
randall.romero@ucr.ac.cr

EC4301 - Macroeconometría
I Semestre 2020

Última actualización: 15 de junio de 2020

UCR
UNIVERSIDAD DE COSTA RICA

**ESCUELA de
ECONOMÍA**
UNIVERSIDAD DE COSTA RICA

Tabla de contenidos

1. Representaciones alternativas de un VAR
2. La función de impulso respuesta
3. Estacionariedad
4. Estimación
5. Causalidad de Granger
6. Pronóstico y la descomposición de varianza de su error

Las tareas del macroeconometrista *

En general, los macroeconometristas hacemos cuatro cosas:

1. describimos y resumimos datos macroeconómicos
2. realizamos pronósticos macroeconómicos
3. cuantificamos lo que sabemos (y lo que no sabemos) acerca de la verdadera estructura de la macroeconomía
4. aconsejamos a quienes toman decisiones de política (y a veces tomamos esas decisiones).

- ▶ En los 1970s, estas cuatro tareas se realizaban usando una variedad de técnicas.
- ▶ Pero después del caos macroeconómico de los 1970s, ninguna de esas técnicas parecía apropiada.
- ▶ En 1980, Sims (1980, pp.15) afirma que
Debe ser factible estimar modelos macro de gran escala como formas reducidas sin restricciones, tratando todas las variable como endógenas.

- ▶ En 1980, Sims propone una nueva técnica macroeconométrica: el vector autorregresivo (VAR)
- ▶ Un VAR es un modelo lineal de n ecuaciones y n variables, en el cual cada variable es explicada por sus propios rezagos, más los valores actuales y rezagados de las otras $n - 1$ variables.
- ▶ Este sencillo esquema provee una manera sistemática de capturar la gran dinámica en series de tiempo múltiples, y las herramientas estadísticas que vienen con los VAR son fáciles de usar y de interpretar.

1. Representaciones alternativas de un VAR

Las series de tiempo

- ▶ En lo que sigue, asumimos que hay n series de tiempo.
- ▶ La observación t se denota y_t , corresponde a un vector columna con n datos (uno por cada serie)

	A	B	C	D	E	F	G
1	fecha	variable1	variable2	variable3			
2	1	-1.4	0.3	2			y_t
3	2	0.6	-1.0	4			1.5
4	3	-0.1	0.9	1			0.9
5	4	0.5	4.1	4			4
6	5	1.6	2.6	5			
7	6	1.5	0.9	4	y'_t		
8	7	-1.1	-0.4	5			
9	8	0.5	4.8	4	y'_{t+2}		
10							

Distintos “sabores” de VAR

- ▶ En general, hay tres variedades de VAR:
 1. VAR en forma reducida
 2. VAR recursivo
 3. VAR estructural
- ▶ Se distinguen por cómo se presentan las relaciones **contemporáneas** entre las variables.

El proceso VAR(p) en forma reducida

- ▶ El VAR reducido expresa cada variable como función lineal de los rezagos de todas las variables en el sistema y de un término de error sin correlación serial.
- ▶ Si hay p rezagos, el VAR reducido se representa

VAR(p)

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

donde Φ_j es la matriz $n \times n$ de coeficientes del rezago j .

- ▶ El término de error cumple:

$$\mathbb{E} \epsilon_t = 0 \quad \text{Var} \epsilon_t = \Omega \quad \text{Cov}[\epsilon_t, \epsilon_s] = 0 \text{ if } t \neq s$$

- ▶ No hay relación contemporánea entre las variables.

El proceso VAR(p) estructural

- ▶ El VAR estructural es similar al reducido, pero las variables tienen relación contemporánea.
- ▶ Para determinarla, se usa teoría económica.
- ▶ Se requiere de supuestos de identificación que permitan interpretar correlaciones como causalidad.
- ▶ Se representa como

$$\Gamma_0 y_t = d + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t$$

donde Γ_j es la matriz $n \times n$ de coeficientes del rezago j .

- ▶ El término de error cumple:

$$\mathbb{E} \varepsilon_t = 0 \quad \text{Cov}[\varepsilon_{it}, \varepsilon_{js}] = \begin{cases} \sigma_{ij}^2, & \text{si } t = s \\ 0, & \text{de lo contrario.} \end{cases}$$

El proceso VAR(p) en forma recursiva

- ▶ El VAR recursivo es un caso particular del VAR estructural, donde las variables tienen relación contemporánea de manera recursiva:
 - ▶ y_{1t} es “exógena”
 - ▶ y_{2t} depende sólo de y_{1t}
 - ▶ y_{3t} depende de y_{1t} y de y_{2t}
 - ▶ y_{nt} depende de todas las demás.
- ▶ Note que el ordenamiento de las variables es importante, habiendo $n!$ formas de ordenarlas. (Por ejemplo, en un VAR de cinco variables, hay $5! = 120$ ordenamientos).
- ▶ Al igual que el VAR estructural, se representa como

$$\Gamma_0 y_t = d + \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t$$

pero con la restricción de que Γ_0 es uni-triangular inferior.

