MULTICOLINEALIDAD: EXTENSIONES

I INTRODUCCION

El problema de la multicolinealidad surge cuando las variables explicativas de un modelo econométrico presentan un grado de correlación tal, que es difícil identificar de manera precisa el efecto individual de cada una de ellas sobre la variable dependiente o explicada.

Para fijar ideas, consideremos el siguiente modelo lineal:

$$Y_t = \beta_1 + \beta_2 X_{t2} + \dots + \beta_k X_{tk} + u_t \quad (1)$$

y supongamos que el siguiente modelo de regresión tiene un R^2 alto:

$$X_{it} = \delta_1 + \delta_2 X_{t2} + \dots + \delta_{i-1} X_{t,i-1} + \delta_{i+1} X_{t,i+1} + \dots + \delta_{k-1} X_{tk} + v_t \quad (2)$$

En dicho caso, X_{it} es aproximadamente una combinación lineal de las variables explicativas restantes. Ello implica que la matriz $\mathbf{X}'\mathbf{X}$ será **aproximadamente** singular.

Siempre y cuando $\mathbf{X}'\mathbf{X}$ NO sea singular, el estimador MICO será único y NO perderá su condición de MELI. Sin embargo, presentará los siguientes problemas:

a) La solución del sistema de ecuaciones normales estará mal definida:

$$(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \quad (3)$$

Si bien el estimador MICO, $\hat{\boldsymbol{\beta}}$, será único, habrá un número de vectores $\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \dots$, que serán una **solución aproximada** a (3).

Como consecuencia, pequeñas variaciones en el tamaño de la muestra—que introducirán sólo leves cambios en $\mathbf{X}'\mathbf{X}$ y $\mathbf{X}'\mathbf{Y}$ —se traducirán en cambios importantes en la magnitud de los componentes del vector $\hat{\boldsymbol{\beta}}$.

b) La casi singularidad de la matriz $\mathbf{X}'\mathbf{X}$ se reflejará en el valor de su determinante, el cual será cercano a cero. Como consecuencia, las varianzas

de los estimadores MICO, provenientes de la matriz $\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$, serán 'grandes'. Ello incrementará el ancho de los intervalos de confianza para el contraste de hipótesis lineales. En consecuencia, tenderemos a aceptar la hipótesis nula. (Es decir, caerá el **poder** del test: la probabilidad de rechazar H_0 cuando ésta es falsa).

Ejemplo

En el modelo lineal simple $Y_t = \beta_1 + \beta_2 X_t + u_t$, la existencia de multicolinealidad equivaldría a que X_t fuera aproximadamente constante. En este caso, un aumento en la variación de X_t contribuiría a una mayor precisión de los estimadores MICO de β_1 y β_2 ♦

II MULTICOLINEALIDAD EXACTA Y APROXIMADA

La multicolinealidad puede tomar dos formas:

- Multicolinealidad Exacta. Ocurre cuando una de las variables explicativas es una combinación lineal determinística de todas las demás (o de un subconjunto de ellas). Es decir, la matriz $\mathbf{X}'\mathbf{X}$ es singular y, por lo tanto, existen infinitas soluciones a las ecuaciones normales.
- Multicolinealidad Aproximada: Ocurre cuando una de las variables explicativas es aproximadamente una combinación lineal de las restantes (o de un subconjunto de ellas). Este es el caso de la ecuación (2).

En la práctica, la **multicolinealidad exacta** es fácilmente detectable porque $\mathbf{X}'\mathbf{X}$ es singular. Por ejemplo, un conjunto de las variables explicativas satisfacen una identidad contable.

En contraste, la **multicolinealidad aproximada** es más difícil de detectar. No obstante, ésta puede traducirse (aunque no necesariamente) en:

- ✓ Un R^2 alto pero con coeficientes de baja significancia estadística.
- ✓ Estimadores MICO con signos y/o magnitudes implausibles, producto de la poca precisión con la cual son estimados.

- ✓ Estimadores MICO muy sensibles a pequeños cambios en el tamaño de la muestra.

2.1 Estimación Bajo Multicolinealidad Exacta

Si bien las ecuaciones normales no tienen una solución única, existen determinadas combinaciones lineales de los parámetros que pueden ser estimadas con toda precisión. Para ilustrar este punto, veamos el siguiente ejemplo (Novales).

Se tiene el modelo:

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + u_t \quad t=1, 2, \dots, T$$

Se sabe que existe una relación lineal exacta entre X_2 y X_3 para todo $t=1, 2, \dots, T$: $X_{t3}=\lambda X_{t2}$. Por ello, el modelo anterior es equivalente a:

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 (\lambda X_{t2}) + u_t \quad t=1, 2, \dots, T$$

Si expresamos el modelo en desvíos con respecto a las medias muestrales, se tiene que:

$$\mathbf{x}'\mathbf{x} = \sum_{t=1}^T x_{t2}^2 \begin{pmatrix} 1 & \lambda \\ \lambda & \lambda^2 \end{pmatrix}$$

donde $x_{t2} \equiv X_{t2} - \bar{X}_2$, $y_t \equiv Y_t - \bar{Y}$.

