Cheat Sheet Adicional

Por Marcelo Moreno - Universidad Rey Juan Carlos

The Econometrics Cheat Sheet Project

Notación matricial MCO

El modelo econométrico general:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i$$

Puede ser escrito en notación matricial como:

$$y = X\beta + u$$

Llamemos $\hat{u}$ al vector de residuos estimados ($\hat{u} \neq u$):

$$\hat{u} = y - X\hat{\beta}$$

El **objetivo** de MCO es **minimizar** la SRC:

$$\min \mathrm{SRC} = \min \sum_{i=1}^{n} \hat{u}_i^2 = \min \hat{u}^\intercal \hat{u}$$

- Definiendo $\hat{u}^\intercal \hat{u}$:

$$\hat{u}^\intercal \hat{u} = (y - X\hat{\beta})^\intercal (y - X\hat{\beta}) =$$
$$= y^\intercal y - 2\hat{\beta}^\intercal X^\intercal y + \hat{\beta}^\intercal X^\intercal X \hat{\beta}$$

- Minimizando $\hat{u}^\intercal \hat{u}$:

$$\frac{\partial \hat{u}^\intercal \hat{u}}{\partial \hat{\beta}} = -2X^\intercal y + 2X^\intercal X\hat{\beta} = 0$$

$$\hat{\beta} = (X^\intercal X)^{-1}(X^\intercal y)$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \dots & \sum x_k \\ \sum x_1 & \sum x_1^2 & \dots & \sum x_1 x_k \\ \vdots & \vdots & \ddots & \vdots \\ \sum x_k & \sum x_k x_1 & \dots & \sum x_k^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y \\ \sum y x_1 \\ \vdots \\ \sum y x_k \end{bmatrix}$$

La segunda derivada $\frac{\partial^2 \hat{u}^\intercal \hat{u}}{\partial \hat{\beta}^2} = X^\intercal X > 0$ (es un mín.)

Matriz de varianzas-covarianzas de $\hat{\beta}$

Tiene la siguiente forma:

$$\mathrm{Var}(\hat{\beta}) = \hat{\sigma}_u^2 \cdot (X^\intercal X)^{-1} =$$

$$= \begin{bmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \dots & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_k) \\ \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \mathrm{Var}(\hat{\beta}_1) & \dots & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \dots & \mathrm{Var}(\hat{\beta}_k) \end{bmatrix}$$

donde: $\hat{\sigma}_u^2 = \frac{\hat{u}^\intercal \hat{u}}{n-k-1}$

Los errores estándar están en la diagonal de:

$$\mathrm{ee}(\hat{\beta}) = \sqrt{\mathrm{Var}(\hat{\beta})}$$

Medidas de error

- $\mathrm{SRC} = \hat{u}^\intercal \hat{u} = y^\intercal y - \hat{\beta}^\intercal X^\intercal y = \sum (y_i - \hat{y}_i)^2$
- $\mathrm{SEC} = \hat{\beta}^\intercal X^\intercal y - n\overline{y}^2 = \sum (\hat{y}_i - \overline{y})^2$
- $\mathrm{STC} = \mathrm{SRC} + \mathrm{SEC} = y^\intercal y - n\overline{y}^2 = \sum (y_i - \overline{y})^2$

Matriz de varianzas-covarianzas de u

Tiene la siguiente forma:

$$\mathrm{Var}(u) = \begin{bmatrix} \mathrm{Var}(u_1) & \mathrm{Cov}(u_1, u_2) & \dots & \mathrm{Cov}(u_1, u_n) \\ \mathrm{Cov}(u_2, u_1) & \mathrm{Var}(u_2) & \dots & \mathrm{Cov}(u_2, u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(u_n, u_1) & \mathrm{Cov}(u_n, u_2) & \dots & \mathrm{Var}(u_n) \end{bmatrix}$$

Cuando no hay heterocedasticidad ni autocorrelación, la matriz de varianzas-covarianzas de u tiene la forma:

$$\mathrm{Var}(u) = \sigma_u^2 \cdot I_n = \begin{bmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \end{bmatrix}$$

donde I_n es una matriz identidad con $n \times n$ elementos.

