# Modelos de vectores autorregresivos VAR

Gabriel V. Montes-Rojas

# Modelos de vectores autorregresivos (VAR) de 2 variables

Consideremos 2 series, $\{y_t\}$ y $\{z_t\}$ que están mutuamente relacionadas. Un modelo VAR consiste en la modelación conjunta de estas series

$$y_t = \gamma_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + w_{yt},$$
$$z_t = \gamma_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + w_{zt},$$

donde se asume:

- $y_t$ y $z_t$ son estacionarias;
- $w_{yt} \sim IID\ (0, \sigma_y^2)$ y $w_{zt} \sim IID\ (0, \sigma_z^2)$ son ruido blanco (en particular, no están autocorrelacionados). Estos se definen como "shocks" que son cambios exógenos con sentido económico. Notar que estos shocks son las contribuciones no predecibles en cada periodo de tiempo al sistema bivariado.

Esto constituye un modelo VAR bivariado de orden 1, VAR(1). Éste es un modelo VAR en forma **estructural** o **primitiva**.

# VAR - ¿Qué es un shock?
## Ramey (2016)

En los modelos macroeconómicos los shocks juegan un rol central para las evaluaciones empíricas. Un listado de lo que un shock debe satisfacer se da a continuación:

- (1) they should be exogenous with respect to the other current and lagged endogenous variables in the model;
- (2) they should be uncorrelated with other exogenous shocks; otherwise, we cannot identify the unique causal effects of one exogenous shock relative to another;
- and (3) they should represent either unanticipated movements in exogenous variables or news about future movements in exogenous variables.

# Modelos de vectores autorregresivos (VAR) de 2 variables

Los coeficientes se interpretan como en los modelos de regresión:

- $-b_{12}$ es el efecto contemporáneo de $z_t$ en $y_t$; $-b_{21}$ es el efecto contemporáneo de $y_t$ en $z_t$;
- $\gamma_{11}$ es el efecto de $y_{t-1}$ en $y_t$; $\gamma_{12}$ es el efecto de $z_{t-1}$ en $z_t$; etc.
- Cada ecuación no puede estimarse por MCO. ¿Por qué? Sesgo de ecuaciones simultáneas. Probar.

# Modelos de vectores autorregresivos (VAR) de 2 variables

El modelo estructural se puede reescribir como

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} w_{yt} \\ w_{zt} \end{bmatrix}$$

$$\boldsymbol{B}_0 \boldsymbol{x}_t = \boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_1 \boldsymbol{x}_{t-1} + \boldsymbol{w}_t$$

donde

$$\boldsymbol{B}_0 = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, \boldsymbol{x}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \boldsymbol{\Gamma}_0 = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \end{bmatrix},$$

$$\boldsymbol{\Gamma}_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \boldsymbol{w}_t = \begin{bmatrix} w_{yt} \\ w_{zt} \end{bmatrix}$$

Premultiplicando por $\boldsymbol{B}_0^{-1}$,

$$\boldsymbol{x}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1 \boldsymbol{x}_{t-1} + \boldsymbol{u}_t,$$

donde $\boldsymbol{A}_0 = \boldsymbol{B}_0^{-1}\boldsymbol{\Gamma}_0$, $\boldsymbol{A}_1 = \boldsymbol{B}_0^{-1}\Gamma_1$, y $\boldsymbol{u}_t = \boldsymbol{B}_0^{-1}\boldsymbol{w}_t$. Esto es un modelo VAR en forma **estándar** o **reducida**. Los parámetros $(\boldsymbol{A}_0, \boldsymbol{A}_1)$ **sí** se pueden estimar por MCO. ¿Por qué?

# Modelos de vectores autorregresivos (VAR) de 2 variables

$$u_{1t} = (w_{yt} - b_{12} w_{zt})/(1 - b_{12} b_{21})$$
$$u_{2t} = (w_{zt} - b_{21} w_{yt})/(1 - b_{12} b_{21})$$

Notar que $u_{1t}$ y $u_{2t}$ tienen media 0 y varianza constante, independiente del tiempo,

$$Var(u_{1t}) = (\sigma_{yt}^2 + b_{12}^2 \sigma_{zt}^2)/(1 - b_{12} b_{21})^2$$
$$Var(u_{2t}) = (\sigma_{zt}^2 + b_{21}^2 \sigma_{yt}^2)/(1 - b_{12} b_{21})^2$$

*Ejercicio: (i) Probar que las autocorrelaciones son 0. (ii) Encontrar la covarianza entre $u_{1t}$ y $u_{2t}$.*

Definir

$$\Sigma_u = \begin{bmatrix} Var(u_{1t}) & Cov(u_{1t}, u_{2t}) \\ Cov(u_{1t}, u_{2t}) & Var(u_{2t}) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

