

Tema 2. Multicolinealidad

Econometría II. 3º o LADE

1 Introducción

Uno de los supuestos en los que descansa el modelo lineal general es que el rango de la matrix de regresores (X) es igual a k , siendo k el número de regresores empleados en la especificación. Esto equivale a decir que la matrix $X'X$, que es una matrix cuadrada de orden k , es de rango completo. Por tanto, existe la matrix inversa $(X'X)^{-1}$ y, en consecuencia, es posible obtener un unico vector de estimadores de los parametros de posicion, determinado por la conocida relacion $\hat{\beta} = (X'X)^{-1} X'Y$.

La paternidad del término multicolinealidad se le adjudica a Ragnar Frisch. Este concepto se asoció originariamente con aquellos casos en los que se incumple la condición de rango comentado con anterioridad y, por tanto, el rango de la matrix X es inferior a k , lo que supone que también el rango de $X'X$ es inferior a k . Por consiguiente, el determinante de esta matrix es igual a 0, no existiendo la inversa de esta matrix. Esto supone que no existe una única solución del sistema mínimo cuadrático ordinario $X'X\beta = X'Y$, por lo que el vector de estimadores mínimo cuadrático ordinario no es único.

Sin embargo, ni esta es una situación que aparezca frecuentemente en la modelización econométrica, ni tampoco es un problema especialmente grave, sobre todo si lo comparamos con los problemas que se producen en aquellos casos en los que el rango de la matrix $X'X$ está próximo a 0, aunque es distinto a esta cantidad. Por tanto, el concepto de multicolinealidad debe asociarse con aquellas situaciones en las que las variables explicativas están linealmente relacionadas entre sí.

En este tema vamos a estudiar los efectos que producen los distintos tipos de multicolinealidad, cómo se puede detectar la presencia de este problema y los pasos que podemos dar para solucionarlo cuando este se presenta.

2 Tipos de multicolinealidad

En este apartado vamos a estudiar el efecto que tiene el diferente grado de correlación entre las variables explicativas del modelo sobre la estimación mínimo cuadrática ordinaria. Vamos a distinguir tres casos distintos: ausencia de colinealidad, colinealidad perfecta y colinealidad aproximada. En los tres casos el planteamiento inicial es el mismo, por cuanto consideramos que la evolución de la variable sometida a estudio viene determinada por la siguiente relacion

$$Y = X\beta + u$$

2.1 Ausencia de colinealidad

En este caso, suponemos que las variables explicativas no tienen ningún tipo de correlación entre ellas. Esto ocurre cuando se cumple que $X_i X_j = 0, \forall i \neq j$. Por tanto, a partir de este resultado, podemos observar que el vector de estimadores mco puede derivarse como sigue:

$$\begin{aligned}
 \hat{\beta} &= (X'X)^{-1} X'Y = \begin{bmatrix} X'_1 X_1 & X'_1 X_2 & \cdots & X'_1 X_k \\ X'_2 X_1 & X'_2 X_2 & \cdots & X'_2 X_k \\ \cdots & \cdots & \cdots & \cdots \\ X'_k X_1 & X'_k X_2 & \cdots & X'_k X_k \end{bmatrix}^{-1} \begin{pmatrix} X'_1 Y \\ X'_2 Y \\ \cdots \\ X'_k Y \end{pmatrix} \\
 &= \begin{bmatrix} X'_1 X_1 & 0 & \cdots & 0 \\ 0 & X'_2 X_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & X'_k X_k \end{bmatrix}^{-1} \begin{pmatrix} X'_1 Y \\ X'_2 Y \\ \cdots \\ X'_k Y \end{pmatrix} \\
 &= \begin{bmatrix} (X'_1 X_1)^{-1} & 0 & \cdots & 0 \\ 0 & (X'_2 X_2)^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & (X'_k X_k)^{-1} \end{bmatrix} \begin{pmatrix} X'_1 Y \\ X'_2 Y \\ \cdots \\ X'_k Y \end{pmatrix} \\
 &= \begin{pmatrix} (X'_1 X_1)^{-1} X'_1 Y \\ (X'_2 X_2)^{-1} X'_2 Y \\ \cdots \\ (X'_k X_k)^{-1} X'_k Y \end{pmatrix}
 \end{aligned}$$

Por tanto, observamos que en este caso la estimación del parámetro β_i depende exclusivamente de las observaciones de la variable X_i y de las observaciones de la variable endógena. Está, en consecuencia, libre de las interacciones entre el conjunto de las variables explicativas. Por tanto, resulta indiferente estimar el modelo conjunto o bien estimar k modelos independientes donde se enfrenta cada una de las variables explicativas frente a la variable endógena.