Cuando hay **heterocedasticidad** y **autocorrelación**, la matriz de varianzas-covarianzas de u tiene la forma:

$$\mathrm{Var}(u) = \sigma_u^2 \cdot \Omega = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_{12}} & \dots & \sigma_{u_{1n}} \\ \sigma_{u_{21}} & \sigma_{u_2}^2 & \dots & \sigma_{u_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_{n1}} & \sigma_{u_{n2}} & \dots & \sigma_{u_n}^2 \end{bmatrix}$$

donde $\Omega \neq I_n$.

- Heterocedasticidad: $\mathrm{Var}(u) = \sigma_{u_i}^2 \neq \sigma_u^2$
- Autocorrelación: $\mathrm{Cov}(u_i, u_j) = \sigma_{u_{ij}} \neq 0, \ \forall i \neq j$

Omisión de variables

Casi siempre es difícil disponer de todas las variables relevantes. Por ejemplo, un modelo con todas las variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v$$

donde $\beta_2 \neq 0$, v el término de error y $\mathrm{Cov}(v|x_1, x_2) = 0$.

El modelo con las variables disponibles:

$$y = \alpha_0 + \alpha_1 x_1 + u$$

donde $u = v + \beta_2 x_2$.

Omisión de variables relevantes causa que los estimadores MCO sean **sesgados** e **inconsistentes**,porque no hay exogeneidad estricta, $\mathrm{Cov}(x_1, u) \neq 0$. Dependiendo de $\mathrm{Corr}(x_1, x_2)$ y el signo de β_2, el sesgo en $\hat{\alpha}_1$ puede ser:

	$\mathrm{Corr}(x_1, x_2) > 0$	$\mathrm{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	sesgo $(+)$	sesgo $(-)$
$\beta_2 < 0$	sesgo $(-)$	sesgo $(+)$

- Sesgo $(+)$: $\hat{\alpha}_1$ será más alto de lo que debería (incluye el efecto de x_2) $\rightarrow \hat{\alpha}_1 > \beta_1$
- Sesgo $(-)$: $\hat{\alpha}_1$ será más bajo de lo que debería (incluye el efecto de x_2) $\rightarrow \hat{\alpha}_1 < \beta_1$

Si $\mathrm{Corr}(x_1, x_2) = 0$, no hay sesgo en $\hat{\alpha}_1$, porque el efecto de x_2 será totalmente recogido por el término de error, u.

Corrección de omisión de variables

Variables proxy

Es el camino cuando la variable relevante no está disponible porque no es observable, y no hay datos disponibles.

- Una **variable proxy** es algo **relacionado** con la variable no observable que tiene datos disponibles.

Por ejemplo, el PIB per capita es una variable proxy para la calidad de vida (no observable).

Instrumental variables

Cuando una variable de interés (x) es observable pero **endógena**, el camino de variables proxy ya no es válido.

- Una **variable instrumental** (VI) **es una variable observable** (z) que está **relacionada** con la variable de interés que es endógena (x), y cumple los **requisitos**:

$$\mathrm{Cov}(z, u) = 0 \rightarrow \text{exogeneidad del instrumento}$$
$$\mathrm{Cov}(z, x) \neq 0 \rightarrow \text{relevancia del instrumento}$$

Variables instrumentales deja la variable omitida en el término de error, pero en vez de estimar el modelo por MCO, utiliza un método que reconoce la omisión de variable. Puede también corregir errores de medida.

- **Mínimos Cuadrados en Dos Etapas** (MC2E) es un método de estimar un modelo con múltiples variables instrumentales . El requisito $\mathrm{Cov}(z, u) = 0$ puede ser relajado, pero debe haber un mínimo de variables que lo satisfacen.