# VAR - Estabilidad y estacionariedad de 2 variables

- Reemplazando $n$ iteraciones llegamos a

$$\boldsymbol{x}_t = (\boldsymbol{I} + \boldsymbol{A}_1 + \boldsymbol{A}_1^2 + ... + \boldsymbol{A}_1^n)\boldsymbol{A}_0 + \sum_{i=0}^{n} \boldsymbol{A}_1^i u_{t-i} + \boldsymbol{A}_{n+1}\boldsymbol{x}_{t-n-1}$$

  Para que el modelo sea estable necesitamos que $lim_{n\to\infty}(\boldsymbol{I} + \boldsymbol{A}_1 + \boldsymbol{A}_1^2 + ... + \boldsymbol{A}_1^n) = (\boldsymbol{I} - \boldsymbol{A}_1)^{-1}$, es decir que la serie converja o que $(\boldsymbol{I} - \boldsymbol{A}_1)$ sea no singular.

- Para modelos bivariados la estabilidad depende de lo que pasa conjuntamente en las dos series. La condición de estabilidad es que las raíces del determinante de $(\boldsymbol{I} - \boldsymbol{A}_1)$ que lo podemos escribir como $(1 - a_{11}r)(1 - a_{22}r) - (a_{12}a_{21}r^2)$ estén por **fuera** del círculo unitario. Sean $r_1$ y $r_2$ la raíces del polinomio (reales o complejas), entonces "fuera del círculo unitario" significa $\sqrt{r_1^2 + r_2^2} > 1$.

- Otra forma de analizar la estabilidad es en términos de autovalores. Si tenemos $\boldsymbol{A}_1 \cdot \boldsymbol{x} = \lambda \boldsymbol{x}$ con $\boldsymbol{x} \neq \boldsymbol{0}$ llamamos a $\lambda$ un autovalor (eigenvalue) de $\boldsymbol{A}_1$. Estos pueden ser números reales o complejos. La condición de estabilidad es que los autovalores de $\boldsymbol{A}_1$ sean menores en valor absoluto a 1 o también **dentro** del círculo unitario.

# VAR - Estabilidad y estacionariedad de 2 variables

- Definamos $\boldsymbol{\mu} = [\bar{y}\ \bar{z}]' = (\boldsymbol{I} - \boldsymbol{A}_1)^{-1}\boldsymbol{A}_0$ donde $\bar{y} = [a_{10}(1 - a_{22}) + a_{12}a_{20}]/\Delta$, $\bar{z} = [a_{20}(1 - a_{11}) + a_{21}a_{10}]/\Delta$, $\Delta = (1 - a_{11})(1 - a_{22}) - a_{12}a_{21}$. Entonces,

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{A}_1^i \boldsymbol{u}_{t-i}$$

  tal que $E[\boldsymbol{x}_t] = \boldsymbol{\mu}$ y $Var[\boldsymbol{x}_t] = (\boldsymbol{I} - A_1^2)^{-1}\boldsymbol{\Sigma}_u$.
- Modelos VAR estables son estacionarios.

# VAR - Estabilidad y estacionariedad de 2 variables

- Otra forma de ver la estabilidad es a través de las siguientes ecuaciones:

$$(1 - a_{11}L)y_t = a_{10} + a_{12}Lz_t + u_{1t}$$

$$(1 - a_{22}L)z_t = a_{20} + a_{21}Lz_t + u_{2t}$$

Entonces, resolviendo

$$y_t = \frac{a_{10}(1 - a_{22}) + a_{12}a_{20} + (1 - a_{22}L)u_{1t} + a_{12}u_{2t}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

$$z_t = \frac{a_{20}(1 - a_{11}) + a_{21}a_{10} + (1 - a_{11}L)u_{2t} + a_{21}u_{1t}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

Notar que ambas series tienen la misma ecuación característica, con lo que se requiere que las raíces del polinomio $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$ estén fuera del círculo unitario.

# VAR - Modelos para $K$ variables, $p$ rezagos

El modelo anterior puede ser generalizado para $K$ variables y $p$ rezagos:

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{A}_1 x_{t-1} + \boldsymbol{A}_2 x_{t-2} + \ldots + \boldsymbol{A}_p \boldsymbol{x}_{t-p} + \boldsymbol{u}_t$$

donde

- $\boldsymbol{x}_t$ es un vector $(K \times 1)$;
- $\boldsymbol{\mu}$ es un vector $(K \times 1)$ de interceptos;
- $\boldsymbol{A}_i$, $i = 1, 2, \ldots, p$ son matrices de $(K \times K)$ de coeficientes;
- $\boldsymbol{u}_t$ es un vector $(K \times 1)$ de errores.

Hay entonces $K + Kp$ parámetros a estimar.
Se puede estimar cada ecuación por separado usando MCO.

