

# Hoja de Referencia Adicional

Por Marcelo Moreno Porras - Universidad Rey Juan Carlos

The Econometrics Cheat Sheet Project

## Notación matricial MCO

El modelo econométrico general:

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

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:

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

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

$$\begin{aligned} \hat{u}^\top \hat{u} &= (y - X\hat{\beta})^\top (y - X\hat{\beta}) \\ &= y^\top y - 2\hat{\beta}^\top X^\top y + \hat{\beta}^\top X^\top X \hat{\beta} \end{aligned}$$

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

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

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

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

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

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

Tiene la siguiente forma:

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \hat{\sigma}_u^2 \cdot (X^\top X)^{-1} \\ &= \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \dots & \text{Cov}(\hat{\beta}_0, \hat{\beta}_k) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \dots & \text{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\hat{\beta}_k, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \dots & \text{Var}(\hat{\beta}_k) \end{bmatrix} \end{aligned}$$

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

Los errores estándar están en la diagonal de:

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

## Medidas de error

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

## Matriz de varianzas-covarianzas de $u$

Tiene la siguiente forma:

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

Bajo no heterocedasticidad y no autocorrelación, la matriz de varianzas-covarianzas:

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

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

Bajo **heterocedasticidad** y **autocorrelación**, la matriz de varianzas-covarianzas:

$$\text{Var}(u) = \sigma_u^2 \cdot \Omega = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} & \dots & \sigma_{u1n} \\ \sigma_{u21} & \sigma_{u2}^2 & \dots & \sigma_{u2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{un1} & \sigma_{un2} & \dots & \sigma_{un}^2 \end{bmatrix}$$

donde  $\Omega \neq I_n$ .

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

## 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  $\text{Cov}(v | x_1, x_2) = 0$ .

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 puede causar que los estimadores MCO sean **sesgados** e **inconsistentes**, porque no hay exogeneidad estricta,  $\text{Cov}(x_1, u) \neq 0$ . Dependiendo de  $\text{Corr}(x_1, x_2)$  y el signo de  $\beta_2$ , el sesgo en  $\hat{\alpha}_1$  puede ser:

	$\text{Corr}(x_1, x_2) > 0$	$\text{Corr}(x_1, x_2) < 0$
$\beta_2 > 0$	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  $\text{Corr}(x_1, x_2) = 0$ , 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**:

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

$\text{Cov}(z, x) \neq 0 \rightarrow$  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. Que  $\text{Cov}(z, u) = 0$  puede ser relajado, pero debe haber un mínimo de variables que lo satisfacen.

El **procedimiento de estimación** de MC2E:

- Estimar un modelo regresando  $x$  por  $z$  usando MCO, obteniendo  $\hat{x}$ :

$$\hat{x} = \hat{\pi}_0 + \hat{\pi}_1 z$$

- 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 **contraste de Hausman** puede usarse para comprobarlo:

$H_0$ : los estimadores MCO son consistentes.

Si  $H_0$  no es rechazada, los estimadores MCO son mejores que MC2E y viceversa.

- Cuando se usan más instrumentos que variables endógenas, el modelo puede estar sobre-identificado; el **contraste de Sargan** puede usarse para comprobarlo:

$H_0$ : todos los instrumentos son válidos.

# Criterio de información

Comparar modelos con diferente número de parámetros ( $p$ ). La fórmula general:

$$Cr(p) = \log\left(\frac{SRC}{n}\right) + c_n \varphi(p)$$

donde:

- SRC 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 mín. el criterio es elegido.

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

- Akaike:  $AIC(p) = \log\left(\frac{SRC}{n}\right) + \frac{2}{n}p$
  - Hannan-Quinn:  $HQ(p) = \log\left(\frac{SRC}{n}\right) + \frac{2 \log(\log(n))}{n}p$
  - Schwarz / Bayesian:  $BIC(p) = \log\left(\frac{SRC}{n}\right) + \frac{\log(n)}{n}p$
- $BIC(p) \leq HQ(p) \leq AIC(p)$

# Contraste de hipótesis no restringido

Alternativa al contraste F cuando hay pocas hipótesis a probar sobre los parámetros. Sean  $\beta_i, \beta_j$  parámetros,  $a, b, c \in \mathbb{R}$  constantes.

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

$$\text{Bajo } H_0: \quad t = \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\text{ee}(a\hat{\beta}_i + b\hat{\beta}_j)}$$
$$= \frac{a\hat{\beta}_i + b\hat{\beta}_j - c}{\sqrt{a^2 \text{Var}(\hat{\beta}_i) + b^2 \cdot \text{Var}(\hat{\beta}_j) + 2ab \text{Cov}(\hat{\beta}_i, \hat{\beta}_j)}}$$

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

# ANOVA

Descomponer STC:

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

- $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$
- $H_1 : \beta_1 \neq 0$  and/or  $\beta_2 \neq 0 \dots$  and/or  $\beta_k \neq 0$

Bajo  $H_0$ :

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

Si  $F > F_{k, n-k-1}$ , existe evidencia para rechazar  $H_0$ .

