MÉTODO DE ALZADO. UN PROCEDIMIENTO ALTERNATIVO PARA LA ESTIMACIÓN DE PARÁMETROS EN PRESENCIA DE COLINEALIDAD.

José Soto Liria
jsoto@cajamar.es
Dr. en Matemáticas
Catedrático de Enseñanza Secundaria.

José García Pérez
nikita97@cajamar.es
Departamento de Economía Aplicada
Universidad de Almería.

Antonio S. Andújar Rodríguez
andujar@ual.es
Departamento de Estadística y Matemática Aplicada
Universidad de Almería.

1 Detección de la colinealidad. El Número Métrico

Cuando se trata de resolver un problema de regresión lineal del tipo

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

con $\boldsymbol{Y} \in \mathbb{R}^n$, $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p]$, $\quad \boldsymbol{X}_i \in \mathbb{R}^n$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ y se aplica el método de MCO se tiene que:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\\ \hat{\boldsymbol{Y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}}\end{aligned}$$

Desde el punto de vista geométrico, el vector $\hat{\boldsymbol{Y}}$ depende linealmente de los vectores columna de la matriz $\boldsymbol{X}$ puesto que es una combinación lineal de los mismos. Ello permite una interpretación vectorial del problema.

Resolver un problema de regresión lineal por el método de MCO equivale, desde el punto de vista vectorial, a encontrar el vector $\hat{\boldsymbol{Y}}$ perteneciente al subespacio vectorial generado por los vectores columnas de la matriz $\boldsymbol{X}$ con la condición de que la distancia euclídea entre $\boldsymbol{Y}$ e $\hat{\boldsymbol{Y}}$ sea mínima.

Por otra parte, el problema de la posible existencia de colinealidad o cuasi colinealidad entre las columnas de la matriz $\boldsymbol{X}$, expresado analíticamente, significa que existe un vector $\boldsymbol{c} \neq \boldsymbol{0}$, $\boldsymbol{c} = (c_1, c_2, \ldots, c_p)'$ tal que

$$c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2 + \ldots + c_p\boldsymbol{X}_p \approx \boldsymbol{0}$$

o, lo que es igual, al menos uno de los $\boldsymbol{X}_i, \ i = 1, 2, \ldots, p$ es cuasi combinación lineal de los demás. Esto, desde el punto de vista geométrico, se puede interpretar como que la distancia entre dicho vector y el subespacio generado por los demás es "muy pequeña". En el caso extremo de que esa distancia sea cero la dependencia lineal es perfecta. Si la distancia es 1 el vector es ortogonal al subespacio generado por los demás (para evitar la influencia de las unidades de medida en la que están expresados los datos, y dado que está generalmente aceptado en la literatura sobre regresión lineal, se consideran los vectores centrados y normalizados a módulo 1).

Para una matriz de vectores $\boldsymbol{X}$ se define el *índice métrico asociado al vector* $\boldsymbol{X}_i$ y se nota $\mathrm{IM}(\boldsymbol{X}_i)$ o simplemente IM_i a la distancia entre $\boldsymbol{X}_i$ y el subespacio vectorial generado por los restantes vectores de $\boldsymbol{X}$. Así mismo, se define el *número métrico asociado a la matriz* $\boldsymbol{X}$ y se nota $\mathrm{NM}(\boldsymbol{X})$ como el menor de los índices métricos asociados a sus vectores columna. Las fórmulas de cálculo correspondientes son

$$\mathrm{IM}_i = \begin{cases} 0 & \text{si} \quad \det(\boldsymbol{X'X}) = 0 \\ \sqrt{\dfrac{\det(\boldsymbol{X'X})}{M_{ii}}} & \text{si} \quad \det(\boldsymbol{X'X}) \neq 0 \end{cases}$$

donde M_{ii} es el menor de orden $p-1$ correspondiente al elemento m_{ii} de la diagonal principal de la matriz $\boldsymbol{M} = \boldsymbol{X'X}$.

$$\mathrm{NM}(\boldsymbol{X}) = \min_i \{\mathrm{IM}_i\} \quad i = 1, \ldots p$$

De este modo se detecta[1] qué columna es la más próxima al subespacio generado por las demás y "qué grado" de proximidad hay.

2 Estimación de parámetros

Cuando se plantea un modelo de regresión lineal pueden aparecer indicadores que muestran la presencia de cuasi-colinealidad entre las variables exógenas. Una vez confirmada su presencia en la muestra, debemos enfrentarnos a la resolución del problema [SGA00].