Ejemplo 1:

Distintas versiones de un VAR

Forma reducida

$$\begin{aligned}y_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\m_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\z_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z\end{aligned}$$

Forma estructural

$$\begin{aligned}y_t + \beta_1 m_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\m_t + \beta_2 z_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\z_t + \beta_3 y_t + \beta_4 m_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z\end{aligned}$$

Forma recursiva

$$\begin{aligned}y_t &= y_0 + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \alpha_{13}z_{t-1} + u_t^y \\ \gamma_1 y_t + m_t &= m_0 + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \alpha_{23}z_{t-1} + u_t^m \\ \gamma_2 y_t + \gamma_3 m_t + z_t &= z_0 + \alpha_{31}y_{t-1} + \alpha_{32}m_{t-1} + \alpha_{33}z_{t-1} + u_t^z\end{aligned}$$

Pasando de un VAR estructural (o recursivo) a uno reducido

- ▶ Si la matriz Γ_0 es invertible (lo que está garantizado en el VAR recursivo), entonces se puede obtener el VAR reducido así:

$$\begin{aligned}\Gamma_0 y_t &= d + \Gamma_1 y_{t-1} + \cdots + \Gamma_p y_{t-p} + \varepsilon_t \\ y_t &= \Gamma_0^{-1} d + \Gamma_0^{-1} \Gamma_1 y_{t-1} + \cdots + \Gamma_0^{-1} \Gamma_p y_{t-p} + \Gamma_0^{-1} \varepsilon_t \\ &= c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t\end{aligned}$$

- ▶ Las matrices de covarianza cumplen

$$\Omega = \Gamma_0^{-1} \Sigma \Gamma_0'^{-1}$$

- Recuerde que

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \Rightarrow \Sigma^{1/2} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}$$

- Así

$$\Omega = \left[\Gamma_0^{-1} \Sigma^{1/2} \right] \left[\Gamma_0^{-1} \Sigma^{1/2} \right]'$$

- En el caso de un VAR recursivo, Γ_0 es triangular inferior, y por tanto $\Gamma_0^{-1} \Sigma^{1/2}$ también lo es.
- Así, hemos obtenido la descomposición de Cholesky de Ω .

Ejemplo 2:

Descomposición de Cholesky

VAR con series como desviación de la media

- ▶ Si el VAR es estacionario, su media $\mathbb{E} y_t \equiv \mu_t$ es constante
- ▶ Así,

$$\begin{array}{rcl} y_t = \Phi_1 y_{t-1} & + \cdots + \Phi_p y_{t-p} & + c + \epsilon_t \\ \mu = \Phi_1 \mu & + \cdots + \Phi_p \mu & + c + \mathbb{E} \epsilon_t \\ \hline y_t - \mu = \Phi_1 (y_{t-1} - \mu) & + \cdots + \Phi_p (y_{t-p} - \mu) & + \epsilon_t \end{array}$$

- ▶ La media del proceso es

$$\mu = (I - \Phi_1 - \Phi_2 \cdots - \Phi_p)^{-1} c$$

- ▶ Definimos $\hat{y} \equiv y_t - \mu$. El VAR(p) con series como desviación de la media es

$$\hat{y}_t = \Phi_1 \hat{y}_{t-1} + \cdots + \Phi_p \hat{y}_{t-p} + \epsilon_t$$

VAR(p) escrito como VAR(1)

$$\begin{bmatrix} \hat{y}_t \\ \hat{y}_{t-1} \\ \hat{y}_{t-2} \\ \vdots \\ \hat{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \Phi_3 & \dots & \Phi_{p-1} & \Phi_p \\ I & 0 & 0 & \dots & 0 & 0 \\ 0 & I & 0 & \dots & 0 & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & \dots & I & 0 \end{bmatrix} \begin{bmatrix} \hat{y}_{t-1} \\ \hat{y}_{t-2} \\ \hat{y}_{t-3} \\ \vdots \\ \hat{y}_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

\hat{Y}_t Φ \hat{Y}_{t-1} ξ_t

así

$$\hat{Y}_t = \Phi \hat{Y}_{t-1} + \xi_t$$

Ejemplo 3:

$\text{VAR}(2)$ a $\text{VAR}(1)$

El VAR(2)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \begin{bmatrix} x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{bmatrix}$$

se escribe como VAR(1)

$$\begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} .5 & .1 & 0 & 0 \\ .4 & .5 & .25 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{xt} \\ \epsilon_{yt} \\ 0 \\ 0 \end{bmatrix}$$

Dinámica de un proceso VAR(1)

- ▶ A menudo necesitamos iterar en la fórmula del AR(1) para analizar su dinámica.
- ▶ En esos casos, es útil observar que:

$$(1 + \Phi L + \Phi^2 L^2 + \dots + \Phi^s L^s) (I - \Phi L) = (I - \Phi^{s+1} L^{s+1})$$

- ▶ Así, si queremos expresar \hat{Y}_t en términos de \hat{Y}_{t-s-1}

$$(I - \Phi L) \hat{Y}_t = \xi_t$$

$$(I - \Phi^{s+1} L^{s+1}) \hat{Y}_t = (1 + \Phi L + \Phi^2 L^2 + \dots + \Phi^s L^s) \xi_t$$

$$\hat{Y}_t - \Phi^{s+1} \hat{Y}_{t-s-1} = \xi_t + \Phi \xi_{t-1} + \dots + \Phi^s \xi_{t-s}$$

Pasado Cuando queremos analizar \hat{Y}_t en función de shocks pasados, utilizamos

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \cdots + \Phi^s \xi_{t-s} + \Phi^{s+1} \hat{Y}_{t-s-1}$$

Futuro Para analizar el efecto de nuevos shocks sobre futuros valores de \hat{Y} , aplicamos L^{-s} a la última ecuación:

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \cdots + \Phi^s \xi_{t-s} + \Phi^{s+1} \hat{Y}_{t-s-1}$$

- Si los eigenvalores de Φ están en el círculo unitario el límite $s \rightarrow \infty$ converge a

$$\hat{Y}_t = \xi_t + \Phi \xi_{t-1} + \Phi^2 \xi_{t-2} + \dots$$

2. La función de impulso respuesta

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

- ▶ La función de **impulso-respuesta** mide la respuesta observada en la variable m -ésima s períodos después ($\hat{Y}_{t+s,m}$) de que se presenta un impulso en la k -ésima variable ($\xi_{t,k}$)
- ▶ Viene dada por

$$\frac{\partial \hat{Y}_{t+s,m}}{\partial \xi_{t,k}} = (\Phi^s)_{km}$$

es decir, por el elemento en la fila k , columna m , de la matriz Φ elevada al número de períodos s .

Interpretando la función de impulso-respuesta

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

- ▶ Suponga que
 - ▶ el sistema estaba en equilibrio en $t - 1$, es decir $\hat{Y}_{t-1} = 0$
 - ▶ hay un shock $v' = [v_1 \ \dots \ v_n]$ a las variables en t , $\xi_t = v$
 - ▶ el shock es transitorio: $0 = \xi_{t+1} = \xi_{t+2} = \dots$
- ▶ En este caso, la desviación del sistema respecto a su equilibrio s períodos después del shock es

$$\hat{Y}_{t+s} = \Phi^s v$$

$$\Omega = PP' = [\Gamma_0^{-1}\Sigma^{1/2}] [\Gamma_0^{-1}\Sigma^{1/2}]' = \Gamma_0^{-1}\Sigma\Gamma_0^{-1}$$

- ▶ En la práctica, estamos interesados en shocks ε_t a las ecuaciones estructurales, en vez de a las ecuaciones reducidas ϵ_t . Esto para tomar en cuenta los efectos contemporáneos del shock.
- ▶ Para calcular las respuestas, nos valemos de $\epsilon_t = \Gamma_0^{-1}\varepsilon_t$ y de la descomposición de Cholesky de la covarianza de los errores reducidos Ω :

Tamaño del impulso	Impulso estructural	Impulso reducido	Respuesta
unitarios	$\varepsilon_t = I$	$v = \Gamma_0^{-1}$	$\Phi^s \Gamma_0^{-1}$
1 st. dev.	$\varepsilon_t = \Sigma^{1/2}$	$v = \Gamma_0^{-1}\Sigma^{1/2}$	$\Phi^s \Gamma_0^{-1}\Sigma^{1/2}$

Ejemplo 4:

Impulso respuesta y descomposición
de Cholesky

Siguiendo con el ejemplo 1, si $y'_t = [m_t \ r_t \ k_t]$ y la matriz de covarianza reducida es

$$\begin{aligned}\Omega &= \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 4.25 & 2.5 \\ -1 & 2.5 & 12.25 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 2 & 0 \\ -1 & 1.5 & 3 \end{bmatrix}}_{P'} \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 2 & 1.5 \\ 0 & 0 & 3 \end{bmatrix}}_{P'} \\ &= \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -1 & 0.75 & 1 \end{bmatrix}}_{\Gamma_0^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix}}_{\Gamma_0'^{-1}}\end{aligned}$$

La respuesta del sistema a un shock en k_t se calcula a partir de ...

- ▶ $v' = [0 \ 2 \ 1.5]$ si el shock es de una desviación estándar.
- ▶ $v' = [0 \ 1 \ 0.75]$ si el shock es unitario.

3. Estacionariedad

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^s \xi_t + \Phi^{s+1} \hat{Y}_{t-1}$$

- ▶ Recuerde que la respuesta del VAR s períodos después a un impulso ξ_t es $\Phi^s \xi_t$.
- ▶ Si el VAR es estacionario, toda respuesta a cualquier impulso en t debe ser transitoria: $\lim_{s \rightarrow \infty} \Phi^s \xi_t = 0$. De lo contrario, la media μ del proceso no sería constante.
- ▶ Es decir, el VAR es estacionario si y sólo si

$$\lim_{s \rightarrow \infty} \Phi^s = 0$$

Nota:

Descomposición espectral de una
matrix

Descomposición espectral de una matrix

Si los eigenvectores de la matriz cuadrada A son linealmente independientes, entonces

Descomposición espectral

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

donde Λ es la matriz diagonal formada por los eigenvalores de A :

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

y las columnas de C son los correspondientes eigenvectores de A .

Si A tiene la descomposición espectral $A = C\Lambda C^{-1}$ es fácil calcular su t -ésima potencia:

$$A^t = C\Lambda^t C^{-1}$$

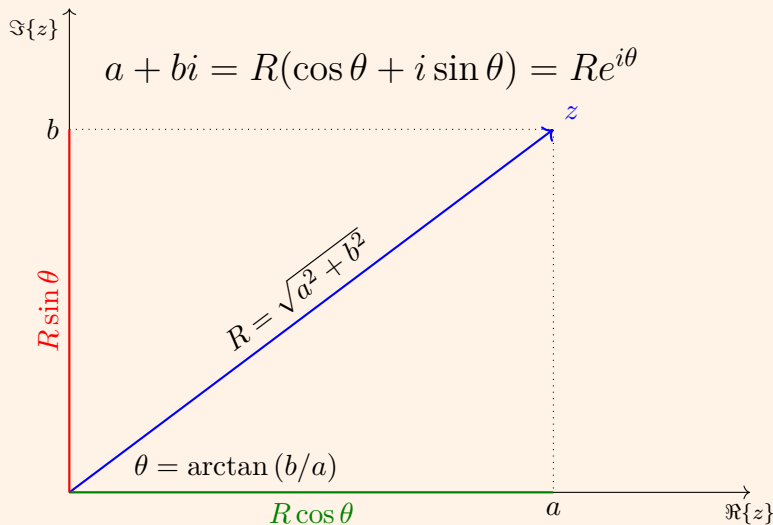
ya que

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & 0 & \dots & 0 \\ 0 & \lambda_2^t & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \lambda_n^t \end{bmatrix}$$