# VAR - Modelos para $K$ variables, $p$ rezagos

Usando el operador de rezagos $L\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$, $L^2\boldsymbol{x}_t = \boldsymbol{x}_{t-2}$, etc. podemos definir

$$\boldsymbol{A}(L)(\boldsymbol{x}_t - \boldsymbol{\mu}) = \boldsymbol{u}_t,$$

donde $A(L) = \boldsymbol{I}_K - \boldsymbol{A}_1 L - \boldsymbol{A}_2 L^2 - ... - \boldsymbol{A}_p L^p$.
Otra forma de verlo es

$$\boldsymbol{A}(L)\boldsymbol{x}_t = \boldsymbol{A}(L)\boldsymbol{\mu} + \boldsymbol{u}_t.$$

La estabilidad del modelo VAR depende de las raíces del polinomio del determinante $det(\boldsymbol{A}(z)) = det(\boldsymbol{I}_K - \boldsymbol{A}_1 r - \boldsymbol{A}_2 r^2 - ... - \boldsymbol{A}_p r^p)$ que tienen que estar por fuera del círculo unitario.

# VAR - Modelos para $K$ variables, $p$ rezagos

El modelo anterior para $K$ variables y $p$ rezagos en forma estructural:

$$\boldsymbol{B}_0 \boldsymbol{x}_t = \boldsymbol{B}_1 \boldsymbol{x}_{t-1} + \boldsymbol{B}_2 \boldsymbol{x}_{t-2} + \ldots + \boldsymbol{B}_p \boldsymbol{x}_{t-p} + \boldsymbol{w}_t$$

donde

- $\boldsymbol{x}_t$ es un vector $(K \times 1)$;
- $\boldsymbol{B}_0$ es una matriz $(K \times K)$;
- $\boldsymbol{B}_i$ es una matriz $(K \times K)$ de coeficientes, $i = 1, 2, \ldots, p$;
- $\boldsymbol{w}_t$ es un vector $(K \times 1)$ de errores estructurales.

$$\boldsymbol{x}_t = \underbrace{\boldsymbol{B}_0^{-1} \boldsymbol{B}_1}_{\boldsymbol{A}_1} \boldsymbol{x}_{t-1} + \underbrace{\boldsymbol{B}_0^{-1} B_2}_{\boldsymbol{A}_2} \boldsymbol{x}_{t-2} + \ldots + \underbrace{\boldsymbol{B}_0^{-1} \boldsymbol{B}_p}_{\boldsymbol{A}_p} \boldsymbol{x}_{t-p} + \underbrace{\boldsymbol{B}_0^{-1} \boldsymbol{w}_t}_{\boldsymbol{u}_t}$$

Notar que los errores en forma reducida cumplen $\boldsymbol{u}_t = \boldsymbol{B}_0^{-1} \boldsymbol{w}_t$.
Podemos normalizar $E(\boldsymbol{w}_t \boldsymbol{w}_t') = \Sigma_w = \boldsymbol{I}_K$ sin perdida de generalidad tal que $E(\boldsymbol{u}_t \boldsymbol{u}_t') = \Sigma_u = \boldsymbol{B}_0^{-1} \boldsymbol{B}_0^{-1'}$.

# VAR - Estimadores MCO

- Dada una muestra de tamaño $T$, definamos

$$\boldsymbol{x}_t = [\boldsymbol{\mu}, \boldsymbol{A}_1, ..., \boldsymbol{A}_p]\boldsymbol{Z}_{t-1} + \boldsymbol{u}_t,$$

donde $\boldsymbol{Z}_{t-1} = (\mathbf{1}_K, \boldsymbol{x}_{t-1}, ..., \boldsymbol{x}_{t-p})$ y $\boldsymbol{u}_t \sim (\mathbf{0}_K, \boldsymbol{\Sigma}_u)$ es ruido blanco. Entonces

$$\hat{\boldsymbol{A}} = [\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{A}}_1, ..., \hat{\boldsymbol{A}}_p] = \left(\sum_{t=1}^{T} x_t Z'_{t-1}\right)\left(\sum_{t=1}^{T} \boldsymbol{Z}_{t-1}\boldsymbol{Z}'_{t-1}\right) = \boldsymbol{X}\boldsymbol{Z}'(\boldsymbol{Z}\boldsymbol{Z}')^{-1},$$

donde $\boldsymbol{X} \equiv [\boldsymbol{x}_1, ..., \boldsymbol{x}_T]$ es $K \times T$ y $\boldsymbol{Z} \equiv [\boldsymbol{Z}_0, ..., \boldsymbol{Z}_{T-1}]$ es $(Kp+1) \times T$.

- Definamos el vector $(pK^2 + K) \times 1$ dado por $\boldsymbol{\alpha} = vec(\boldsymbol{A})$ con $\boldsymbol{A} = [\boldsymbol{\mu}, \boldsymbol{A}_1, ..., \boldsymbol{A}_p]$, entonces

$$\sqrt{T}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}),$$

donde $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}} = plim(\frac{1}{T}\boldsymbol{Z}\boldsymbol{Z}')^{-1} \otimes \boldsymbol{\Sigma}_u$.