  El **procedimiento de estimación** de MC2E:

  1. Estimar un modelo regresando x por z usando MCO, obteniendo $\hat{x}$:
  $$\hat{x} = \hat{\pi}_0 + \hat{\pi}_1 z$$
  2. Reemplazar x por $\hat{x}$ en el modelo final y estimarlo por MCO:
  $$y = \beta_0 + \beta_1 \hat{x} + u$$

  Hay algunas cosas importantes sobre MC2E:

  – MC2E son menos eficientes que MCO cuando las variables explicativas son exógenas. El **test de Hausman** puede usarse para comprobarlo:

  H_0: los estimadores MCO son consistentes.

  Si H_0 es aceptada, los estimadores MCO son mejores que MC2E y viceversa.

  – Pueden haber algunos instrumentos (o todos) que no sean válidos. Esto se conoce como sobre-identificación, el **test de Sargan** puede usarse para comprobarlo:

  H_0: todos los instrumentos son válidos.

Criterio de información

Es usado para comparar modelos con diferente número de parámetros (p). La fórmula general:

$$\mathrm{Cr}(p) = \log(\tfrac{\mathrm{SRC}}{n}) + c_n \varphi(p)$$

donde:

- SRC es la Suma de Residuos Cuadráticos de un modelo de orden p.
- c_n es una secuencia indexada por el tamaño muestral.
- $\varphi(p)$ es una función que penaliza órdenes grandes de p.

Interpretado como el tamaño relativo de información perdida por el modelo. Orden p que min. el criterio es elegido.

Hay diferentes funciones $c_n\varphi(p)$:

- Akaike: $\mathrm{AIC}(p) = \log(\frac{\mathrm{SRC}}{n}) + \frac{2}{n}p$
- Hannan-Quinn: $\mathrm{HQ}(p) = \log(\frac{\mathrm{SRC}}{n}) + \frac{2\log(\log(n))}{n}p$
- Schwarz: $\mathrm{Sc}(p) = \log(\frac{\mathrm{SRC}}{n}) + \frac{\log(n)}{n}p$

$\mathrm{Sc}(p) \leq \mathrm{HQ}(p) \leq \mathrm{AIC}(p)$

Contraste de hipótesis no restrigido

Es una alternativa al contraste F cuando hay pocas hipótesis a probar sobre los parámetros. Sean β_i, β_j parámetros, $a, b, c \in \mathbb{R}$ constantes.

- $H_0 : a\beta_i + b\beta_j = c$
- $H_1 : a\beta_i + b\beta_j \neq c$

$$\text{Bajo } H_0\text{:} \quad t = \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\sqrt{\mathrm{Var}(a\hat{\beta}_i + b\hat{\beta}_j)}}$$

$$= \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\sqrt{a^2\mathrm{Var}(\hat{\beta}_i) + b^2 \cdot \mathrm{Var}(\hat{\beta}_j) \pm 2ab\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j)}}$$

Si $|t| > |t_{n-k-1,\alpha/2}|$, existe evidencia para rechazar H_0.

ANOVA

Descomponer la suma total de cuad. en suma de residuos cuad. y suma explicada de cuad.: $\mathrm{STC} = \mathrm{SRC} + \mathrm{SEC}$

Origen var.	Suma Cuad.	df	Suma Cuad. Prom.
Regresión	SEC	k	SEC$/k$
Residuos	SRC	$n-k-1$	SRC$/(n-k-1)$
Total	STC	$n-1$	

El estadístico F:

$$F = \frac{\text{SCP de SEC}}{\text{SCP de SRC}} = \frac{\mathrm{SEC}}{\mathrm{SRC}} \cdot \frac{n-k-1}{k} \sim F_{k,n-k-1}$$

Si $F > F_{k,n-k-1}$, existe evidencia para rechazar H_0: No hay diferencia entre las medias de los grupos.

Forma funcional incorrecta

Para comprobar si la **forma funcional** de un modelo es correcta, podemos usar el **Ramsey's RESET** (Regression Specification Error Test). Prueba el modelo original vs. un modelo con variables en potencias.