Este caso, conocido comúnmente como caso de regresores ortogonales, sería el caso ideal en cuanto a la interpretación de los coeficientes, en el sentido de que la estimación de cada uno de los parámetros se encuentra libre de las interacciones entre las explicativas. Desde el punto de vista práctico no es muy interesante por cuanto no es habitual la ausencia de correlación entre las variables explicativas del modelo.

2.2 Colinealidad perfecta

Frente al caso anterior donde no existía correlación entre los regresores del modelo, supongamos aquel caso donde existe una relación lineal entre un sub-

conjunto de las variables explicativas. Desde el punto de vista matemático, esta relación se puede formular como sigue:

$$a_1 X_1 + a_2 X_2 + \dots + a_k X_k = 0$$

donde $a_i (i = 1, 2, \dots, k)$ son parámetros tales que al menos uno de ellos es distinto de 0. En estas condiciones, existe una combinación lineal entre las observaciones de las variables explicativas del modelo. Por tanto, esto supone que los k vectores que componen la matriz X no son linealmente independientes, lo que conlleva que su rango será, necesariamente inferior a k (mantenemos el supuesto de que $T > k$). Entonces, como el rango de la matriz $X'X$ coincide con el rango de la matriz X , la matriz $X'X$, que es cuadrada de orden k , no será de rango completo. Como consecuencia de este resultado, $\det X'X = 0$, por lo que esta matriz será singular, no siendo posible de invertir la matriz inversa $(X'X)^{-1}$. Al no ser posible de invertir esta matriz, el sistema de ecuaciones $X'X\beta = X'Y$ carece de una única solución. Por tanto, el vector de estimadores mínimo cuadrático ordinario no está definido en este tipo de situaciones.

En realidad, la afirmación anterior no es totalmente cierta y es necesario matizarla. Supongamos que queremos estimar el siguiente modelo:

$$y_t = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

donde, además, sabemos que $x_{1i} = k x_{2i}$, donde k es un parámetro que toma un valor fijo. La presencia de una combinación lineal en la matriz X es evidente, por lo que su rango será igual a 1. De ahí podemos deducir que la matriz $X'X$ será singular, por lo que no existe la matriz $(X'X)^{-1}$. De ahí que no este definido el vector de estimadores mínimo cuadrático ordinario, por lo que no podemos estimar el modelo anterior. No obstante, debemos indicar que el uso de matrices inversas generalizadas permite invertir matrices que no son singulares. Esto nos permitiría estimar alguno de los parámetros del modelo. En cualquier caso, una simple transformación nos lleva al siguiente modelo:

$$\begin{aligned} y_t &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i = \beta_1 (k x_{2i}) + \beta_2 x_{2i} + u_i \\ &= (\beta_1 k + \beta_2) x_{2i} + u_i = b x_{2i} + u_i \end{aligned}$$

Por tanto, es sencillo comprobar que podemos ofrecer una estimación del parámetro $\hat{b} = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2}$. Sin embargo, debemos tener en cuenta que $\hat{b} = \beta_1 k + \beta_2$ por lo que somos incapaces de distinguir los efectos individuales de las variables x_1 y x_2 . Este es el verdadero problema causado por la presencia de colinealidad perfecta: no somos capaces de identificar los efectos de cada una de las variables explicativas sobre la variable endógena. Sí que podemos estimar el efecto conjunto, que viene a expresar un promedio de los individuales.

Aparentemente, la presencia de colinealidad perfecta supone un problema dramático para la estimación mínimo cuadrático ordinaria. Primero, por que

no parece tener solución y, segundo, porque no está clara la forma de detectar su presencia. Sin embargo, desde el punto de vista empírico la presencia de colinealidad perfecta no supone un problema irresoluble, ni mucho menos. En primer lugar, la detección de este tipo de problemas es inmediata. Al incluir una combinación lineal de variables explicativas, el determinante de la matriz $X'X$ se iguala a 0 y es imposible estimar el modelo. Por tanto, no corremos el riesgo de caer inadvertidamente en este problema. Una vez que observamos que tenemos un problema el siguiente paso es solucionarlo.

Ejemplos: ratios, trampa de las variables ficticias

2.3 Colinealidad aproximada

Los dos casos que hemos estudiado con anterioridad son muy relevantes desde el punto de vista teórico. No obstante, desde el punto de vista práctico su incidencia no es tan importante por cuanto la presencia de regresores ortogonales es harto difícil, salvo por propia construcción del investigador, y la presencia de colinealidad exacta es producida por una incorrecta especificación. Entre medio de ambos casos extremos tenemos el caso de colinealidad aproximada. Desde el punto de vista teórico, este caso viene representado por la siguiente igualdad:

$$a_1 X_1 + a_2 X_2 + \dots + a_k X_k \simeq 0$$

donde $a_i (i = 1, 2, \dots, k)$ son parámetros tales que al menos uno de ellos es distinto de 0. Como podemos comprobar, no existe una relación lineal exacta entre las variables explicativas del modelo, aunque los regresores están relacionados. El grado de dependencia puede ser variable, por lo que los resultados que vamos a presentar son válidos tanto si la relación lineal entre las variables es muy elevada o si esta es escasa. En cualquier caso, es fácil entender que este es el caso más habitual dentro de la investigación econométrica.