# VAR - Estimadores MCO

- Un estimador consistente de la varianza-covarianza de los errores es

$$\hat{\boldsymbol{\Sigma}}_u = \frac{\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}'}{T - Kp - 1},$$

donde $\hat{\boldsymbol{U}} = \boldsymbol{X} - \hat{\boldsymbol{A}}\boldsymbol{Z}$ son los residuos MCO.

- Si los errores se asumen normales, $\boldsymbol{u}_t \sim iid(\boldsymbol{0}_K, \boldsymbol{\Sigma}_u)$, entonces el estimador de máxima verosimilitud coincide con MCO, y

$$\hat{\boldsymbol{\Sigma}}_u^{MV} = \frac{\hat{\boldsymbol{U}}\hat{\boldsymbol{U}}'}{T}.$$

- Método delta: si tenemos una función de dimensión $N$ $\phi(\boldsymbol{A}) : \Re^{pK^2+K} \to \Re^N$ (por ejemplo, las funciones impulso respuesta), entonces usando $\hat{\phi} = \phi(\hat{\boldsymbol{A}})$,

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \frac{\partial \phi}{\partial \boldsymbol{\alpha}'} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}} \frac{\partial \phi'}{\partial \boldsymbol{\alpha}}),$$

donde $\frac{\partial \phi}{\partial \boldsymbol{\alpha}'}$ es una matriz $N \times (pK^2 + K)$ de derivadas de $\boldsymbol{\alpha} \equiv vec(\boldsymbol{A})$.

# VAR - Definiciones de 2 variables

**(Variables predeterminadas)**

*$y_t$ está predeterminada para $z_t$ si $b_{12} = 0$.*

**(Variables estrictamente exógenas)**

*$y_t$ es estrictamente exógena a $z_t$ si $b_{12} = \gamma_{12} = 0$.*

# VAR - Causalidad de Granger de 2 variables

- Un contraste de causalidad muy usado en series de tiempo es el de causalidad en el sentido de Granger.
- En un modelo de dos variables, $\{y_t\}$ no causa $\{z_t\}$ si y solo si los coeficientes $A_{21}(L)$ son cero.
- Si todas las variables son estacionarias el contraste de Granger es

$$a_{21}(1) = a_{21}(2) = ... = a_{21}(p) = 0$$

- Más formalmente, la no causalidad en el sentido de Granger implica

$$E(z_{t+1}|z_t) = E(z_{t+1}|z_t, y_t)$$

  ($y_t$ no agrega información para predecir $z_{t+1}$)
- La generalización para $n$ variables, la variable $j$ no causa $i$ en el sentido de Granger si todos los coeficientes $A_{ij}(L)$ son 0.
- La causalidad o no de Granger no implica causalidad o no contemporánea, y viceversa.

# VAR - Causalidad de Granger

Granger and Newbold (1977, p. 225): "Possibly cause is too strong a term, or one too emotionally laden, to be used. A better term might be temporally related, but since cause is such a simple term we shall continue to use it".
Granger (1980, p. 333): "provided I define what I personally mean by causation, I can use the term. I could, if I so wish, replace the word cause throughout my lecture by some other words, such as oshkosh or snerd, but what would be gained? It is like saying that whenever I use x, you would prefer me to use z."

# Matrices descomponibles

- Una **permutación** es una función uno-a-uno del conjunto $\{1, 2, \ldots, n\}$ a sí mismo. Sirve para intercambiar el orden de un conjunto ordenado. Una **matriz de permutación** $\boldsymbol{P}$ permuta las columnas (o filas) de una matriz identidad, ej. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Para estas matrices $\boldsymbol{P}^{-1} = \boldsymbol{P}'$.
- Si a una matriz cualquiera $\boldsymbol{A}$ le aplicamos $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ entonces quedan intercambiadas las columnas y las filas (simultáneamente) de $\boldsymbol{A}$. Si $\boldsymbol{A}_1$ es la matriz de rezagos entonces $\boldsymbol{P}^{-1}\boldsymbol{x}_t = \boldsymbol{P}^{-1}\boldsymbol{A}_1\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{x}_{t-1} + \boldsymbol{P}^{-1}\boldsymbol{u}_{t-1}$ reordena las variables endógenas.
- Sea una matriz $n \times n$ $\boldsymbol{A}$, se la llama **descomponible** si existe una matriz de permutación $\boldsymbol{P}$ tal que $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix}$, donde $\boldsymbol{A}_{11}$ y $\boldsymbol{A}_{22}$ son submatrices cuadradas. Si esto no es posible, entonces $\boldsymbol{A}$, se la define como **indescomponible**. Si $\boldsymbol{A}_1$ se puede reducir, entonces hay dos tipos de variables: unas (tipo 2) que no se ven afectadas (en el sentido de Granger) por las otras (tipo 1).
- Si además tenemos que existe una matriz de permutación $\boldsymbol{P}$ tal que $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix}$, esta matriz se define como **completamente descomponible**.