Claramente, la matriz $\mathbf{x}'\mathbf{x}$ tiene rango igual a 1. Ello implica que es singular. Entonces, el sistema de ecuaciones normales se reduce a una sola ecuación:

$$\sum_{t=1}^T x_{t2} y_t = \hat{\beta}_2 + \lambda \hat{\beta}_3 \sum_{t=1}^T x_{t2}^2$$

la cual tiene infinitas soluciones para $\hat{\beta}_2$ y $\hat{\beta}_3$. Sin embargo, existe una única solución para la combinación lineal $\beta_2 + \lambda\beta_3$, dada por:

$$\hat{\beta}_2 + \lambda \hat{\beta}_3 = \frac{\sum_{t=1}^T X_{t2} Y_t}{\sum_{t=1}^T X_{t2}^2}$$

- ¿Por qué? Porque el modelo se reduce a:

$$Y_t = \beta_1 + (\beta_2 + \beta_3 \lambda) X_{t2} + u_t \quad t=1, 2, \dots, T$$

Entonces, podemos estimar sin problemas la combinación lineal de los parámetros β_2 y β_3 , $\beta_2 + \lambda\beta_3$.

2.2 Detección y Estimación Bajo Multicolinealidad Aproximada

Señalamos anteriormente que es posible que la multicolinealidad aproximada se materialice en un R^2 alto pero con coeficientes de baja significancia estadística, en estimadores con signos y/o magnitudes implausibles que pueden ser, además, altamente sensibles al tamaño de la muestra.

Sin embargo, la multicolinealidad aproximada NO necesariamente conlleva a los síntomas anteriormente señalados. Por ello, se han diseñado mecanismos formales para su detección.

2.2.1 Método Formales de Detección

a) Correlación entre Variables Explicativas

Particionemos la matriz $\mathbf{X}$ como sigue:

$$\mathbf{X} = (\mathbf{x}_i \quad \mathbf{X}_i) \quad (4)$$

donde $\mathbf{x}_i$ es el vector columna de observaciones de la i -ésima variable explicativa, y $\mathbf{X}_i$ es la matriz $T \times (k-1)$ de observaciones de las variables restantes. Entonces, la matriz $\mathbf{X}'\mathbf{X}$ puede escribirse como:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{x}_i' \\ \mathbf{X}_i' \end{pmatrix} (\mathbf{x}_i \quad \mathbf{X}_i) = \begin{pmatrix} \mathbf{x}_i' \mathbf{x}_i & \mathbf{x}_i' \mathbf{X}_i \\ \mathbf{X}_i' \mathbf{x}_i & \mathbf{X}_i' \mathbf{X}_i \end{pmatrix} \quad (5)$$

Si utilizamos la propiedad de la inversa de una matriz particionada (véase apéndice), tenemos que el elemento (1, 1) de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ es:

$$(\mathbf{x}_i' \mathbf{x}_i - \mathbf{x}_i' \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{x}_i)^{-1} = (\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i)^{-1}$$

donde $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i'$.

Por lo tanto, la varianza de $\hat{\beta}_i$ estará dada por:

$$\text{Var}(\hat{\beta}_i) = \sigma^2 (\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i)^{-1} = \frac{\sigma^2}{\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i} \quad (6)$$

dado que $\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i$ es un escalar.

Notemos que $\mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i$ se puede descomponer en lo siguiente:

$$\begin{aligned} \mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i &= \mathbf{x}_i' \mathbf{x}_i - \mathbf{x}_i' \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{x}_i \\ &= \mathbf{x}_i' \mathbf{x}_i - \mathbf{x}_i' \hat{\mathbf{x}}_i \\ &= \mathbf{x}_i' \mathbf{x}_i - \hat{\mathbf{x}}_i' \hat{\mathbf{x}}_i \end{aligned}$$

donde $\hat{\mathbf{x}}_i = \mathbf{X}_i (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{x}_i$ es el vector de valores **predichos** de $\mathbf{x}_i$ resultante de correr una regresión de $\mathbf{x}_i$ en los regresores restantes del modelo, $\mathbf{X}_i$.