En la medida que no tenemos una relación lineal, esto supone que el rango de la matriz X es igual a k , a la vez que el rango de la matriz $X'X$ es también igual a k . Por tanto, la matriz $X'X$ es de rango completo y, en consecuencia, el determinante de esta matriz es distinta de 0, por lo que es una matriz no singular. Esto supone que existe la matriz $(X'X)^{-1}$, de ahí que el vector de estimadores MCO existe y es único. En suma, desde este punto de vista, no existe ningún problema en estimar el modelo, en el sentido de que la estimación MCO es insesgada, eficiente y consistente, por cuanto no se incumplen ninguno de los supuestos básicos. Ahora bien, esto no quiere decir que no existan problemas en estimar un modelo donde las variables explicativas presentan un elevado grado de correlación. Para entender cuáles son estos problemas, podemos plantear el modelo lineal general de la siguiente manera:

$$Y = X\beta + u = X_1 b_1 + X_2 b_2 + u$$

donde X_1 es un vector $T \times 1$ y X_2 es una matriz de orden $T \times (k-1)$. Al mismo tiempo, b_1 es un escalar y b_2 es un vector de orden $(k-1) \times 1$. A partir de esta

particion, el vector de estimadores mco del modelo se puede expresar como sigue:

$$\hat{\beta} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = (X'X)^{-1} X'Y = \begin{bmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{bmatrix}^{-1} \begin{bmatrix} X'_1Y \\ X'_2Y \end{bmatrix}$$

Esta estimacion es insesgada, como hemos comentado. La varianza de b_1 es igual a:

$$Var(\hat{b}_1) = \sigma^2 (X'X)^{-1}_{11} \quad (1)$$

El elemento $(X'X)^{-1}_{11}$ se corresponde con el primer bloque de la matriz $(X'X)^{-1}$. En este caso, este elemento es un escalar. Aplicando tanto resultado procedentes de la inversion de matrices particionadas, como a partir de la aplicacion directa del teorema de Frisch-Waugh-Lowell, se demuestra que este elemento es igual a:

$$\begin{aligned} (X'X)^{-1}_{11} &= \left[X'_1X_1 - X'_1X_2 (X'_2X_2)^{-1} X'_2X_1 \right]^{-1} = \left\{ X'_1 \left[I - X_2 (X'_2X_2)^{-1} X'_2 \right] X_1 \right\}^{-1} \\ &= (X'_1M_{X_2}X_1)^{-1} \end{aligned} \quad (2)$$

donde la matriz M_{X_2} es una matriz de proyecciones tal que aplicada a un vector nos proporciona los residuos de la relacion entre X_2 y ese vector de residuos. El ejemplo mas claro es el propio modelo lineal general, donde demostramos que $\hat{u} = MY$. Al mismo tiempo, la suma residual del modelo lineal general tambien la podemos expresar en funcion de esta matrix de proyecciones: $\hat{u}'\hat{u} = Y'MY$. Si comparamos este resultado con el expresado en la ecuacion (2) debemos concluir que esta cantidad coincide con la suma residual del siguiente modelo:

$$X_1 = X_2\gamma + v \quad (3)$$

Entonces, la expresion (1) queda como sigue:

$$Var(\hat{b}_1) = \sigma^2 (X'_1M_{X_2}X_1)^{-1} \quad (4)$$

Si en la expresion anterior multiplicamos y dividimos por la misma cantidad, tenemos que:

$$Var(\hat{b}_1) = \sigma^2 \frac{1}{(X'_1M_{X_2}X_1)^{-1}} \frac{\sum (X_1 - \bar{X}_1)^2}{\sum (X_1 - \bar{X}_1)^2} \quad (5)$$