# VAR - Identificación de 2 variables

El problema de identificación tiene que ver con la posibilidad o no de despejar los parámetros estructurales a partir de la forma reducida.
Supongamos el VAR(1) para dos variables.

- El modelo estructural tiene 10 parámetros a estimar: $(b_{10}, b_{20}, b_{12}, b_{21}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \sigma_y, \sigma_z)$.
- El modelo reducido tiene 9 parámetros a estimar: $(a_{10}, a_{20}, a_{11}, a_{12}, a_{21}, a_{22}, \sigma_1, \sigma_2, \sigma_{12})$.
- Los parámetros no se pueden identificar a menos que impongamos una restricción adicional.
- Para un modelo bivariadosolo hace falta una restricción.)

# VAR - Identificación basada en restricciones de corto plazo de 2 variables

Una solución es descomponer los residuos en **forma triangular**, imponiendo 0s en la matriz $B$ en algunos elementos fuera de la diagonal. Esto se llama la descomposición de **Cholesky** o el método **recursivo** de Sims (1980).

- Por ejemplo podemos imponer $b_{21} = 0$: $z_t$ puede tener un efecto contemporáneo en $y_t$ (se dice que $z$ causa $y$ contempóraneamente), pero $y_t$ no afecta $z_t$.
- Entonces, tenemos

$$
\begin{aligned}
u_{1t} &= w_{yt} - b_{12} w_{zt} \\
u_{2t} &= w_{zt}
\end{aligned}
$$

tal que tenemos 3 ecuaciones con 3 incógnitas

$$
\begin{aligned}
var(u_1) &= \sigma_y^2 + b_{12}\sigma_z^2 \\
var(u_2) &= \sigma_z^2 \\
cov(u_1, u_2) &= -b_{12}\sigma_z^2
\end{aligned}
$$

- Todos los parámetros del modelo estructural se pueden encontrar. Hacer para un caso empírico.

# Descomposición de Cholesky

En general usando el modelo general para $K$ variables hay que resolver
$\mathbf{\Sigma}_u = E(\boldsymbol{uu}') = \boldsymbol{B}_0^{-1}E(\boldsymbol{ww}')\boldsymbol{B}_0^{-1'} = \boldsymbol{B}_0^{-1}\mathbf{\Sigma}_w\boldsymbol{B}_0^{-1'}$.

*En matemáticas, la factorización o* **descomposición de Cholesky** *toma su nombre del matemático André-Louis Cholesky, quien encontró que una matriz simétrica definida positiva (ej. $\mathbf{\Sigma}_u$) puede ser descompuesta como el producto de una matriz triangular inferior (ej. $\boldsymbol{B}_0^{-1}\mathbf{\Sigma}_w^{1/2}$) y la traspuesta de la matriz triangular inferior (ej. $\mathbf{\Sigma}_w^{1/2}\boldsymbol{B}_0^{-1'}$). La matriz triangular inferior es el triángulo de Cholesky de la matriz original positiva definida. Es una manera de resolver sistemas de ecuaciones matriciales.*

# VAR - Identificación basada en restricciones de corto plazo

- Notemos que hay $K^2 - K$ elementos en $\boldsymbol{B}_0$ a identificar. Además tenemos $K$ varianzas de $w_t$ en $\boldsymbol{\Sigma}_w$. Entonces hay que identificar $K^2$ parámetros.
- Tenemos $(K^2 + K)/2$ elementos adicionales en $\boldsymbol{\Sigma}_u$.
- Entonces tenemos que imponer $K^2 - (K^2 + K)/2 = (K^2 - K)/2$ restricciones. Así, $\hat{\boldsymbol{B}}_0^{-1} = chol(\hat{\boldsymbol{\Sigma}}_u)$, donde $chol()$ es la descomposición triangular de Cholesky.
- Las restricciones pueden basarse en varias cosas:
- Rezagos en como se transmite la información. Variables ajustan más rápido que otras.
- Restricciones físicas. Variables reales se mueven más lento que monetarias.
- Restricciones institucionales o de la estructura de mercado. Ej: oferta inelástica de corto plazo.
- Valores de parámetros exógenos al modelo.

# Christiano, Eichenbaum, y Evans (1996)

Christiano, L. J., M. Eichenbaum, and C. Evans (1996): "The effects of monetary policy shocks: Evidence from the flow of funds," *Review of Economics and Statistics*, 78(1), 16-34.