H_0: el modelo está correctamente especificado.

Procedimiento del contraste:

1. Estimar el modelo original y obtener $\hat{y}$ y R^2:
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$
2. Estimar un nuevo modelo añadiendo potencias de $\hat{y}$ y obtener el nuevo R^2_{new}:
$$\tilde{y} = \hat{y} + \tilde{\gamma}_2 \hat{y}^2 + \cdots + \tilde{\gamma}_l \hat{y}^l$$
3. Definir el estadístico de contraste, bajo $\gamma_2 = \cdots = \gamma_l = 0$ como hipótesis nula:
$$F = \frac{R^2_{\text{new}} - R^2}{1 - R^2_{\text{new}}} \cdot \frac{n-(k+1)-l}{l} \sim F_{l,n-(k+1)-l}$$

Si $F > F_{l,n-(k+1)-l}$, hay evidencia para rechazar H_0.

Regresión logística

Cuando hay una variable dependiente binaria (0, 1), el modelo de regresión lineal ya no es válido, podemos usar la regresión logística en su lugar. Por ejemplo, un **modelo logit**:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i + u_i)}} = \frac{e^{\beta_0 + \beta_1 x_i + u_i}}{1 + e^{\beta_0 + \beta_1 x_i + u_i}}$$

donde $P_i = \mathrm{E}(y_i = 1 \mid x_i)$ and $(1 - P_i) = \mathrm{E}(y_i = 0 \mid x_i)$

La **razón de probabilidades** (a favor de $y_i = 1$):

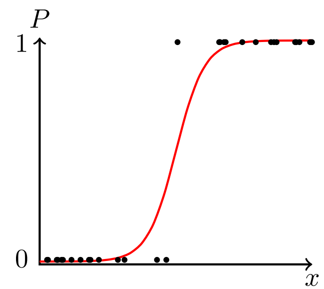
$$\frac{P_i}{1 - P_i} = \frac{1 + e^{\beta_0 + \beta_1 x_i + u_i}}{1 + e^{-(\beta_0 + \beta_1 x_i + u_i)}} = e^{\beta_0 + \beta_1 x_i + u_i}$$

Tomando el logaritmo natural de la razón de probabilidades, se obtiene el **logit**:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 x_i + u_i$$

P_i se encuentra entre 0 y 1, pero L_i va desde $-\infty$ a $+\infty$.

Si L_i es positivo, significa que cuando x_i incrementa, la probabilidad de que $y_i = 1$ incrementa, y viceversa.



Definiciones estadísticas

Sean ξ, η variables aleatorias, $a, b \in \mathbb{R}$ constantes, y P denota probabilidad.

Media

Definición: $E(\xi) = \sum_{i=1}^{n} \xi_i \cdot P[\xi = \xi_i]$

Media poblacional: $\mathrm{E}(\xi) = \frac{1}{N}\sum_{i=1}^{N} \xi_i$

Media muestral: $\mathrm{E}(\xi) = \frac{1}{n}\sum_{i=1}^{n} \xi_i$

Algunas propiedades:

- $\mathrm{E}(a) = a$
- $\mathrm{E}(\xi + a) = \mathrm{E}(\xi) + a$
- $\mathrm{E}(a \cdot \xi) = a \cdot \mathrm{E}(\xi)$
- $\mathrm{E}(\xi \pm \eta) = \mathrm{E}(\xi) + \mathrm{E}(\eta)$
- $\mathrm{E}(\xi \cdot \eta) = \mathrm{E}(\xi) \cdot \mathrm{E}(\eta)$ sólo si ξ y η son independientes.
- $\mathrm{E}(\xi - \mathrm{E}(\xi)) = 0$
- $\mathrm{E}(a \cdot \xi + b \cdot \eta) = a \cdot \mathrm{E}(\xi) + b \cdot \mathrm{E}(\eta)$