Pero, de lo anterior,

$$\begin{aligned} \mathbf{x}_i' \mathbf{M}_i \mathbf{x}_i &= \sum_{t=1}^T x_{ti}^2 - \sum_{t=1}^T \hat{x}_{ti}^2 \\ &= \sum_{t=1}^T x_{ti}^2 - \sum_{t=1}^T \hat{x}_{ti}^2 + T\bar{x}_i^2 - T\bar{x}_i^2 = \sum_{t=1}^T (x_{ti} - \bar{x}_i)^2 - \sum_{t=1}^T (\hat{x}_{ti} - \bar{x}_i)^2 \\ &= \sum_{t=1}^T (x_{ti} - \bar{x}_i)^2 (1 - R_i^2) \end{aligned}$$

donde $R_i^2 = \frac{\sum_{t=1}^T (\hat{x}_{ti} - \bar{x})^2}{\sum_{t=1}^T (x_{ti} - \bar{x})^2}$. Con lo cual,

$$\text{Var}(\hat{\beta}_i) = \frac{\sigma^2}{\sum_{t=1}^T (x_{ti} - \bar{x})^2 (1 - R_i^2)} \quad (7)$$

Por lo tanto, si R_i^2 es alto, la varianza de $\hat{\beta}_i$ será ‘grande’. De ello, un método de detección de la multicolinealidad aproximada consiste en estimar regresiones de **cada una** de las variables explicativas en las restantes. Un R^2 alto en alguna de estas regresiones indicaría que alguna de las variables explicativas está causando problemas.

El valor más pequeño de la expresión (7) se obtiene cuando $R_i^2=0$. Sea $\text{Var}(\hat{\beta}_i^0)$ el valor de la varianza de $\hat{\beta}_i$ cuando $R_i^2=0$. Definamos τ como:

$$\tau = \frac{\text{Var}(\hat{\beta}_i)}{\text{Var}(\hat{\beta}_i^0)} = \frac{1}{1 - R_i^2} \quad (8)$$

Entonces τ es una medida de qué tan severa es la dependencia lineal de x_i con respecto a $\mathbf{X}_i$. Por ejemplo, si $R_i^2=0.5$, $\tau=2$; si $R_i^2=0.999$, $\tau=1000$. En la práctica, un $\tau > 10$ ($R_i^2 > 0.9$), se considera problemático.

b) Tamaño de la Matriz $\mathbf{X}'\mathbf{X}$

Una medida de qué tan problemática es la multicolinealidad aproximada es el número de condición de la matriz. En particular, para una matriz cuadrada, $\mathbf{A}$, éste se define como:

$$\gamma = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \quad (9)$$

Para matrices no cuadradas, tal como $\mathbf{X}_{n \times k}$, calculamos el número de condición de $\mathbf{X}'\mathbf{X}$. Debido a que los valores propios son sensibles a la escala de las variables, se recomienda dividir cada columna, $\mathbf{x}_i$, por su norma, $\sqrt{\mathbf{x}_i' \mathbf{x}_i}$, a fin de que tenga un largo igual a 1.

El hecho de que el valor propio más pequeño sea cercano a cero en relación al más grande significa que la matriz es cuasi singular. Una matriz con un número de condición 'grande' (>20) es difícil de invertir de manera exacta.

Definición: El **error cuadrático medio** (ECM) de un estimador se define como:

$$\begin{aligned} \text{ECM} &= E(\hat{\beta} - \beta)^2 = \text{Var}(\hat{\beta}) + \text{Sesgo}(\hat{\beta})^2 \quad \text{si } \hat{\beta} \text{ es un escalar} \\ &= \text{Var}(\hat{\beta}) + \text{Sesgo}(\hat{\beta}) \text{Sesgo}(\hat{\beta})' \quad \text{si } \hat{\beta} \text{ es un vector} \end{aligned}$$

2.2.2 Posibles Soluciones para la Multicolinealidad Aproximada

a) Pre-estimación: Esto hace alusión al caso en que se cuenta con información a priori que permite reducir la colinealidad entre algún par o conjunto de variables. Supongamos el siguiente ejemplo para ilustrar este punto:

$$\ln(Y_t) = \beta_1 + \beta_2 \ln(K_t) + \beta_3 \ln(L_t) + u_t$$

donde Y_t = nivel de producción, K_t = factor capital, L_t = factor trabajo.

Supongamos que K_t y L_t son altamente colineales. Sin embargo, se sabe, por estudios previos, que la función de producción presenta rendimientos constantes a escala. Esto es, $\beta_2 + \beta_3 = 1$. Por lo tanto, el modelo puede ser simplificado a:

$$\ln(Y_t/L_t) = \beta_1 + \beta_2 \ln(K_t/L_t) + u_t$$

Desgraciadamente, en la práctica no es usual contar con este tipo de información.

b) Estimador “Cresta” (*Ridge*)

Una solución que se ha propuesto para reducir el grado de colinealidad entre las columnas de $\mathbf{X}$ es el estimador cresta:

$$\hat{\beta}_r = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{Y} \quad (10)$$

donde $\mathbf{D}$ es una matriz diagonal que contiene los elementos de la diagonal de $\mathbf{X}'\mathbf{X}$ y r es un escalar escogido arbitrariamente. En la práctica, r se escoge de modo tal que $\hat{\beta}_r$ sea estable frente a pequeñas variaciones en éste. (Se sugiere partir de un valor de $r \approx 0.01$).