- Para identificar los shocks monetarios (cambios en la tasa de interés) se necesitan asumir restricciones en el VAR estructural.
- Descomposición triangular de Cholesky. Se asume una jerarquía en los efectos contemporáneos.
- Tenemos tres variables: producto, inflación y tasa de interés: $(y, \pi, r)$. Entonces se asume que $y$ (variable real) no cambia contemporáneamente ante un shock en $r$ o en $\pi$. $\pi$ cambia ante un shock en $y$ pero no en $r$. $r$ cambia ante un shock en $y$ y $\pi$.
- Entonces un shock monetario es un cambio en $r$, medido directamente de los residuos.

# VAR - Funciones impulso respuesta de 2 variables

- Los modelos VAR se pueden escribir como modelos MA de infinitos rezagos usando el resultado $\boldsymbol{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{A}_1^i \boldsymbol{u}_{t-i}$. En particular,

$$\boldsymbol{u}_t = \sum_{i=0}^{\infty} \boldsymbol{\theta}(i) \boldsymbol{w}_{t-i}$$

donde $\boldsymbol{\theta}(i) = \boldsymbol{A}_1^i \boldsymbol{B}_0^{-1} = \begin{bmatrix} \theta_{11}(i) & \theta_{12}(i) \\ \theta_{21}(i) & \theta_{2}(i) \end{bmatrix}$, tal que $\boldsymbol{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \boldsymbol{\theta}(i) \boldsymbol{w}_{t-i}$.

- Los elementos de la matriz $\theta_i$ son los multiplicadores de impacto que determinan las **funciones de impulso respuesta**. Por ejemplo, $\theta_{12}(0)$ es el efecto contemporáneo de un shock en $w_z$ en $y$; $\theta_{12}(1)$ es el efecto con un rezago de un shock en $w_z$ en $y$, (o sea $w_{zt}$ en $y_{t+1}$).

# VAR - Funciones impulso respuesta

- En general se definen las funciones impulso respuesta como $\theta_{jk,i} = \frac{\partial y_{j,t+i}}{\partial w_{k,t}}$, donde $j,k = 1,2,...,K$. Es decir como los efectos de los shocks estructurales.
- Definamos $\mathbf{\Theta}_i = [\theta_{jk,i}]$.
- El efecto acumulado $h$ periodos de un shock en la variable $k$ sobre $j$ es $\sum_{i=0}^{h} \theta_{jk,i}$.
- Dado que los shocks no pueden tener un efecto permanente (para variables estacionarias), debe darse que $lim_{i\to\infty}\theta_{jk,i} = 0$.
- El efecto acumulado en el largo plazo de un shock en la variable $k$ sobre $j$ es $\sum_{i=0}^{\infty} \theta_{jk,i}$.
- También debemos tener $\sum_{i=0}^{\infty} \theta_{jk,i}^2 < \infty$.

# VAR - Funciones impulso respuesta

Para el modelo general de dimensión $K$ y $p$ rezagos definamos
$X_t = \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{U}_t$,
donde

$$\boldsymbol{X}_t = \begin{pmatrix} x_t \\ \vdots \\ x_{p-p+1} \end{pmatrix}, \boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \dots & \boldsymbol{A}_{p-1} & A_p \\ \boldsymbol{I}_K & 0 & \dots & 0 & 0 \\ 0 & \boldsymbol{I}_K & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \boldsymbol{I}_K \end{bmatrix}, \boldsymbol{U}_t = \begin{pmatrix} \boldsymbol{u}_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

La restricción de estabilidad $det(\boldsymbol{A}(r)) = det(\boldsymbol{I}_K - \boldsymbol{A}_1 r - \dots - \boldsymbol{A}_p r^p) \neq 0, \forall r, |r| < 1$, implica $det(\boldsymbol{I}_{Kp} - \boldsymbol{A}r) \neq 0, \forall r, |r| < 1$.

# VAR - Funciones impulso respuesta

Un modelo VAR(p) estable puede ser representado como la suma ponderada de las innovaciones presentes y pasadas.

$$\boldsymbol{x}_t = \boldsymbol{A}(L)^{-1}\boldsymbol{u}_t = \sum_{i=0}^{\infty} \boldsymbol{J}\boldsymbol{A}^i\boldsymbol{J}'\boldsymbol{J}\boldsymbol{U}_{t-i} = \sum_{i=0}^{\infty} \boldsymbol{\Theta}_i\boldsymbol{u}_{t-i},$$

donde $\boldsymbol{J} \equiv [\boldsymbol{I}_K, \boldsymbol{0}_{K\times K(p-1)}]$.
También podemos escribir

$$\boldsymbol{X}_{t+i} = \boldsymbol{A}^{i+1}\boldsymbol{X}_{t-1} + \sum_{j=0}^{i} \boldsymbol{A}^j\boldsymbol{U}_{t+i-j}$$

tal que

$$\boldsymbol{x}_{t+i} = \boldsymbol{J}\boldsymbol{A}^{i+1}\boldsymbol{X}_{t-1} + \sum_{j=0}^{i} \boldsymbol{J}\boldsymbol{A}^j\boldsymbol{J}'\boldsymbol{u}_{t+i-j}.$$