Si tenemos en cuenta que $\sum (X_1 - \bar{X}_1)^2$ es la suma total del modelo (2), entonces el coeficiente de determinación del modelo se define como sigue:

$$R_{X_2}^2 = 1 - \frac{SR}{ST} = 1 - \frac{X_1' M_{X_2} X_1}{\sum (X_1 - \bar{X}_1)^2} \quad (6)$$

utilizando este resultado, es directo probar que:

$$Var(\hat{b}_1) = \frac{1}{1 - R_{X_2}^2} \frac{\sigma^2}{\sum (X_1 - \bar{X}_1)^2}$$

donde el ratio $\frac{1}{1 - R_{X_2}^2}$ se le conoce con el nombre de factor incrementador de la varianza. A partir de la expresión anterior es fácil observar cual es el daño que puede causar la presencia de un alto grado de correlación. En el caso en el que no exista correlación entre las variables explicativas, entonces resulta que $Var(\hat{b}_1) = \frac{\sigma^2}{\sum (X_1 - \bar{X}_1)^2}$, que coincide con la varianza en el caso de regresores ortogonales. En el caso extremo en el que existe un elevado de correlación, entonces $R_{X_2}^2 \rightarrow 1$, por lo que $Var(\hat{b}_1) \rightarrow \infty$. Esto supone que, aunque la estimación puntual es insesgada, la estimación por intervalo puede llegar a tener límites infinitos, por lo que en la práctica, cualquier valor puede ser igualmente válido. De ahí que, cuanto mayor es el grado de correlación, menor es la confianza sobre la interpretación económica.

Para ver que efectos tiene esta falta de credibilidad sobre la estimación puntual de los parámetros, podemos observar el siguiente ejemplo. Supongamos las siguientes ecuaciones normales:

$$\begin{pmatrix} 200 & 150 \\ 150 & 113 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 350 \\ 263 \end{pmatrix}$$

La solución de este sistema es $\hat{\beta}_1 = \hat{\beta}_2 = 1$. Si eliminamos, por ejemplo, una observación podemos llegar a las siguientes ecuaciones normales:

$$\begin{pmatrix} 199 & 149 \\ 149 & 112 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 347.5 \\ 261.5 \end{pmatrix}$$

La solución del nuevo sistema es $\hat{\beta}_1 = -\frac{1}{2}$ y $\hat{\beta}_2 = 3$. Como vemos, la simple variación de una observación, que provoca cambios aparentemente no trascendentes en las ecuaciones, supone un cambio importante en las estimaciones de los parámetros. La explicación de por qué se está produciendo este resultado lo encontramos en el coeficiente de correlación lineal entre las dos variables que componen el sistema. Este coeficiente es igual a 0.998. Es evidente que el grado

de correlación entre las dos variables es elevado. En este tipo de situaciones, cualquier tipo de análisis estructural del modelo nos llevaría hacia interpretaciones cuando menos dudosas de los efectos de las variables explicativas sobre la variable en estudio. El problema que se nos plantea es cual es el límite de colinealidad que podemos admitir entre las variables explicativas para asumir los resultados de la estimación. Es evidente que cierto grado de correlación siempre existirá, pero ¿hasta donde admitimos su presencia? Desde este punto de vista, el problema de multicolinealidad no es un problema que deba analizarse en tiempo discreto (¿existe o no existe?) sino en tiempo continuo (¿es asumible el grado de colinealidad?). En consecuencia, cuando hablemos de problemas de colinealidad debemos entender que existe un elevado grado de colinealidad entre las variables, también se conoce con el nombre de colinealidad grave. En la siguiente sección vamos a analizar diversos métodos que nos pueden ayudar a discernir cuando es grave el problema de colinealidad.

3 Detección de problemas de colinealidad aproximada grave

En la sección anterior hemos comprobado que el verdadero problema relacionado con la presencia de colinealidad entre los regresores del modelo es que este sea su ciente elevado como para causar una seria distorsión en la estimación por intervalo de los parámetros de posición. Desde este punto de vista, el problema de colinealidad aproximada es un problema de grado. Lo que hay que determinar es cuando esta relación entre las variables explicativas del modelo es grave o no. En la literatura existen diversas medidas para determinar la gravedad de este problema. A continuación, vamos a citar algunas de ellas.

3.1 Métodos basados en el grado de correlación lineal de las variables

Esta primera familia de métodos está basada en el cálculo del coeficiente de correlación lineal para todos los pares de variables que componen la especificación del modelo. Dadas dos variables cualesquiera, x_i y x_j , este coeficiente se define de la siguiente manera:

$$r_{ij} = \frac{\sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{i=1}^T (x_{it} - \bar{x}_i)^2} \sqrt{\sum_{i=1}^T (x_{jt} - \bar{x}_j)^2}}$$

Si alguno, o varios, de estos pares de variables presenta un valor del coeficiente de correlación lineal elevado, entonces podemos concluir que existen problemas graves de colinealidad aproximada.

La cuestión que se nos plantea es determinar cuando podemos concluir que la multicolinealidad es grave o no. Esto supone que debemos arbitrar un límite para el coeficiente de correlación lineal tal que si alguno de los coeficientes es superior podamos concluir que existe problemas graves de colinealidad. La decisión de dónde situar este punto de corte no es sencilla. Por ejemplo, si $r_{ij} = 0.8$ ¿es eso colinealidad grave? ¿Debemos considerar grave el problema sólo si $r_{ij} > 0.9$? Este es uno de los problemas de medir la gravedad del problema a partir del coeficiente de correlación lineal.