# Datos de panel

Observaciones de  $n$  entidades durante  $T$  períodos.

$$y_{it} = X_{it}\beta + \alpha_i + u_{it}$$

$\alpha_i$  es heterogeneidad no observada invariante en el tiempo.

## Pooled OLS

- Aplicar MCO a los datos directamente.
- Supuesto:  $\alpha_i$  es constante.

**Modelo de efectos fijos** (within estimator)

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_{it})\beta + (\alpha_i - \bar{\alpha}_i) + (u_{it} - \bar{u}_i)$$

- Se realiza centrado para eliminar  $\alpha_i$ .
- Controla efectos específicos de la entidad no observados.
- Supuesto:  $\text{Corr}(X_{it}, \alpha_i) \neq 0$ .

**Modelo de variable ficticia de mín. cuad.** (LSDV)

Se agregan variables ficticias para cada entidad y/o período de tiempo para capturar los efectos fijos.

**Modelo de primeras diferencias**

$$y_{it} - y_{i,t-1} = (X_{it} - X_{i,t-1})\beta + (\alpha_i - \alpha_i) + (u_{it} - u_{i,t-1})$$

- Se realizan primeras diferencias para eliminar  $\alpha_i$ .
- Supuesto:  $\text{Corr}(u_{it} - u_{i,t-1} \mid X_{it} - X_{i,t-1}) = 0$ .

**Modelo de efectos aleatorios**

$$y_{it} = X_{it}\beta + \alpha_i + \epsilon_{it} \text{ donde } u_{it} = \alpha_i + \epsilon_{it}$$

- Supuesto:  $\text{Corr}(X_{it}, \alpha_i) = 0$ .

# Regresión logística

Variable dependiente binaria (0, 1). **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  $P_i = E(y_i = 1 \mid x_i)$  y  $(1 - P_i) = E(y_i = 0 \mid x_i)$

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

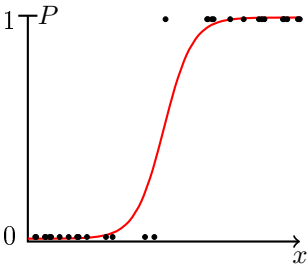
$$\frac{P_i}{1 - P_i} = \frac{1 + e^{\beta_0 + \beta_1 x_i + u_i}}{1 + e^{-(\beta_0 + \beta_1 x_i + u_i)}} = e^{\beta_0 + \beta_1 x_i + u_i}$$

Tomando el logaritmo natural de la razón, 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.



# Forma funcional incorrecta

**Ramsey's RESET** (Regression Specification Error Test).

$H_0$ : el modelo está correctamente especificado.

1. Estimar el modelo original y obtener  $\hat{y}$  y  $R^2$ :  
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$
  2. Estimar modelo con potencias de  $\hat{y}$  y obtener  $R^2_{\text{new}}$ :  
$$\tilde{y} = \hat{y} + \tilde{\gamma}_2 \hat{y}^2 + \dots + \tilde{\gamma}_l \hat{y}^l$$
  3. Estadístico, bajo  $\gamma_2 = \dots = \gamma_l = 0$  como  $H_0$ :  
$$F = \frac{R^2_{\text{new}} - R^2}{1 - R^2_{\text{new}}} \cdot \frac{n - (k+1) - l}{l} \sim F_{l, n - (k+1) - l}$$
- Si  $F > F_{l, n - (k+1) - l}$ , hay evidencia para rechazar  $H_0$ .

# Definiciones estadísticas

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

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

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

Propiedades de la media:

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

**Varianza**  $\text{Var}(\xi) = E[(\xi - E(\xi))^2]$

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

Propiedades de la varianza:

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

**Covarianza**  $\text{Cov}(\xi, \eta) = E[(\xi - E(\xi)) \cdot (\eta - E(\eta))]$

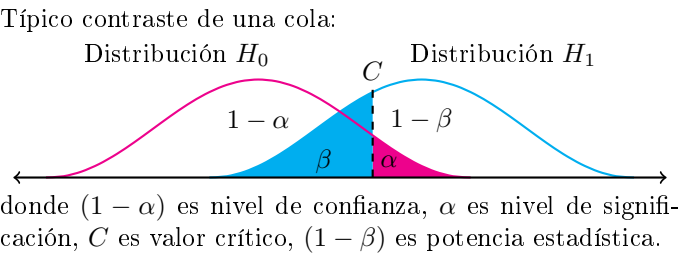
Covarianza muestral:  $\frac{\sum_{i=1}^n (\xi_i - E(\xi)) \cdot (\eta_i - E(\eta))}{n - 1}$