Nota:

Números complejos

Representación de números complejos



Multiplicación de números complejos

- ▶ Si $z = Re^{i\theta}$ y $w = Se^{i\varphi}$, entonces su producto es

$$zw = RSe^{i(\theta+\varphi)}$$

- ▶ Así, si elevamos z a la n -ésima potencia:

$$z^n = \left(Re^{i\theta}\right)^n = R^n e^{in\theta}$$

- ▶ Es decir

$$\lim_{n \rightarrow \infty} z^n = 0 \Leftrightarrow |R| < 1$$

Ejemplo de potencia de números complejo

Condiciones para la estacionariedad

- ▶ Y es estacionario si y solo si todos los eigenvalores de Φ están dentro del círculo unitario.
- ▶ Los eigenvalores λ de Φ satisfacen:

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

Ejemplo 5:

Estacionariedad de un VAR(2)

Para el VAR(2) del ejemplo 1:

$$\begin{aligned} 0 &= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \lambda^2 - \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \lambda - \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} \lambda^2 - .5\lambda & -.1\lambda \\ -.4\lambda - .25 & \lambda^2 - .5\lambda \end{bmatrix} \right| \\ &= (\lambda^2 - .5\lambda) (\lambda^2 - .5\lambda) - (-.4\lambda - .25) (-.1\lambda) \\ &= \lambda^4 - \lambda^3 + 0.21\lambda^2 - 0.025\lambda \end{aligned}$$

Las raíces de este polinomio son $\lambda_1 = 0$, $\lambda_2 = 0.7693$, $\lambda_3 = 0.1154 + 0.1385i$, y $\lambda_4 = 0.1154 - 0.1385i$, todas ellas dentro del círculo unitario. Por lo tanto, el VAR es estacionario.

Ejemplo 6:
Dinámica de un VAR

- ▶ Hasta el momento hemos estudiado
 - ▶ cómo determinar si un VAR es estacionario
 - ▶ la función de impulso respuesta
- ▶ En el cuaderno de Jupyter 02 Simulacion de un VAR.ipynb se introduce un ejemplo para analizar un VAR.
- ▶ En el cuaderno de Jupyter 03 VAR(1) clase.ipynb se presentan 5 modelos VAR reducidos. Acá definimos una **clase** para representar un VAR.

4. Estimación

Especificación de un VAR

- ▶ Para especificar un VAR, hay dos cosas que escoger:
 - ▶ Las n series que conforman el VAR
 - ▶ El número p de rezagos a ser incluidos en el modelo
- ▶ En contraste con la metodología de Box-Jenkins, el modelo VAR no es parsimonioso \Rightarrow un VAR está sobreparametrizado:
 - ▶ hay $n + pn^2$ parámetros.
 - ▶ muchos de sus coeficientes no son significativos
 - ▶ regresores posiblemente son altamente colineales, estadísticos t no confiables.
- ▶ No obstante, en un VAR los coeficientes individuales no son de interés. Interesa determinar la dinámica del modelo (impulso respuesta) y la causalidad (Granger).

- ▶ El VAR reducido es un sistema de ecuaciones SUR, donde todas las ecuaciones tienen los mismos regresores.
- ▶ Por lo tanto, la estimación OLS ecuación por ecuación es insesgada, eficiente y consistente.
- ▶ La matriz de covarianza de los errores se estima a partir de los residuos

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

Si las series son $I(1)$, ¿hay que diferenciarlas para estimar el VAR?
Hay un debate respecto:

NO: Algunos aconsejan no diferenciar las series, porque se pierde información, alegando que la meta del VAR es determinar relaciones entre variables, no los parámetros mismos.

SÍ: Otros advierten que si las series son integradas, el VAR en niveles no es consistente con el proceso generador de datos. En este caso se aconseja estudiar la cointegración de las series \Rightarrow VECM

Ejemplo 7:

Propiedades de la estimación OLS de un VAR



VAR insesgado.ipynb

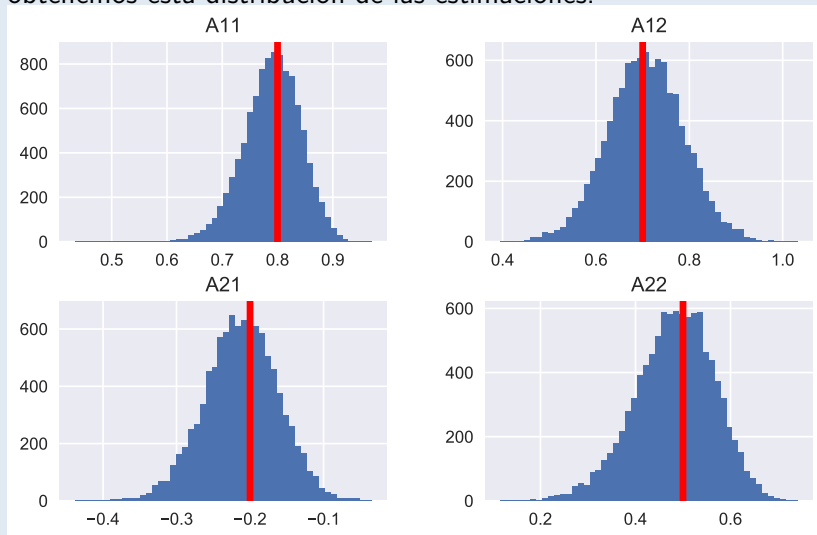
- ▶ En el cuaderno de Jupyter [VAR insesgado.ipynb](#) se presenta un ejercicio de simulación para ilustrar que el estimador OLS de un VAR es insesgado y consistente.
- ▶ Se simula el VAR(1)

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 0.8 & 0.7 \\ -0.2 & 0.5 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{x,t} \\ \epsilon_{y,t} \end{pmatrix}$$

$\underbrace{\hspace{1.5cm}}_{Y_t} \qquad \underbrace{\hspace{1.5cm}}_{\Phi} \qquad \underbrace{\hspace{1.5cm}}_{Y_{t-1}} \qquad \underbrace{\hspace{1.5cm}}_{\epsilon_t}$

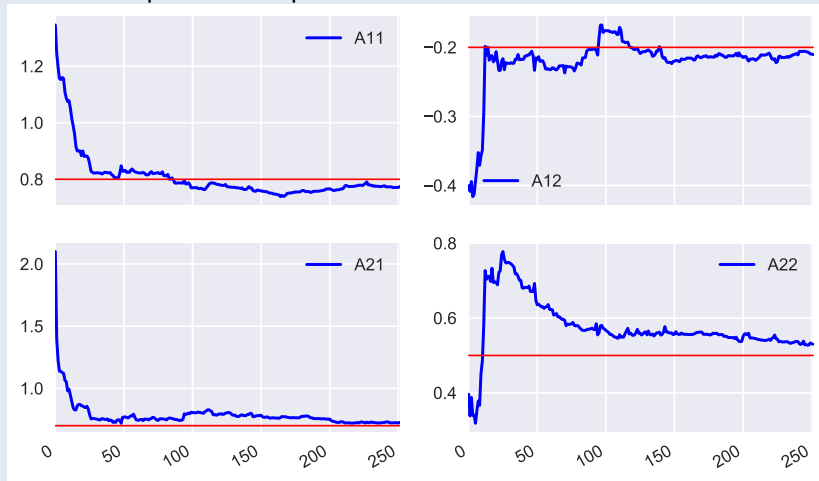
$$\Omega = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}$$

Con un experimento de Monte Carlo de 10.000 repeticiones obtenemos esta distribución de las estimaciones.



Consistente

Simulando una sola muestra, se estima el VAR agregando una observación a la vez. Las estimaciones convergen hacia los verdaderos parámetros poblacionales.



Escogiendo el óptimo número de rezagos del VAR

- ▶ La selección del número de rezagos p es crítica.
 - ▶ Si p es demasiado pequeño, el modelo está mal especificado
 - ▶ Si p es demasiado grande, se desperdician grados de libertad y se introduce colinealidad.
- ▶ Una opción: dejar que p difiera para cada ecuación y variable
⇒ el VAR no podría estimarse con OLS.
- ▶ Por tanto, se determina un sólo p para todas las variables y ecuaciones, usando dos enfoques distintos
 - ▶ Descartando los últimos rezagos si no son significativos
 - ▶ Utilizando un criterio de selección que contraponga el ajuste a los grados de libertad.

Escogiendo el p óptimo con pruebas de significancia

- ▶ Suponga que queremos decidir entre p y q número de rezagos, con $q < p$.
- ▶ Obtenemos $\hat{\Omega}$ para ambos modelos, usando la misma muestra.
- ▶ Comprobamos la significancia conjunta de los rezagos $q + 1, \dots, p$ con la prueba de razón de verosimilitud

$$(T - 1 - pn) \left(\ln \left| \hat{\Omega}_q \right| - \ln \left| \hat{\Omega}_p \right| \right) \overset{asym}{\sim} \chi^2(n^2(p - q))$$

- ▶ Si no se rechaza la hipótesis nula, puede escogerse q rezagos.
- ▶ Limitaciones:
 - ▶ Resultado asintótico, pero muestras usualmente pequeñas
 - ▶ Resultado sensible a escogencia de p y q .

Escogiendo el p óptimo con criterios de selección

- ▶ En la práctica, el número óptimo de rezagos se escoge minimizando el valor de AIC o de SBC:

$$\text{AIC} = T \ln |\hat{\Omega}| + 2N$$

$$\text{SBC} = T \ln |\hat{\Omega}| + N \ln T$$

donde $N = n^2p + n$ es el número total de parámetros estimados

- ▶ Rezagos adicionales reducen $\hat{\Omega}$, pero incrementando N

Usando un VAR para describir los datos

- ▶ La práctica usual en el análisis VAR es reportar
 - ▶ pruebas de causalidad de Granger
 - ▶ funciones de impulso respuesta
 - ▶ descomposición de varianza del error de pronóstico
- ▶ Usualmente no se reportan los coeficientes estimados ni el R^2 , porque no son tan informativos.