En ocasiones se toma como límite el coeficiente de determinación del modelo, de tal forma que si para algún par de valores i, j se cumple que $r_{ij}^2 > R^2$, entonces se considera que existen problemas graves de colinealidad. Si, por el contrario, $r_{ij}^2 \leq R^2$, entonces se considera que los problemas no son graves. Este límite puede provocar ciertas situaciones contradictorias. Por ejemplo, supongamos que el coeficiente de correlación lineal entre dos variables explicativas es 0.95 y que el coeficiente de determinación del modelo es 0.91. En este caso, la conclusión a la que llegamos es que $r_{ij}^2 = .9025 < R^2 = 0.91$, por lo que no existen problemas graves de colinealidad. Esta conclusión es ciertamente contraintuitiva, por cuanto la correlación entre ambas variables es incuestionable.

Otra debilidad que presentan los métodos basados en el grado de correlación lineal entre pares de variables es que pueden no detectar problemas en aquellos casos en los que el patrón de colinealidad afecta a más de 2 variables. Por ejemplo, es posible que el patrón de colinealidad venga determinado por la relación $x_4 = x_2 + x_3$, donde x_2 y x_3 son ortogonales entre sí. La correlación entre x_4 y sus dos componentes no será muy elevada, por lo que el coeficiente de correlación lineal no indicará un alto grado de colinealidad. Sin embargo, la inclusión conjunta de las tres variables en la especificación supone la presencia de problemas de colinealidad perfecta, lo que invalida el uso del coeficiente de correlación lineal en este tipo de escenarios.

3.2 Regresiones auxiliares

Para superar los problemas anteriores, se puede realizar algunas regresiones auxiliares. La idea es que si existe un alto grado de colinealidad, entonces una de las variables explicativas tiene que estar muy relacionada con el resto. Por tanto, una regresión entre una de las variables explicativas y el resto de las mismas puede detectar esta relación. La propuesta de Farrar-Glauber es interesante en este sentido. Estos autores proponen estimar el siguiente modelo:

$$x_{si} = \gamma_1 + \gamma_2 x_{2i} + \dots + \gamma_{s-1} x_{s-1,i} + \gamma_s x_{s+1,i} + \dots + \gamma_{k-1} x_{ki} + v_i \quad (7)$$

y, a partir de su estimación, podemos calcular el estadístico del análisis de la varianza:

$$FG = \frac{R_1^2}{1 - R_1^2} \frac{T - (k - 1)}{k - 2}$$

donde R_1^2 es el coeficiente de determinación del modelo (7). Bajo la hipótesis nula $H_0 : \gamma_2 = \gamma_3 = \dots = \gamma_{k-1} = 0$, este estadístico sigue una distribución $F_{k-1, T-(k-1)}$. Si el estadístico de Farrar-Glauber es superior al valor crítico teórico, entonces rechazamos la hipótesis nula y, por tanto, concluimos que existen problemas graves de colinealidad. Si, por contra, se acepta la hipótesis nula, debemos concluir que no existen problemas graves de colinealidad.

Relacionado con este método está el cálculo del coeficiente que se conoce como factor incrementador de la varianza.

3.3 Métodos basados en la matrix $X'X$

Cómo hemos visto con anterioridad, en el caso extremo en el que tenemos una combinación lineal entre las variables explicativas del modelo, entonces el determinante de la matriz $X'X$ es igual a 0. Si esta relación no es exacta, sino aproximada, el valor del determinante está próximo a 0, aunque es matemáticamente distinto a este valor. La conclusión a la que llegamos es que cuanto más próximo se encuentre el determinante de la matriz $X'X$ a 0, mayor es el grado de colinealidad entre las variables.

Tomando como punto de partida esta premisa, una primera posibilidad para medir el grado de colinealidad entre las variables es estudiar el valor de $|X'X|$. Cuanto más próximo se encuentre a 0, mayor es el grado de colinealidad. Este método presenta dos problemas. El primero es donde ponemos el punto de corte para determinar si los problemas de colinealidad son graves. Por ejemplo, si $|X'X| = 0.1$, ¿debemos concluir que existen problemas de colinealidad grave? ¿Y si $|X'X| = 0.05$? Esta primera decisión no es sencilla. Además, existe un segundo problema. Si modificamos las unidades en las que están medidas las variables, podemos alterar el valor de $|X'X|$ haciendo que éste se aleje o aproxime hacia 0. Por ejemplo, si tenemos datos medidos en pesetas y el valor del determinante $|X'X|$ toma el valor, digamos, k , si transformamos todas las variables en millones de pesetas, el valor del determinante es entonces $k/10^6$. Por tanto, las unidades en las que están medidas de las variables pueden alterar nuestra decisión sobre la presencia de problemas de colinealidad grave.