Aunque el estimador cresta es sesgado, tiene una matriz varianza-covarianza menor a la de MICO:

$$E(\hat{\beta}_r) = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{X}\beta \quad \text{Var}(\hat{\beta}_r) = (\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1} \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + r\mathbf{D})^{-1}$$

Ello implica que $\hat{\beta}_r$ puede ser preferible a MICO, términos de ECM.

c) Método de Componentes Principales

La idea detrás del método de componentes principales radica en extraer de la matriz $\mathbf{X}$ un número de variables que expliquen la mayor parte de la variación en $\mathbf{X}$. (Esta última se define como $\text{tr}(\mathbf{X}'\mathbf{X})$, $\text{tr} \equiv \text{traza}$).

❖ La pregunta relevante es: ¿qué combinaciones lineales proporcionan el mejor ajuste para todas las columnas de $\mathbf{X}$?

Definamos $\mathbf{Z}$ como la matriz que reúne a estas nuevas variables. Sea, en particular, $\mathbf{z}_1$, vector $T \times 1$, el primer componente de dicha matriz:

$$z_{t1} = c_{11}x_{t1} + c_{21}x_{t2} + \dots + c_{k1}x_{tk} \quad t=1, 2, \dots, T \quad (11)$$

En notación matricial:

$$\mathbf{z}_1 = \mathbf{X} \mathbf{c}_1 \quad (12)$$

Normalizamos $\mathbf{z}_1$, dado que otra combinación lineal de $\mathbf{c}$ conduciría al mismo R^2 en una regresión de la i -ava columna de $\mathbf{X}$, $\mathbf{x}_i$, en $\mathbf{z}_1$:

$$\mathbf{z}_1' \mathbf{z}_1 = 1 \quad (13)$$

Supongamos que corremos una regresión de $\mathbf{x}_i$ en $\mathbf{z}_1$, $i=1, 2, \dots, k$. La suma de errores al cuadrado de dicha regresión es:

$$\hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i = \mathbf{x}_i' (\mathbf{I}_T - \mathbf{z}_1 (\mathbf{z}_1' \mathbf{z}_1)^{-1} \mathbf{z}_1') \mathbf{x}_i = \mathbf{x}_i' (\mathbf{I}_T - \mathbf{z}_1 \mathbf{z}_1') \mathbf{x}_i \quad (14)$$

dado que $\mathbf{z}_1' \mathbf{z}_1 = 1$.

Si tomamos todos los vectores $\mathbf{x}_i$ simultáneamente, buscamos **minimizar**:

$$\begin{aligned} \sum_{i=1}^k \hat{\mathbf{u}}_i' \hat{\mathbf{u}}_i &= \sum_{i=1}^k \mathbf{x}_i' (\mathbf{I}_T - \mathbf{z}_1 \mathbf{z}_1') \mathbf{x}_i = \text{tr}(\mathbf{X}' (\mathbf{I}_T - \mathbf{z}_1 \mathbf{z}_1') \mathbf{X}) \\ &= \text{tr}(\mathbf{X}' \mathbf{X}) - \text{tr}(\mathbf{X}' \mathbf{z}_1 \mathbf{z}_1' \mathbf{X}) \end{aligned}$$

sujeto a $\mathbf{z}_1' \mathbf{z}_1 = 1$.

Dado que $\text{tr}(\mathbf{X}' \mathbf{X})$ no involucra al vector $\mathbf{c}_1$, minimizar la expresión anterior es equivalente a **maximizar**:

$$L = \text{tr}(\mathbf{X}' \mathbf{z}_1 \mathbf{z}_1' \mathbf{X}) + \lambda_1 (1 - \mathbf{z}_1' \mathbf{z}_1) \quad (15)$$

Dado que $\mathbf{z}_1 = \mathbf{X} \mathbf{c}_1$ y por las propiedades de traza, el lagrangeano anterior se reduce a:

$$L = \mathbf{c}_1' (\mathbf{X}' \mathbf{X})^2 \mathbf{c}_1 + \lambda_1 (1 - \mathbf{c}_1' (\mathbf{X}' \mathbf{X}) \mathbf{c}_1) \quad (16)$$

con condiciones de primer orden:

$$\frac{\partial L}{\partial \mathbf{c}_1} = 2(\mathbf{X}' \mathbf{X})^2 \mathbf{c}_1 - 2\lambda_1 (\mathbf{X}' \mathbf{X}) \mathbf{c}_1 = \mathbf{0} \quad (i)$$

$$\frac{\partial L}{\partial \lambda} = 1 - \mathbf{c}_1' (\mathbf{X}' \mathbf{X}) \mathbf{c}_1 = 0 \quad (ii)$$

La condición (i) puede ser re-escrita como:

$$(\mathbf{X}'\mathbf{X}) \mathbf{c}_1 = \lambda_1 \mathbf{c}_1 \quad (i')$$

lo cual indica que $\mathbf{c}_1$ es un vector propio de $\mathbf{X}'\mathbf{X}$.