Varianza

Definición: $\mathrm{Var}(\xi) = \mathrm{E}(\xi - \mathrm{E}(\xi))^2$

Varianza poblacional: $\mathrm{Var}(\xi) = \dfrac{\sum_{i=1}^{N}(\xi_i - \mathrm{E}(\xi))^2}{N}$

Varianza muestral: $\mathrm{Var}(\xi) = \dfrac{\sum_{i=1}^{n}(\xi_i - \mathrm{E}(\xi))^2}{n-1}$

Algunas propiedades:

- $\mathrm{Var}(a) = 0$
- $\mathrm{Var}(\xi + a) = \mathrm{Var}(\xi)$
- $\mathrm{Var}(a \cdot \xi) = a^2 \cdot \mathrm{Var}(\xi)$
- $\mathrm{Var}(\xi \pm \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta) \pm 2 \cdot \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Var}(a \cdot \xi \pm b \cdot \eta) = a^2 \cdot \mathrm{Var}(\xi) + b^2 \cdot \mathrm{Var}(\eta) \pm 2ab \cdot \mathrm{Cov}(\xi, \eta)$

Covarianza

Definición: $\mathrm{Cov}(\xi, \eta) = \mathrm{E}[(\xi - E(\xi)) \cdot (\eta - E(\eta))]$

Covarianza poblacional:

$$\frac{\sum_{i=1}^{N}(\xi_i - \mathrm{E}(\xi)) \cdot (\eta_i - \mathrm{E}(\eta))}{N}$$

Covarianza muestral:

$$\frac{\sum_{i=1}^{n}(\xi_i - \mathrm{E}(\xi)) \cdot (\eta_i - \mathrm{E}(\eta))}{n-1}$$

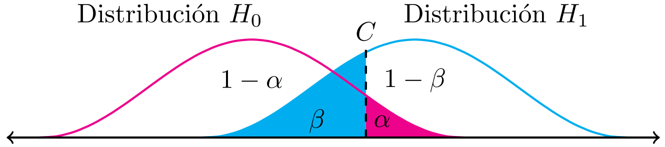
Algunas propiedades:

- $\mathrm{Cov}(\xi, a) = 0$
- $\mathrm{Cov}(\xi + a, \eta + b) = \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Cov}(a \cdot \xi, b \cdot \eta) = ab \cdot \mathrm{Cov}(\xi, \eta)$
- $\mathrm{Cov}(\xi, \xi) = \mathrm{Var}(\xi)$
- $\mathrm{Cov}(\xi, \eta) = \mathrm{Cov}(\eta, \xi)$

Contraste de hipótesis (extra)

	H_0 verdadera	H_0 falsa
Rechazar H_0	Falso positivo Error Tipo I (α)	Verdadero pos. ($1-\beta$)
Aceptar H_0	Verdadero neg. ($1-\alpha$)	Falso negativo Error Tipo II (β)

Típico contraste de una cola:



donde $(1-\alpha)$ es nivel de confianza, α es nivel de significación, C es valor crítico, $(1-\beta)$ es potencia estadística.

Bootstraping

Problema - Aprox. asint. a las distribuciones de los estadísticos de contraste no funcionan en muestras pequeñas.
Solución - Boostrap es básicamente muestreo con reemplazo. Los datos observados se tratan como una población y se extraen varias muestras para recalcular un estimador o estadístico varias veces (mejora la precisión).

VAR (Vector Autoregressive)

Un modelo VAR captura **interacciones dinámicas** entre series temporales. El VAR(p):

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + B_0 x_t + \cdots + B_q x_{t-q} + CD_t + u_t$$

donde:

- $y_t = (y_{1t}, \ldots, y_{Kt})^\intercal$ es un vector de K series temporales observables endógenas.
- A_i's son $K \times K$ matrices de coeficientes.
- $x_t = (x_{1t}, \ldots, x_{Mt})^\intercal$ es un vector de M series temporales observables exógenas.
- B_j's son $K \times M$ matrices de coeficientes.
- D_t es un vector que contiene todos los términos deterministas, que pueden ser: una constante, tendencia lineal, variables estacionales binarias, y/o cualquier otra variable ficticia especificada por el usuario.
- C es una matriz de coeficientes de dimensión apropiada.
- $u_t = (u_{1t}, \ldots, u_{Kt})^\intercal$ es un vector de K series de ruido blanco.