Para superar este problema podemos tipificar las variables explicativas, esto es, restarles a cada una de ellas su media muestral y dividir este resultado por la desviación típica. En estas condiciones, el valor del determinante no depende de las unidades de medida utilizadas y, además, se encuentra acotado entre 0 y 1. No obstante, queda sin resolver la cuestión de cuando podemos considerar que el determinante de la matriz se encuentra lo suficientemente próximo a 0 cómo para admitir la presencia de problemas graves de colinealidad.

Un tercer método para determinar la presencia de problemas de colinealidad graves es el del número condición. Para comprender este método, debemos considerar que, al ser la matriz $X'X$ simétrica, su determinante es igual al producto de sus valores propios.

$$|X'X| = \prod_{i=1}^k \lambda_i$$

Si el determinante de esta matriz es igual a 0, entonces al menos uno de los valores propios tiene que ser igual a 0. EN el caso de que tengamos un alto grado de colinealidad entre un grupo de variables explicativas, entonces el determinante tiende hacia 0, por lo que al menos uno de los valores propios tiene que estar proximo a 0. De nuevo, el problema es determinar cuando un valor propio esta lo su cientemente proximo a 0 como para considerar que tenemos problemas graves de colinealidad.

Para solucionar este problema, lo que se hace habitualmente es formar el ratio entre el mayor y el menor de los valores propios. Cuanto mayor es el valor de este ratio, mayor seran los problemas de colinealidad. La raiz cuadrada de este ratio se le conoce como el numero condicion:

$$n^\circ \text{condicion} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

Si el valor de este cociente es mayor que 25 entonces se considera que existen problemas graves de colinealidad.

3.4 Contradiccion entre los estadisticos t y F

Como hemos visto con anterioridad, una de las consecuencias de la presencia de un alto grado de colinealidad es que la varianza del vector de estimadores puede tomar valores elevados. En consecuencia, la estimacion por intervalo de estos parametros puede ser poco precisa. Esto supone que cualquier valor real puede estar incluido dentro de este intervalo de con anza, lo que sin duda incluye el valor 0. En consecuencia, podemos aceptar la hipotesis nula de no signi catividad de los parametros.

Por otro lado, es posible que la especi cacion del modelo incluya todas las variables relevantes, por lo que no hay omision de variables relevantes. Por consiguiente, el coe ciente de determinacion del modelo puede tomar valores elevados, haciendo que el valor del estadistico del analisis de la varianza nos conduzca a rechazar la hipotesis nula de no signi catividad conjunta de los parametros.

En suma, tenemos de un lado que las variables no son individualmente signi cativas, pero conjuntamente si. Este tipo de contradicciones son habituales en aquellos casos en los que existe un alto grado de colinealidad entre las variables explicativas. Por tanto, si se observa este tipo de efectos, debemos sospechar que existen estos problemas. Sin embargo, no es la unica causa, por lo que no existe una relacion bi-univoca entre contradiccion y colinealidad grave.

4 Soluciones

En las secciones anteriores hemos discutidos los problemas que causa la presencia de colinealidad entre los regresores del modelo, así como los medios para detectarlos. En esta sección vamos a presentar distintos métodos para intentar solucionar estos problemas. Todos ellos están perfectamente diseñados desde el punto de vista teórico. No obstante, debemos indicar de antemano que ninguno de ellos supone una solución óptima para los problemas de colinealidad, aunque en la literatura se utilizan habitualmente.

a) Una primera solución es aumentar la muestra. La idea es que si las nuevas observaciones rompen el patrón de colinealidad, entonces los problemas de colinealidad grave pueden disiparse. Esta posible solución es poco razonable en la práctica. No tiene sentido guardarse algunas observaciones para que, en el caso de que tengamos problemas de colinealidad, poderlos solucionar. En la mayoría de los trabajos aplicados, la disponibilidad muestral es uno de los mayores problemas contra los que debe luchar el investigador.

b) Introducción de información a priori

El razonamiento en el que se basa este método reside en el hecho de que si conocemos los valores de los parámetros del modelo, podemos incluir esta información en la especificación del modelo empírico. De esta manera, no es necesario estimar este parámetro y, por tanto, podemos eliminar los problemas de colinealidad. Por ejemplo, si tenemos la siguiente relación:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad (8)$$

donde conocemos que $x_{2t} \simeq kx_{3t}$. Si suponemos conocido el valor del parámetro β_3 podemos incluir esta información en el modelo anterior, de forma que éste queda:

$$\begin{aligned} y_t - \beta_3 x_{3t} &= \beta_1 + \beta_2 x_{2t} + u_t \Rightarrow \\ y_t^* &= \beta_1 + \beta_2 x_{2t} + u_t \end{aligned}$$