# VAR - Funciones impulso respuesta

Entonces tenemos las funciones impulso respuesta como:

$$
\begin{aligned}
\boldsymbol{\Theta}_0 &= \boldsymbol{B}_0^{-1} \\
\boldsymbol{\Theta}_1 &= \boldsymbol{JA}J'\boldsymbol{B}_0^{-1} \\
\boldsymbol{\Theta}_2 &= \boldsymbol{JA}^2\boldsymbol{J}'\boldsymbol{B}_0^{-1} \\
&\vdots
\end{aligned}
$$

# VAR - Descomposición de varianza

- Usando la forma de VMA anterior podemos encontrar el *error de predicción h-periodos adelante* ($h$-period forecast error)

$$\mathbf{x}_{t+h} - E_t \mathbf{x}_{t+h} = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \mathbf{w}_{t+h-i}$$

- La varianza error de predicción $h$-periodos adelante es

$$\boldsymbol{\Sigma}(h) = var\left(\sum_{i=0}^{h-1} \boldsymbol{\theta}_i \mathbf{w}_{t+h-i}\right) = \sum_{i=0}^{h-1} \boldsymbol{\theta}_i \mathbf{I}_K \boldsymbol{\theta}_i'$$

- Por ejemplo para el ejemplo de dos variables, tomemos la variable $y$, y tenemos

$$\begin{aligned} y_{t+h} - E_t y_{t+h} \quad &= \theta_{11,0} w_{y,t+h} + \theta_{11,1} w_{y,t+h-1} + \ldots + \theta_{11,h-1} w_{y,t+1} + \\ &\phi_{12,0} w_{z,t+h} + \theta_{12,1} w_{z,t+h-1} + \ldots + \theta_{12,h-1} w_{z,t+1} \end{aligned}$$

$$\sigma_y^2(n) = \sigma_y^2(\theta_{11}^2(0) + \theta_{11}^2(1) + \ldots + \theta_{11}^2(h-1)) + \sigma_z^2(\theta_{12}^2(0) + \theta_{12}^2(1) + \ldots + \theta_{12}^2(h-1))$$

- Entonces podemos computar las proporciones que se deben a cada shock (forecast error variance decomposition):

$$\frac{\sigma_y^2(\theta_{11}^2(0) + \theta_{11}^2(1) + \ldots + \theta_{11}^2(h-1))}{\sigma_y^2(h)}$$

$$\frac{\sigma_z^2(\theta_{12}^2(0) + \theta_{12}^2(1) + \ldots + \theta_{12}^2(h-1))}{\sigma_y^2(h)}$$

# VAR - Identificación basada en restricciones de largo plazo

- El modelo se puede escribir como $B(L)y_t = w_t$ donde $\boldsymbol{B}(L) = \boldsymbol{B}_0 - \boldsymbol{B}_1 L - \boldsymbol{B}_2 L^2 - ... - \boldsymbol{B}_p L^p = \boldsymbol{B}_0 \boldsymbol{A}(L)$. Entonces, $\boldsymbol{y}_t = \boldsymbol{B}(L)^{-1}\boldsymbol{w}_t = \Theta(L)w_t$. Supongamos la normalización $\boldsymbol{w}_t \sim (\boldsymbol{0}_K, \boldsymbol{I}_K)$, tal que $\boldsymbol{\Sigma}_u = \boldsymbol{B}_0^{-1}\boldsymbol{B}_0^{-1'}$.
- Los efectos de largo plazo se pueden definir como

$$\boldsymbol{\Theta}(1) = \boldsymbol{B}(1)^{-1} = \boldsymbol{A}(1)^{-1}\boldsymbol{B}_0^{-1}.$$

  Entonces restricciones en $\boldsymbol{\Theta}(1)$ pueden identificar $\boldsymbol{B}_0$.
- En particular,

$$\boldsymbol{\Sigma}_u = \boldsymbol{B}_0^{-1}\boldsymbol{B}_0^{-1'} = \boldsymbol{A}(1)\Theta(1)\Theta(1)'A(1)' \Rightarrow \boldsymbol{A}(1)^{-1}\boldsymbol{\Sigma}_u[\boldsymbol{A}(1)^{-1}]' = \boldsymbol{\Theta}(1)\boldsymbol{\Theta}(1)'$$

  Entonces tenemos que imponer $K(K-1)/2$ restricciones para identificar los parámetros de $\boldsymbol{\Theta}(1)$. Y luego $\boldsymbol{B}_0^{-1} = \boldsymbol{A}(1)\boldsymbol{\Theta}(1)$.