Ahora, notemos que al pre-multiplicar ambos lados de (i') por $\mathbf{c}_1'(\mathbf{X}'\mathbf{X})$, llegamos a:

$$\mathbf{c}_1' (\mathbf{X}'\mathbf{X})^2 \mathbf{c}_1 = \lambda_1 \mathbf{c}_1' (\mathbf{X}'\mathbf{X}) \mathbf{c}_1 = \lambda_1 \quad (17)$$

Es decir, si queremos maximizar nuestra función objetivo $\mathbf{c}_1' (\mathbf{X}'\mathbf{X})^2 \mathbf{c}_1$, entonces debemos escoger el vector propio asociado al valor propio mayor. (La matriz $\mathbf{X}'\mathbf{X}$ es positiva definida, por lo cual sus valores propios son positivos). Por lo tanto, $\mathbf{z}_1$ es el primer componente principal de $\mathbf{X}$.

Definamos ahora $\mathbf{z}_2 = \mathbf{X}\mathbf{c}_2$. Deseamos maximizar $\mathbf{c}_2' (\mathbf{X}'\mathbf{X})^2 \mathbf{c}_2$, sujeto a que $\mathbf{c}_2' \mathbf{c}_2 = 1$ y $\mathbf{c}_1' \mathbf{c}_2 = 0$ (esto es, $\mathbf{c}_1$ y $\mathbf{c}_2$ no están correlacionados). El lagrangeano se reduce a:

$$L = \mathbf{c}_2' (\mathbf{X}'\mathbf{X})^2 \mathbf{c}_2 + \lambda_2 (1 - \mathbf{c}_2' (\mathbf{X}'\mathbf{X}) \mathbf{c}_2) - \mu \mathbf{c}_1' \mathbf{c}_2 \quad (18)$$

Después de hacer el álgebra correspondiente, se llega a que:

$$(\mathbf{X}'\mathbf{X}) \mathbf{c}_2 = \lambda_2 \mathbf{c}_2$$

Ello implica que λ_2 corresponde al segundo valor propio más grande y, por lo tanto, $\mathbf{z}_2$ es el segundo componente principal de $\mathbf{X}$. Podemos continuar análogamente para cada una de los k valores propios de $\mathbf{X}'\mathbf{X}$, y reunirlos en la matriz $\mathbf{C}$, donde $\mathbf{C} = (\mathbf{c}_1 \mathbf{c}_2 \dots \mathbf{c}_k)$. Con ello,

$$\mathbf{Z} = \mathbf{XC} \quad \mathbf{Z}'\mathbf{Z} = \mathbf{C}'\mathbf{X}'\mathbf{XC} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_k \end{pmatrix}$$

Si el rango de $\mathbf{X}$ es $r < k$, $k-r$ valores propios son iguales a cero. Aún si $\mathbf{X}$ no es singular, pero casi, algunos de los valores propios serán cercanos a cero. Por ello, un número menor de componentes principales explicará la mayor parte de la variación en $\mathbf{X}$. La idea, entonces, es utilizar un número $L < k$ de componentes principales, de modo tal que $\sum_{i=1}^L \lambda_i / \sum_{i=1}^k \lambda_i \approx 1$.

Sea $\mathbf{C}_L$ la matriz que reúne a los vectores propios asociados a los L valores propios más grandes, donde $\mathbf{Z}_L = \mathbf{X}\mathbf{C}_L$. El paso siguiente consiste en correr una regresión de $\mathbf{Y}$ en $\mathbf{Z}_L$, para obtener el siguiente estimador:

$$\hat{\mathbf{a}} = (\mathbf{Z}_L' \mathbf{Z}_L)^{-1} \mathbf{Z}_L' \mathbf{Y} = \Lambda_L^{-1} \mathbf{C}_L' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}_{\text{MICO}} = \mathbf{C}_L' \hat{\boldsymbol{\beta}}_{\text{MICO}} \quad (19)$$

dado que $(\mathbf{X}'\mathbf{X})\mathbf{C}_L = \mathbf{C}_L \Lambda_L \Leftrightarrow \mathbf{C}_L'(\mathbf{X}'\mathbf{X}) = \Lambda_L \mathbf{C}_L'$.