Es evidente que ahora han desaparecido los problemas de colinealidad aproximada. No obstante, debemos indicar que en la práctica no es probable que podamos utilizar este método, por cuanto nunca vamos a conocer a priori el verdadero valor de los parámetros poblacionales del modelo, por lo que desde el punto de vista práctico no tiene mucha aplicabilidad.

c) Eliminación de variables

Otra posibilidad es la de eliminar una de las variables. Es decir, si, por ejemplo, el proceso generador de los datos coincide con (8), donde $x_{2t} \simeq kx_{3t}$, podríamos estimar el siguiente modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t$$

Es evidente que la estimación mco de este modelo es sesgada, por cuanto estamos eliminando una de las variables que aparecen en el proceso generador de los datos. Sin embargo, también es cierto que, al desaparecer el patrón de colinealidad, entonces existe una reducción evidente en la varianza del estimador. La idea ahora es la de comparar el error cuadrático medio de cada uno de los estimadores. Recordemos que, dado un estimador $\hat{\theta}$ del parámetro θ , el error cuadrático medio se define así:

$$ECM(\hat{\theta}) = \text{sesgo}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

Si el error cuadrático medio de los estimadores del modelo (8) es superior al obtenido cuando se omite una variable relevante, entonces resulta rentable, en términos de eficiencia, la eliminación de esta variable.

El problema que presenta este método es que en la práctica desconocemos los valores de los parámetros poblacionales y, en consecuencia, el cálculo del error cuadrático medio es imposible.

d) Ridge regresion o regresión cresta.

Este método tiene ciertas similitudes con el anterior, en el sentido de que está basado en la introducción de un sesgo en la estimación del vector de parámetros de posición. Se espera que, a cambio, su varianza disminuya, por lo que el error cuadrático medio se reduce. En concreto, Hoerl y Kennard (1970a,b) definen el siguiente estimador:

$$b(k) = (X'X + cI_k)^{-1}X'Y$$

donde c es una constante estrictamente positiva, que se conoce habitualmente como el parámetro de sesgo. Dado un valor del parámetro k no estocástico, tenemos que:

$$\begin{aligned} E[b(k)] &= (X'X + cI_k)^{-1}X'E(Y) = (X'X + cI_k)^{-1}X'E(X\beta + u) \\ &= (X'X + cI_k)^{-1}X'X\beta + (X'X + cI_k)^{-1}X'E(u) \\ &= (X'X + cI_k)^{-1}X'X\beta \end{aligned}$$

Es evidente que si $c \neq 0$, entonces $E[b(k)] \neq \beta$, por lo que la estimación del vector de parámetros de posición está sesgada. Por otro lado, la matriz de varianzas y covarianzas de este vector de estimadores es igual a:

$$\begin{aligned} \text{Var}[b(k)] &= (X'X + cI_k)^{-1}X'E(uu')X(X'X + cI_k)^{-1} = \\ &= \sigma^2 (X'X + cI_k)^{-1}X'X(X'X + cI_k)^{-1} \end{aligned}$$

Si tenemos en cuenta que $\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, es directo comprobar que $\text{Var}(\hat{\beta}) - \text{Var}[b(k)]$ es una matriz semidefinida positiva. En consecuencia,

la reducción en la varianza puede compensar la creación del sesgo y, en consecuencia, el estimador $b(k)$ puede tener menor error cuadrático medio que el estimador mco. Si esto ocurre, sería preferible el uso del estimador $b(k)$.

Desde el punto de vista teórico, la discusión sobre este estimador cresta es elegante (ver Judge et al.). Sin embargo no tiene mucha aplicabilidad dado que desconocemos cuál es el sesgo real que estamos conociendo y, por tanto, cuál es la posible ganancia en términos de error cuadrático medio. Otra cuestión que di culta su aplicabilidad es determinar el valor más adecuado del parámetro c .

e) Convivir con el problema.

Según algunos autores, el problema de colinealidad es cosustancial con las muestras disponibles en economía y, en consecuencia, siempre van a existir problemas de colinealidad. En consecuencia, el investigador debe aprender a convivir con esta situación (sic).

5 Colinealidad aproximada: de ciencia muestral o problema de especificación

En el apartado anterior, hemos presentado una serie de recetas que se encuentran en la literatura encaminadas a mitigar los efectos producidos por la presencia de colinealidad aproximada grave entre las variables explicativas del modelo. Como hemos visto, ninguno de los métodos ofrece serias garantías de poder solucionar el problema, al menos desde el punto de vista práctico. Todos, eso sí, suponen desarrollos teóricos correctos e, incluso, algunos de ellos elegantes. Sin embargo, su aplicabilidad es prácticamente nula.