5. Causalidad de Granger

- ▶ Las pruebas de causalidad de Granger determinan si los valores rezagados de una variable ayudan a predecir otra variable.
- ▶ Si los rezagos de A no ayudan a predecir B , entonces coeficientes de los rezagos de A en la ecuación de B son ceros.
- ▶ Si alguna variable no es causada en el sentido de Granger por ninguna otra, **no implica que la variable sea exógena**. Bien podría haber causalidad contemporánea.

Ejemplo 8:

Causalidad de Granger

En el modelo VAR(2) con variables y_t, m_t :

$$\begin{aligned}y_t &= c^y + \alpha_{11}y_{t-1} + \alpha_{12}m_{t-1} + \beta_{11}y_{t-2} + \beta_{12}m_{t-2} + u_t^y \\m_t &= c^m + \alpha_{21}y_{t-1} + \alpha_{22}m_{t-1} + \beta_{21}y_{t-2} + \beta_{22}m_{t-2} + u_t^m\end{aligned}$$

las pruebas de causalidad de Granger son:

Hipótesis nula	Restricciones
m no causa y	$\alpha_{12} = \beta_{12} = 0$
y no causa m	$\alpha_{21} = \beta_{21} = 0$

Si se rechaza, por ejemplo, la primera hipótesis, decimos que m causa y en el sentido de Granger.

6. Pronóstico y la descomposición de varianza de su error

$$\hat{Y}_{t+s} = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^{s-1} \xi_{t+1} + \Phi^s \hat{Y}_t$$

- ▶ Suponga que Φ ha sido estimado con datos hasta $t = T$.
- ▶ El mejor pronóstico del sistema s períodos adelante es

$$\mathbb{E} \left[\hat{Y}_{T+s} \mid \hat{Y}_T \right] = \Phi^s \hat{Y}_T$$

- ▶ El error de pronóstico es

$$\hat{Y}_{T+s} - \mathbb{E} \left[\hat{Y}_{T+s} \mid \hat{Y}_T \right] = \xi_{t+s} + \Phi \xi_{t+s-1} + \cdots + \Phi^{s-1} \xi_{t+1}$$

- ▶ y su varianza (MSPE) es

$$\text{Var} \left[\hat{Y}_{T+s} \mid \hat{Y}_T \right] = \Omega + \Phi \Omega \Phi' + \cdots + \Phi^{s-1} \Omega \Phi'^{s-1}$$

Pronósticos de largo plazo con VAR

$$\mathbb{E} \left[\hat{Y}_{T+s} \mid \hat{Y}_T \right] = \Phi^s \hat{Y}_T$$

► Recuerde que:

► $\hat{y} \equiv y_t - \mu$

► si el VAR es estacionario, $\lim_{s \rightarrow \infty} \Phi^s = 0$.

► Entonces el pronóstico de largo plazo

$$\lim_{s \rightarrow \infty} \mathbb{E} \left[Y_{T+s} - \mu \mid \hat{Y}_T \right] = \lim_{s \rightarrow \infty} \Phi^s \hat{Y}_T = 0$$

► es decir

$$\lim_{s \rightarrow \infty} \mathbb{E} \left[y_{T+s} \mid \hat{Y}_T \right] = \mu$$

en el largo plazo el VAR volverá a su equilibrio estacionario.

Descomposición de la varianza del shock reducido

- ▶ Recuerde que los errores reducidos están relacionados con los estructurales por $\epsilon_t = \Gamma_0^{-1} \varepsilon_t$.
- ▶ Sea $\Gamma_0^{-1} \equiv A = [a_1 \ \dots \ a_n]$, con a_i la i -ésima columna de A .
- ▶ Entonces

$$\begin{aligned}\epsilon_t &= a_1 \varepsilon_{1t} + a_2 \varepsilon_{2t} + \dots + a_n \varepsilon_{nt} \\ \Omega &= \mathbb{E}[\epsilon_t \epsilon_t'] = \sigma_1^2 a_1 a_1' + \sigma_2^2 a_2 a_2' + \dots + \sigma_n^2 a_n a_n' \\ &= \sum_{j=1}^n \sigma_j^2 a_j a_j'\end{aligned}$$

Ejemplo 9:

Descomposición de Ω

Tomando Ω del ejemplo 1

$$\Omega = \begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 4.25 & 2.5 \\ -1 & 2.5 & 12.25 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ -1 & 0.75 & 1 \end{bmatrix}}_{\Gamma_0^{-1}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix}}_{\Gamma_0'^{-1}}$$

Entonces $a'_1 = [1 \quad 0.5 \quad -1]$, $a'_2 = [0 \quad 1 \quad 0.75]$, $a'_3 = [0 \quad 0 \quad 1]$ y

$$\begin{aligned} \Omega &= 1 \begin{bmatrix} 1 \\ 0.5 \\ -1 \end{bmatrix} [1 \quad 0.5 \quad -1] + 4 \begin{bmatrix} 0 \\ 1 \\ 0.75 \end{bmatrix} [0 \quad 1 \quad 0.75] + 9 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [0 \quad 0 \quad 1] \\ &= \underbrace{1}_{\sigma_1^2} \underbrace{\begin{bmatrix} 1 & 0.5 & -1 \\ 0.5 & 0.25 & -0.5 \\ -1 & -0.5 & 1 \end{bmatrix}}_{a_1 a'_1} + \underbrace{4}_{\sigma_2^2} \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0.75 \\ 0 & 0.75 & 0.5625 \end{bmatrix}}_{a_2 a'_2} + \underbrace{9}_{\sigma_3^2} \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{a_3 a'_3} \end{aligned}$$

Observe que esta descomposición depende del ordenamiento de las variables.