Entonces $\hat{\mathbf{Y}} = \mathbf{Z}_L \hat{\mathbf{a}} = \mathbf{X}\mathbf{C}_L \hat{\mathbf{a}} \equiv \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{CP}}$

donde el **estimador de componentes principales** viene dado por:

$$\hat{\boldsymbol{\beta}}_{\text{CP}} = \mathbf{C}_L \mathbf{C}_L' \hat{\boldsymbol{\beta}}_{\text{MICO}} \quad (20)$$

El estimador de componentes principales es sesgado, pero tiene una varianza menor a la de MICO:

$$\text{Var}(\hat{\boldsymbol{\beta}}_{\text{CP}}) = \sigma^2 \mathbf{C}_L \mathbf{C}_L' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}_L \mathbf{C}_L' \quad (21)$$

Por ello, puede ser preferible a éste, según el criterio de ECM. Sin embargo, el estimador de componentes principales no está libre de problemas:

- ✓ Difícil de interpretar .(Es una mezcla de los coeficientes originales).
- ✓ Los componentes principales NO son escogidos en base a ninguna relación entre $\mathbf{X}$ e $\mathbf{Y}$.
- ✓ Como sabemos, la magnitud de los valores propios es sensible a la escala de las $\mathbf{X}$'s, por lo cual se recomienda normalizar cada columna. Sin embargo, ello afecta la interpretación de los resultados.

d) Exclusión de Variables

Supongamos que tenemos el siguiente modelo econométrico:

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + u_t \quad (22)$$

y que X_{t2} y X_{t3} están altamente correlacionadas. Una posibilidad es eliminar una de las variables del modelo, digamos X_{t3} . Sin embargo, ello generará sesgo en el estimador MICO de β_2 proveniente del modelo:

$$Y_t = \beta_1 + \beta_2 X_{t2} + v_t \quad (23)$$

a menos que X_2 sea **ortogonal** a X_3 :

$$E(\tilde{\beta}_2) = \beta_2 + \beta_3 \frac{\sum_{t=1}^T X_{t2} X_{t3}}{\sum_{t=1}^T X_{t2}^2} \quad (24)$$

donde las letras minúsculas denotan variables en desvíos. Por otra parte,

$$\text{Var}(\tilde{\beta}_2) = \frac{\sigma^2}{\sum_{t=1}^T X_{t2}^2} \quad \text{Var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{t=1}^T X_{t2}^2 (1 - r_{23}^2)}$$

donde $\hat{\beta}_2$ es el estimador MICO proveniente del modelo (22) y

$r_{23} = \frac{\sum_{t=1}^T X_{t2} X_{t3}}{\sqrt{\sum_{t=1}^T X_{t2}^2 \sum_{t=1}^T X_{t3}^2}}$, coeficiente de correlación muestral entre X_2 y X_3 .

Claramente, $\text{Var}(\tilde{\beta}_2) < \text{Var}(\hat{\beta}_2)$. Por lo tanto, los estimadores deben ser comparados en términos de ECM:

$$\boxed{\frac{\text{ECM}(\tilde{\beta}_2)}{\text{ECM}(\hat{\beta}_2)} = 1 + r_{23}^2 (t_{\beta_3}^2 - 1)} \quad (25)$$

$$\text{donde } t_{\beta_3}^2 = \frac{\beta_3^2}{\text{Var}(\hat{\beta}_3)} = \frac{\beta_3^2}{\sigma^2} \sum_{t=1}^T x_{t3}^2 (1 - r_{23}^2).$$

Por lo tanto, el estimador sesgado es preferible si $|t| < 1$. El estimador escogido mediante esta estrategia se denomina estimador de **pre-test**, puesto que su elección se sustenta en un test t de una regresión previa. El problema radica en que t no es conocido, puesto que ni β_3 ni σ^2 lo son. Por lo tanto, la decisión debe basarse en $\hat{t}$. Defínase

$$\begin{aligned} \bar{\beta} &= \tilde{\beta}_2 & \text{si } \hat{t}^2 &= \frac{\hat{\beta}_3^2}{\hat{\sigma}^2} \sum_{t=1}^T x_{t3}^2 (1 - r_{23}^2) < 1 \\ &= \hat{\beta}_2 & \text{si } \hat{t}^2 &\geq 1 \end{aligned} \quad (26)$$

Se ha argüido que $\hat{\beta}_2$ es preferible a $\bar{\beta}$, a menos que se crea firmemente, a priori, que $t^2 < 1$.

e) Imposición de Restricciones Lineales

Una forma de reducir la colinealidad aproximada es mediante la imposición de restricciones lineales sobre un conjunto de parámetros del modelo. Sabemos que, bajo los supuestos del modelo lineal general, el estadígrafo para contrastar un conjunto de J restricciones lineales, $H_0: \mathbf{R}\beta = \mathbf{q}$, viene dado por:

$$F = \frac{(\mathbf{R}\hat{\beta} - \mathbf{q})' (\hat{\sigma}^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q})}{J} = \frac{(\hat{\mathbf{u}}_r' \hat{\mathbf{u}}_r - \hat{\mathbf{u}}_{nr}' \hat{\mathbf{u}}_{nr}) / J}{\hat{\mathbf{u}}_{nr}' \hat{\mathbf{u}}_{nr} / (T - k)}$$

siendo la regla de decisión: rechace H_0 si el F calculado supera a $F(J, T-k)$ al $(1-\alpha)$ % de confianza.