Sobre este respecto debemos hacer un primer comentario. La mayoría de estas aportaciones se realizan entre finales de los años 60 y primeros de los 70. Por tanto, se trata de aportaciones muy pioneras dentro de la econometría y que no tienen en cuenta los recientes avances en, por ejemplo, series temporales, modelización económica, etc. Por ejemplo, para todos estos autores la presencia de correlación entre los regresores es un problema. Sin embargo, hoy en día, sabemos que si las series son integradas, como sucede en la mayoría de los modelos macroeconómicos, la presencia de correlación entre los regresores es algo necesario y casi conveniente para poder modelizar el comportamiento de largo plazo de las variables.

La cuestión que queremos introducir aquí, y que nos parece clave para entender los problemas de correlación, es que los problemas de colinealidad no son tanto una cuestión de ciencia muestral, como de mala especificación del modelo. Para entender este punto, supongamos que queremos modelizar el comportamiento las importaciones españolas de manufacturas, por ejemplo. De acuerdo con nuestros conocimientos teóricos sobre el comportamiento de esta variable, parece acertado intentar explicar su evolución a partir de una variable de precio y otra variable de escala, que recoja el efecto de la demanda. Para la primera, podríamos utilizar un índice de tipo de cambio real efectivo entre la peseta y el conjunto de monedas de los principales países con los que comercia

España (TCRE). En el segundo caso, algunos autores consideran como variable de escala el PIB, mientras que otros consideran como variable más adecuada la Demanda Nacional (DN), de nida como suma de Consumo más Inversión. Supongamos que, para salir de dudas acerca de la idoneidad de la variable de escala, estimamos un modelo con las tres mencionadas variables. Los resultados para el periodo T=1964-1995 es el siguiente:

$$\begin{aligned} \ln \widehat{ME}_t &= (0.81)1.98 - (-9.11)8.91 \ln \text{PIB}_t + (11.63)9.90 \ln \text{DN}_t - (-3.20)1.30 \ln \text{TCRE}_t \\ R^2 &= 0.979, r = 0.9972 \end{aligned}$$

Los resultados de la anterior estimación muestran un alto grado de ajuste para la muestra utilizada. Además, la estimación de la elasticidad precio entra dentro de los límites aceptables (-1.30). Sin embargo, las elasticidades de de las variables de escala resultan extrañas. Por ejemplo, la elasticidad del PIB es negativa, lo que no tiene mucho sentido desde el punto económico. Además, su valor absoluto es realmente elevado. Por tanto, existen dudas razonable sobre la interpretación de este parámetro. De igual forma, la estimación de la elasticidad de DN resulta difícil de interpretar. En esta caso, el signo es acorde a la teoría económica. Sin embargo, su magnitud es desorbitada. Debemos pensar que implica que un incremento de un 1% en la demanda nacional, supone un incremento de casi el 10% en las importaciones. Este resultado no es propio de una economía estable, por lo que su interpretación económica no es sencilla.

Una explicación de estos resultados la podemos dar desde el punto de vista de la posible presencia de un elevado grado de correlación entre PIB y DN. En concreto, el coeficiente de correlación lineal entre ambas variables es igual a 0.9972. Es indudable que ambas variables están ofreciendo una explicación de la evolución de las importaciones desde una óptica muy similar. Por tanto, es lógico que no seamos capaces de discriminar cuál es el efecto individual de cada una de las variables sobre la evolución de las importaciones y la forma de declarar la presencia de este problema es la existencia de unas estimaciones de nula interpretación económica.

Si queremos solucionar el problema, parece adecuado utilizar sólo una variable de escala para explicar la evolución de la variable endógena. Utilizando, por ejemplo, el PIB, la estimación que obtenemos es la siguiente:

$$\begin{aligned} \ln \widehat{ME}_t &= (-4.73) - 18.73 + (13.70)2.43 \ln \text{PIB}_t - (-2.07)1.98 \ln \text{TCRE}_t \\ R^2 &= 0.882 \end{aligned}$$

El ajuste es ahora inferior al anterior, lo que indica que existen problemas sin resolver. Sin embargo, las estimaciones de las elasticidades resulta interpretable desde el punto de vista económico. En este caso, la elasticidad del PIB es 2.43, un valor que es aproximadamente concordante con el de previos trabajos.

En resumen, vemos que el problema de multicolinealidad es muchas ocasiones es generado por una incorrecta especificación del modelo. En este caso,

una sobreparametrización puede generar un alto grado de colinealidad entre los regresores, dificultando la interpretación económica del modelo estimado, por cuanto las estimaciones de los parámetros de posición carecen de precisión. La misión del investigador es detectar y corregir esta sobreespecificación.