Descomposición de la varianza del pronóstico

- ▶ Sustituyendo $\Omega = \sum_{j=1}^n \sigma_j^2 a_j a_j'$ en $\text{Var} [\hat{Y}_{T+s} | \hat{Y}_T]$ tenemos

$$\text{Var} [\hat{Y}_{T+s} | \hat{Y}_T] = \sum_{j=1}^n \sigma_j^2 \left\{ a_j a_j' + \Phi a_j a_j' \Phi' + \cdots + \Phi^{s-1} a_j a_j' \Phi'^{s-1} \right\}$$

- ▶ Con esta expresión, podemos cuantificar la contribución del j -ésimo shock estructural al error cuadrático medio del pronóstico s -períodos adelante.
- ▶ Observe que esto asume que conocemos los parámetros del modelo! **No toma en cuenta los errores de estimación.**

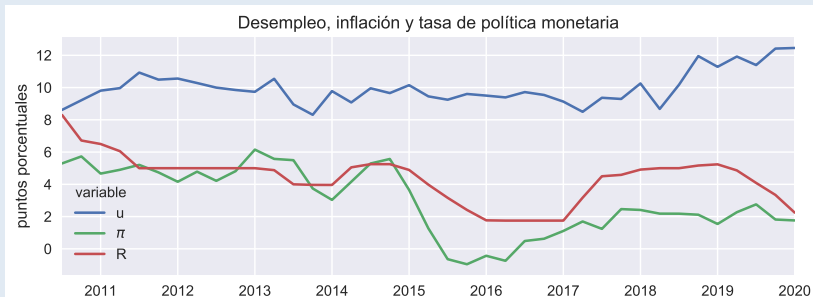
Ejemplo 10:

Un modelo VAR para la política
monetaria de Costa Rica

- ▶ Supongamos que queremos ver el papel que juega la tasa de política monetaria R del BCCR sobre el desempleo u y la inflación π en nuestro país.
- ▶ Para ello estimaremos un VAR:

$$\begin{bmatrix} u_t \\ \pi_t \\ R_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \sum_{i=1}^p \underset{3 \times 3}{\Phi_i} \begin{bmatrix} u_{t-i} \\ \pi_{t-i} \\ R_{t-i} \end{bmatrix} + \begin{bmatrix} \epsilon_t^u \\ \epsilon_t^\pi \\ \epsilon_t^R \end{bmatrix}$$

- ▶ Contamos con una muestra de 39 observaciones trimestrales, de 2010-III a 2020-I.



- El primer paso es escoger el número de rezagos p del VAR:

	AIC	BIC	FPE	HQIC
0	1.227	1.360	3.411	1.273
1	-1.779	-1.246*	0.1693	-1.595
2	-2.121*	-1.188	0.1219*	-1.799*
3	-2.007	-0.6741	0.1411	-1.547
4	-2.002	-0.2689	0.1511	-1.404

- En este ejemplo escogeremos un solo rezago, en parte porque tenemos una muestra muy pequeña.

Results for equation u

	coefficient	std. error	t-stat	prob
const	2.305139	1.265376	1.822	0.069
L1.u	0.708084	0.125437	5.645	0.000
L1.\$\pi\$	-0.127880	0.087455	-1.462	0.144
L1.R	0.243191	0.124252	1.957	0.050

Results for equation \$\pi\$

	coefficient	std. error	t-stat	prob
const	-0.062525	1.588872	-0.039	0.969
L1.u	-0.003928	0.157505	-0.025	0.980
L1.\$\pi\$	0.827925	0.109813	7.539	0.000
L1.R	0.119255	0.156017	0.764	0.445

Results for equation R

	coefficient	std. error	t-stat	prob
const	2.282890	0.964744	2.366	0.018
L1.u	-0.164187	0.095635	-1.717	0.086
L1.\$\pi\$	0.091391	0.066677	1.371	0.170
L1.R	0.753419	0.094732	7.953	0.000

Correlation matrix of residuals

	u	\$\pi\$	R
u	1.000000	-0.031396	-0.003364
\$\pi\$	-0.031396	1.000000	0.307269
R	-0.003364	0.307269	1.000000

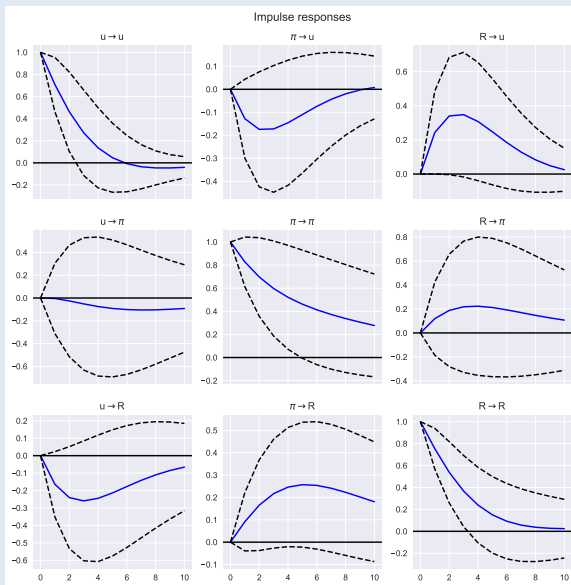
- ▶ Al parecer, ninguna variable del sistema causa a otra en el sentido de Granger.