# VAR - Identificación basada en restricciones de largo plazo

- Esta propuesta se debe a Blanchard, O.J., y D. Quah (1989) "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 654-673.
- En su paper tienen un shock de oferta y uno de demanda, y el de demanda tiene un efecto nulo en la oferta en el "largo plazo".
- Supongamos que $\boldsymbol{x}_t = (\Delta gdp_t, ur_t)'$, crecimiento y desempleo. La primera variable tiene asociado un shock de oferta, $w_t^{AS}$ (shock tecnológico $\rightarrow$ permanente). La segunda un shock de demanda, $w_t^{AD}$ (shocks transitorios).
- Entonces se asume que

$$\boldsymbol{\Theta}(1) = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}$$

# Modelos de vectores autorregresivos (VAR)

En STATA, los modelos VAR se pueden implementar fácilmente. Consideremos un modelo con 3 variables: x,y,z endógenas y una variable exógena w. Ver:
`https://www.stata.com/manuals13/tsvarintro.pdf`
`https://www.stata.com/manuals13/tsvar.pdf`

# Modelos de vectores autorregresivos (VAR)

- `varsoc x y z` /*para seleccionar la cantidad de rezagos*/
- Los criterios de información se pueden usar en procesos multivariados para elegir el número de lags/rezagos. Tienen una lógica parecida al $R^2$ ajustado, tratan de maximizar la predicción o minimizar los errores al cuadrado, pero penalizan por variables/parámetros incluidos.
- Para un modelo de $p$ rezagos hay $pK$ regresores en cada ecuación, entonces $pK^2$ parámetros en $\boldsymbol{A}$. Si agregamos un intercepto en cada ecuación, tenemos $pK^2 + K$.
- Criterio de información de Akaike/Akaike information criterion (AIC): $AIC(p) = ln\left(det(\boldsymbol{\Sigma}_u^2)\right) + \frac{2}{T}(pK^2 + K)$
- Criterio de información bayesiano/Bayesian information criterion (BIC) o Schwarz criterion (SBC, SBIC): $SIC(p) = ln\left(det(\boldsymbol{\Sigma}_u^2)\right) + \frac{ln(T)}{T}(pK^2 + K)$.
- Notar que $AIC \geq BIC$. Entonces en general AIC selecciona más parámetros que SIC.
- Criterio de Hannan-Quinn (HQC): $HQC(p) = ln\left(det(\boldsymbol{\Sigma}_u^2)\right) + \frac{2ln(ln(T))}{T}(pK^2 + K)$.

# Modelos de vectores autoregresivos (VAR)

- `var x y z, lags(1/p) exog(w)` /*donde p es la cantidad de rezagos a usar*/
- `varlmar` /*para chequear que los residuos no tienen autocorrelación*/
- `varnorm` /*para chequear la normalidad del modelo*/
- `varstable` /*para chequear que los procesos son estables*/
  Para realizar inferencia con modelos VAR se requiere que las variables sean estacionarias en covarianzas. Para ello se debe chequear que todos los autovalores (eigenvalues) del modelo VAR sean menores que 1 en módulo.
- `vargranger` para análisis de causalidad de Granger.

# Modelos de vectores autoregresivos (VAR)

- Para modelos estructurales hay que usar `svar`.
- `https://www.stata.com/manuals13/tsvarsvar.pdf`

# Ejemplo empírico

Para ilustrar los modelos VAR consideremos el ejemplo de Lütkepohl (1993, Introduction to Multiple Time Series Analysis, New York:Springer, pp77-78). Los datos consisten en 3 variables, trimestrales de 1960 a 1982, de Alemania Occidental:

`dln_inv`: primera diferencia del log de inversión
`dln_inc`: primera diferencia del log del ingreso
`dln_consump`: primera diferencia del log de consumo

En STATA:

```
webuse lutkepohl2, clear
varsoc dln_inv dln_inc dln_consump
var dln_inc dln_consump, exog(dln_inv)
```

# Predicción para modelos VAR

Para predecir se pueden usar los mismos comandos que para procesos univariados: `fcast` y `predict`.

- `fcast compute hat, step(8)`

  Esto crea nuevas variables que empezarán con "hat" para todas las variables endógenas en el modelo.
- `fcast graph`

# Funciones de impulso respuesta (impulse response function)

Para estimaciones macroéconomicas es muy importante estudiar los efectos dinámicos de una varible sobre otras. En particular, se busca estimar el efecto de un shock en una variable sobre otras.

- Consideremos el siguiente ejemplo:

```
var x y z, lags(p)
irf create order1, set(myirf1)
irf graph oirf, impulse(y) response(z)
```

# Lecturas sugeridas

- Enders, W. "Applied Econometric Time Series", caps. 5,6.
- Johnston, J. y Dinardo, J. "Econometric Methods", cap. 9.
- Kilian, L. y Lütkepohl, H. "Structural Vector Autoregressive Analysis", caps. 2,3,4,7,8.
- Lütkepohl, H. "New Introduction to Multiple Time Series Analysis".
- Tsay, R. "Analysis of Financial Time Series", cap. 8.