- ❖ **¿Qué pasa si las restricciones son rechazadas estadísticamente?** En dicho caso, el estimador restringido (es decir, aquel que incorpora las restricciones) es **sesgado**. Sin embargo, es más **eficiente** que MICO. Por lo tanto, puede ser preferible en términos de ECM.

Para ello, recordemos que la solución al problema de optimización (véase capítulo 7, Greene):

$$\text{Min}_{\tilde{\beta}} S(\tilde{\beta}) = (\mathbf{Y} - \mathbf{X}\tilde{\beta})'(\mathbf{Y} - \mathbf{X}\tilde{\beta}) \quad \text{s.a.} \quad \mathbf{R}\tilde{\beta} = \mathbf{q}$$

da origen al estimador MICO restringido:

$$\begin{aligned} \hat{\beta}_r &= \hat{\beta}_{nr} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{R}\hat{\beta}_{nr} - \mathbf{q}) \\ \text{Var}(\hat{\beta}_r) &= \text{Var}(\hat{\beta}_{nr}) - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \quad (27)$$

donde $\hat{\beta}_{nr} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\text{Var}(\hat{\beta}_{nr}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

➤ ¿En qué circunstancias, entonces, será $\hat{\beta}_r$ preferible a $\hat{\beta}_{nr}$, en términos de ECM? Este tópico fue estudiado por Wallace y Toro-Vizcarrondo. Ellos concluyeron que si H_0 es rechazada, el estadígrafo F tiene una distribución F no central, $F(J, T-k, 1/2)$, donde 1/2 es el parámetro de no centralidad. Los hallazgos de W-TV se pueden resumir como sigue:

Si el F computado $> F_{1-\alpha}(J, T-k, 1/2)$, $\hat{\beta}_{nr}$ es preferible a $\hat{\beta}_r$ en términos de ECM. De lo contrario, $\hat{\beta}_r$ es preferible a $\hat{\beta}_{nr}$, ya que:

$$\mathbf{c}'\text{ECM}(\hat{\beta}_r)\mathbf{c} \leq \mathbf{c}'\text{ECM}(\hat{\beta}_{nr})\mathbf{c} \quad \forall \mathbf{c}_{k \times 1} \neq \mathbf{0} \quad (28)$$

APENDICE

1) **Matrices Particionadas:** El particionamiento de una matriz consiste en agrupar los elementos de ésta en sub-matrices. Por ejemplo, supongamos una matriz $A_{3 \times 3}$ que particionamos de la siguiente forma:

$$A = \begin{pmatrix} 1 & 2 & 7 \\ 9 & 4 & 3 \\ 1 & 5 & 8 \end{pmatrix}$$

Esto es, $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$, con $\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 \\ 9 & 4 \end{pmatrix}$, $\mathbf{A}_{12} = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$, $\mathbf{A}_{21} = (1 \ 5)$, $\mathbf{A}_{22} = (8)$.

a) Suma y Multiplicación de Matrices Particionadas

Sea $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ y $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$, entonces:

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}$$

con $\mathbf{A}_{ij}$ y $\mathbf{B}_{ij}$, matrices conformables para la suma y multiplicación.

b) Transpuesta

$$\mathbf{A}' = \begin{pmatrix} \mathbf{A}_{11}' & \mathbf{A}_{21}' \\ \mathbf{A}_{12}' & \mathbf{A}_{22}' \end{pmatrix}$$

Por ejemplo, para el caso particular $\mathbf{A} = (\mathbf{A}_1 \ \mathbf{A}_2)$, se tiene que $\mathbf{A}' = \begin{pmatrix} \mathbf{A}_1' \\ \mathbf{A}_2' \end{pmatrix}$.

c) Inversa: para una matriz particionada 2x2 se tiene:

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{F}_1 & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_2 \end{pmatrix}$$

donde $\mathbf{F}_1 = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$, $\mathbf{F}_2 = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$

Para el caso particular de una matriz diagonal en bloque se tiene:

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{pmatrix}$$

c) Determinante

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{22} \parallel \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}| = |\mathbf{A}_{11} \parallel \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}|$$

2) Valores y Vectores Propios de una Matriz

Supongamos una matriz cuadrada, $\mathbf{A}$, $n \times n$. Las soluciones al siguiente conjunto de ecuaciones:

$$\mathbf{A}\mathbf{c} = \lambda\mathbf{c} \quad (1)$$

se denominan, respectivamente, vector propio, $\mathbf{c}_{n \times 1}$, y valor propio, λ (escalar).