Dependiente Explicativa	u	π	R
u	0.000	0.980	0.089
π	0.147	0.000	0.173
R	0.053	0.446	0.000

Valores p de hipótesis de causalidad.

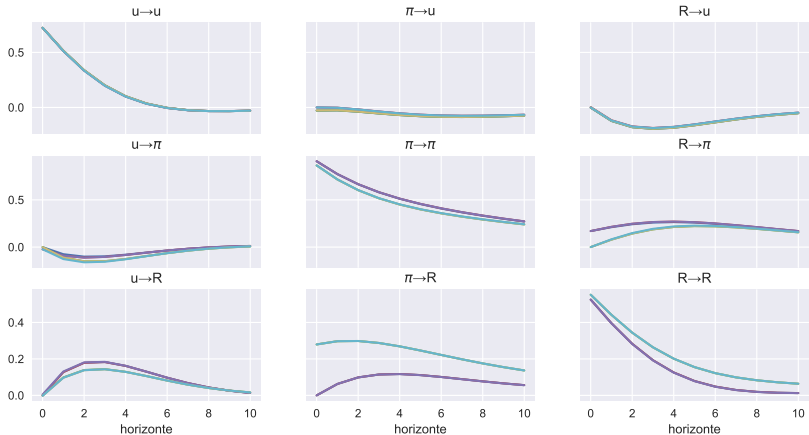
- ▶ Esto no necesariamente implica que no haya relación de causalidad entre las variables: podría haber causalidad contemporánea.

► Funciones de impulso respuesta, con impulsos unitarios.

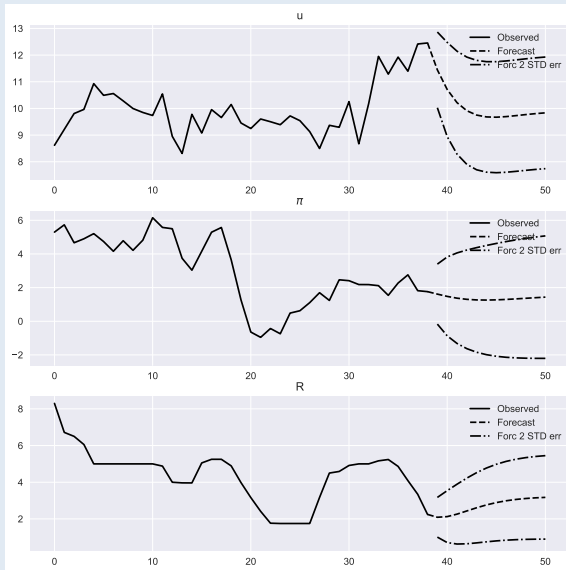


- Funciones de impulso respuesta, con impulsos ortogonales.
- El resultado depende del ordenamiento de las variables en el sistema (Choleski).

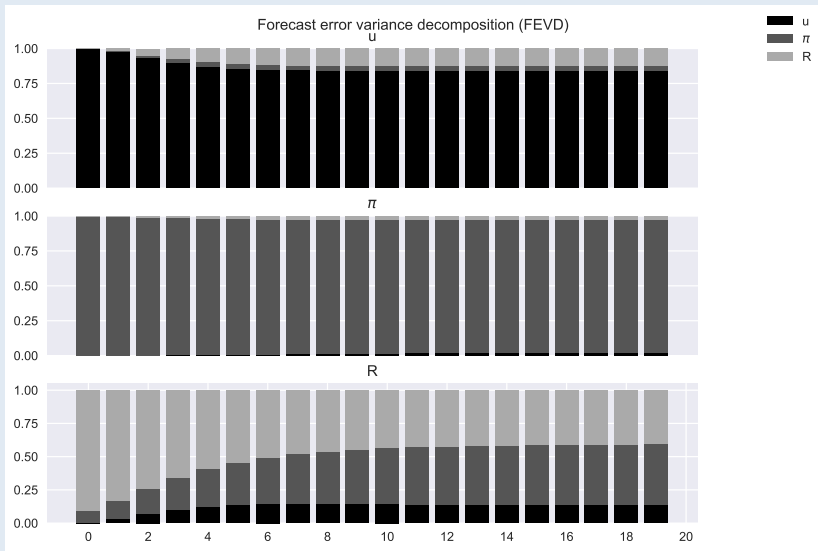
Impulso respuesta, todos los posibles ordenamientos de Choleski







- En un VAR estacionario, los pronósticos siempre convergen a la media de largo plazo de cada variable.



► Descomposición de la varianza de pronóstico.



-  Enders, Walter (2014). *Applied Econometric Time Series*. 4ª ed. Wiley. ISBN: 978-1-118-80856-6.
-  Hamilton, James M. (1994). *Time Series Analysis*. Princeton University Press. ISBN: 0-691-04289-6.
-  Sims, Christopher A. (ene. de 1980). "Macroeconomics and Reality". En: *Econometrica* 48.1.
-  Stock, James H. y Mark W. Watson (2001). "Vector Autoregressions". En: *Journal of Economic Perspectives* 15.4 (Fall), págs. 101-115.