Propiedades de la covarianza:

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

# Contraste de hipótesis

	$H_0$ verdadera	$H_0$ falsa
Rechazar $H_0$	Falso positivo Error Tipo I ( $\alpha$ )	Verdadero pos. ( $1 - \beta$ )
No rechazar $H_0$	Verdadero neg. ( $1 - \alpha$ )	Falso negativo Error Tipo II ( $\beta$ )



# Bootstrapping

**Problema** - Aprox. asint. a las distribuciones de los estadísticos de contraste no funcionan en muestras pequeñas.  
**Solución** - Bootstrap es 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( $p$ ):  
$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + B x_t + C D_t + u_t$$

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

**Condición de estabilidad:**

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

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 <sub>$t$</sub> ):

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

Visualizando las ecuaciones por separado:

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

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

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).

# SVAR (Structural VAR)

En un modelo VAR, las interpretaciones de causalidad no son explícitas, y los resultados pueden variar según el orden de las variables. Un SVAR extiende el VAR al imponer restricciones sobre las matrices  $A$  y/o  $B$ . Esto permite una interpretación causal y un análisis de shocks sin necesidad de depender de un orden arbitrario.

Por ejemplo, un modelo SVAR( $p$ ) básico:

$$A y_t = A[A_1, \dots, A_p] y_{t-1} + B \varepsilon_t$$

donde:

- $u_t = A^{-1} B \varepsilon_t$
- $A, B$  son  $(K \times K)$  matrices.

# 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( $p - 1$ ):

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + B x_t + C D_t + u_t$$

donde:

- $\Delta y_t = (\Delta y_{1t}, \dots, \Delta y_{Kt})^\top$  es un vector de  $K$  series temporales observables endógenas.
- $\Pi y_{t-1}$  es la parte **largo plazo**.
  - $\Pi = -(I_K - A_1 - \dots - A_p)$  para  $i = 1, \dots, p - 1$
  - $\Pi = \alpha \beta^\top$
  - $\alpha$  es la **matriz de carga** ( $K \times r$ ). Representa la velocidad de ajuste.
  - $\beta$  es la **matriz de cointegración** ( $K \times r$ ).
  - $\beta^\top y_{t-1}$  es la **ecuación de cointegración**. Representa el equilibrio a largo plazo.
  - $\text{rg}(\Pi) = \text{rg}(\alpha) = \text{rg}(\beta) = r$  es el **rango cointegrador**.
- $\Gamma_i = -(A_{i+1} + \dots + A_p)$  para  $i = 1, \dots, p - 1$  son los parámetros a **corto plazo**.
- $x_t, B, C, D_t$  y  $u_t$  son como en VAR.

Por ejemplo, un VECM con tres variables endógenas ( $K = 3$ ), dos retardos ( $p = 2$ ) y dos relaciones cointegradoras ( $r = 2$ ):

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t$$

donde:

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

$$e c_{1,t-1} = \beta_{11} y_{1,t-1} + \beta_{21} y_{2,t-1} + \beta_{31} y_{3,t-1}$$

$$e c_{2,t-1} = \beta_{12} y_{1,t-1} + \beta_{22} y_{2,t-1} + \beta_{32} y_{3,t-1}$$

y

$$\Gamma_1 \Delta y_{t-1} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-1} \\ \Delta y_{2,t-1} \\ \Delta y_{3,t-1} \end{bmatrix} \quad u_t = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Visualizando las ecuaciones por separado:

$$\Delta y_{1t} = \alpha_{11} e c_{1,t-1} + \alpha_{12} e c_{2,t-1} + \gamma_{11} \Delta y_{1,t-1} + \gamma_{12} \Delta y_{2,t-1} + \gamma_{13} \Delta y_{3,t-1} + u_{1t}$$

$$\Delta y_{2t} = \alpha_{21} e c_{1,t-1} + \alpha_{22} e c_{2,t-1} + \gamma_{21} \Delta y_{1,t-1} + \gamma_{22} \Delta y_{2,t-1} + \gamma_{23} \Delta y_{3,t-1} + u_{2t}$$

$$\Delta y_{3t} = \alpha_{31} e c_{1,t-1} + \alpha_{32} e c_{2,t-1} + \gamma_{31} \Delta y_{1,t-1} + \gamma_{32} \Delta y_{2,t-1} + \gamma_{33} \Delta y_{3,t-1} + u_{3t}$$