Dado que el vector $k\mathbf{c}$ también satisface la ecuación anterior, donde k constante cualquiera, el vector $\mathbf{c}$ es normalizado:

$$\mathbf{c}'\mathbf{c} = 1 \quad (2)$$

La ecuación (1) puede ser re-escrita como sigue:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0} \quad (3)$$

A fin de que el sistema homogéneo (3) tenga una solución distinta del vector $\mathbf{0}$, tiene que ser el caso que la matriz $\mathbf{A} - \lambda\mathbf{I}$ sea singular:

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \quad (4)$$

La ecuación (4), llamada '**ecuación característica**' es un polinomio de orden n en λ . Los valores propios pueden repetirse, y no son necesariamente números reales. (Únicamente si la matriz $\mathbf{A}$ es simétrica, los valores propios son todos reales).

Ejemplo (Greene)

Sea $\mathbf{A} = \begin{pmatrix} 5 & 1 \\ 2 & 4 \end{pmatrix}$. La ecuación característica viene dada por:

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 5 - \lambda & 1 \\ 2 & 4 - \lambda \end{vmatrix} = \lambda^2 - 9\lambda + 18 = (\lambda - 6)(\lambda - 3) = 0$$

cuyas soluciones son $\lambda=6$ y $\lambda=3$.

Los vectores característicos pueden ser encontrados a partir de (3):

$$\begin{pmatrix} 5 - \lambda & 1 \\ 2 & 4 - \lambda \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Para $\lambda=6$, se tiene el par de ecuaciones (linealmente dependientes, por construcción):

$$-c_1 + c_2 = 0 \quad 2c_1 - 2c_2 = 0 \quad \Leftrightarrow \quad c_1 = c_2 \quad (\text{i})$$

Para $\lambda=3$, en tanto, se tiene el par de ecuaciones:

$$2c_1 + c_2 = 0 \quad 2c_1 + c_2 = 0 \quad \Leftrightarrow \quad c_1 = (-1/2) c_2 \quad (\text{ii})$$

A fin de determinar los elementos de cada vector, se impone la normalización $\mathbf{c}'\mathbf{c}=1$ ($\Leftrightarrow c_1^2 + c_2^2 = 1$). Con ello,

$$\lambda=6, \mathbf{c} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}; \lambda=3, \mathbf{c} = \begin{pmatrix} 1/5 \\ -2/5 \end{pmatrix}.$$

2.1 Resultados Importantes

a) Sea $\mathbf{A}$ una matriz simétrica $n \times n$, entonces la matriz $\mathbf{A}$ es igual:

$$\mathbf{A} = \mathbf{C}' \mathbf{\Lambda} \mathbf{C} \quad (5)$$

donde la matriz $\mathbf{C}=(\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_n)$ reúne a los n vectores propios, y la matriz $\mathbf{\Lambda}$ reúne a los n valores propios (en el mismo orden) en una matriz diagonal:

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

La expresión (5) se denomina “**descomposición espectral**” de $\mathbf{A}$.

Para una matriz simétrica, los vectores propios son ortogonales. Esto es,

$$\mathbf{c}_i' \mathbf{c}_j = 0 \quad \forall i \neq j$$

con lo cual, $\mathbf{C}\mathbf{C}' = \mathbf{C}'\mathbf{C} = \mathbf{I}$, $\mathbf{C}' = \mathbf{C}^{-1}$.

- b) El **rango de una matriz simétrica** es igual al número de valores propios distintos de cero.
- c) La **traza de una matriz simétrica** (cuadrada) es igual a la suma de sus valores propios.
- d) El **determinante** de una **matriz simétrica** (cuadrada) es igual al producto de sus valores propios. En particular, si uno de los valores propios es igual a cero, la matriz es singular.
- e) Para cualquier matriz simétrica no singular, $\mathbf{A}$, se tiene que sus **potencias** vienen dada por:

$$\mathbf{A}^k = \mathbf{C}\mathbf{\Lambda}^k\mathbf{C}' \quad k = \dots, -2, -1, 0, 1, 2, \dots$$

Si, además, $\mathbf{A}$ es **positiva definida** (lo cual implica que todos sus valores propios son positivos):

$$\mathbf{A}^r = \mathbf{C}\mathbf{\Lambda}^r\mathbf{C}' \quad r, \text{ número real cualquiera.}$$

Para mayores detalles, véase Greene, capítulo 3 o Johnston, capítulo 5 ♦