El proceso es **estable** si:

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{para} \quad |z| \leq 1$$

esto es, **no hay raíces** en y sobre el círculo unitario complejo.

Por ejemplo, un modelo VAR con dos variables endógenas ($K=2$), dos retardos ($p=2$), una variable exógena contemporánea ($M=1$), constante (const) y tendencia (Tend$_t$):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{11,2} & a_{12,2} \\ a_{21,2} & a_{22,2} \end{bmatrix} \cdot \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} \cdot [x_t] + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \cdot \begin{bmatrix} \text{const} \\ \text{Tend}_t \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Visualizando las ecuaciones por separado:

$$y_{1t} = a_{11,1} y_{1,t-1} + a_{12,1} y_{2,t-1} + a_{11,2} y_{1,t-2} + a_{12,2} y_{2,t-2} + b_{11} x_t + c_{11} + c_{12} \text{Tend}_t + u_{1t}$$
$$y_{2t} = a_{21,1} y_{2,t-1} + a_{22,1} y_{1,t-1} + a_{21,2} y_{2,t-2} + a_{22,2} y_{1,t-2} + b_{21} x_t + c_{21} + c_{22} \text{Tend}_t + u_{2t}$$

Si hay una raíz unitaria, el determinante es cero para $z=1$, entonces una o todas las variables son integrados y el modelo VAR ya no es apropiado (es inestable).

VECM (Vector Error Correction Model)

Si existen **relaciones cointegradoras** en un sistema de variables, la forma VAR no es la más conveniente. Es mejor usar un VECM, esto es, el VAR en niveles sustrayendo y_{t-1} de ambos lados. El VECM($p-1$):

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + B_0 x_t + \cdots + B_q x_{t-q} + CD_t + u_t$$

donde:

- y_t, x_t, D_t y u_t son como especificados en VAR.
- $\Pi = -(I_K - A_1 - \cdots - A_p)$ para $i = 1, \ldots, p-1$; Πy_{t-1} es referido como la parte a **largo plazo**.
- $\Gamma_i = -(A_{i+1} + \cdots + A_p)$ para $i = 1, \ldots, p-1$ es referido como parámetros a **corto plazo**.
- A_i, B_j y C son matrices de coeficientes de dimensiones apropiadas.

Si el proceso VAR(p) es inestable (no hay raíces), Π puede ser escrito como el producto de $(K \times r)$ matrices α (**matriz de carga**) y β (**matriz de cointegración**) con $\text{rg}(\Pi) = \text{rg}(\alpha) = \text{rg}(\beta) = r$ (**rango cointegrador**) como $\Pi = \alpha\beta^\intercal$.

- $\beta^\intercal y_{t-1}$ contiene las relaciones cointegradoras.

Por ejemplo, si hay tres variables endógenas ($K=3$) con dos relaciones cointegradoras ($r=2$), la parte a largo plazo del VECM:

$$\Pi y_{t-1} = \alpha\beta^\intercal y_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} = \begin{bmatrix} \alpha_{11} ec_{1,t-1} + \alpha_{12} ec_{2,t-1} \\ \alpha_{21} ec_{1,t-1} + \alpha_{22} ec_{2,t-1} \\ \alpha_{31} ec_{1,t-1} + \alpha_{32} ec_{2,t-1} \end{bmatrix}$$

donde:

$$ec_{1,t-1} = \beta_{11} y_{1,t-1} + \beta_{21} y_{2,t-1} + \beta_{31} y_{3,t-1}$$
$$ec_{2,t-1} = \beta_{12} y_{1,t-1} + \beta_{22} y_{2,t-1} + \beta_{32} y_{3,t-1}$$