En la actualidad, los métodos existentes para la estimación de parámetros en presencia de colinealidad los podemos agrupar en dos clases:

1. Aquellos que van directamente a resolver el problema algebraico olvidándose de la muestra, entre los que podemos citar a los estimadores cresta Hoerl y Kennard [HK70a,

[1] Véase [GAS99]

HK70b], García Ferrer [Gar77], Brown y Beattie [BB75], MacDonal y Galarneau [MG75], Polverini [Pol78], Hemmerle [Hem75], Hocking, Speed y Lynn [HSL76], Fourgeaud, Gourieux y Pradel [FGP84], Casella [Cas85], etc..

2. Los estimadores que proponen resolver el problema actuando sobre la muestra en el sentido de 'eliminar' parte de los datos muestrales. En estos métodos podrían incluirse los estimadores por componentes principales — Kendall [Ken57], Massy [Mas65], Farrar y Glaubert [FG67], Silvey [Sil69], Johnston [Joh89] — e inversa generalizada — Rao [Rao62], D. W. Marquard [Mar70] —.

En este trabajo se propone un nuevo método[2] que se detalla en el epígrafe siguiente y que se podría incluir en el el segundo de los grupos citados, puesto que intenta buscar la solución del problema allí donde se genera, es decir, en la muestra, pero en lugar de eliminar datos que podrían contener información relevante lo que se propone es conservar la información disponible y modificar ligeramente aquella de las variables que es la máxima responsable de la colinealidad —que se habrá detectado a través del número métrico—, de modo que se conserven todas las buenas cualidades relacionadas con su presencia en la muestra y que se manifiestan al aplicar MCO —como un elevado R^2— y se mitiguen o eliminen aquellas que provocan los problemas relacionados con la estimación de parámetros en presencia de colinealidad —como la no significatividad de algunos de los coeficientes parciales de pendiente—.

3 Método de Alzado

A través del número métrico es posible determinar cuál de las variables exógenas es "la mayor" causante de la cuasi-colinealidad siendo dicha variable es la "más próxima" , en sentido euclídeo, al subespacio generado por las demás. Se puede considerar, sin pérdida de generalidad, que esa variable es $\boldsymbol{X}_1$.
Para corregir la cuasi-colinealidad, en principio, tenemos dos opciones:

a) Eliminar la variable $\boldsymbol{X}_1$ si se considera que su aportación a la variable endógena $\boldsymbol{Y}$ está recogida por el resto de las variables $\{\boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p\}$ al ser cuasi-combinación

[2] Ponencia presentada en la XV Reunión de Asepelt-España, [SGA01]

lineal de ellas, Figura 1.

b) Si la opción a) no parece adecuada porque tenemos fundado interés en que dicha variable permanezca en el modelo, el método que se propone consiste en modificar ligeramente $\boldsymbol{X}_1$ en el sentido de "alejarla" del subespacio $\ll \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p \gg$. De este modo, se define: $\tilde{\boldsymbol{X}}_1 = \boldsymbol{X}_1 + \lambda_1 \boldsymbol{e}_1$, $\lambda_1 > 0$ para algún λ_1 elegido adecuadamente y

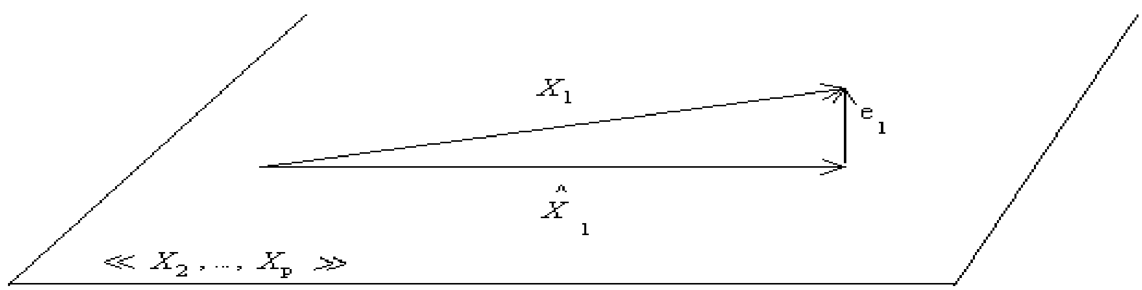

Figura 1: Vector cuasi-combinación lineal de otros

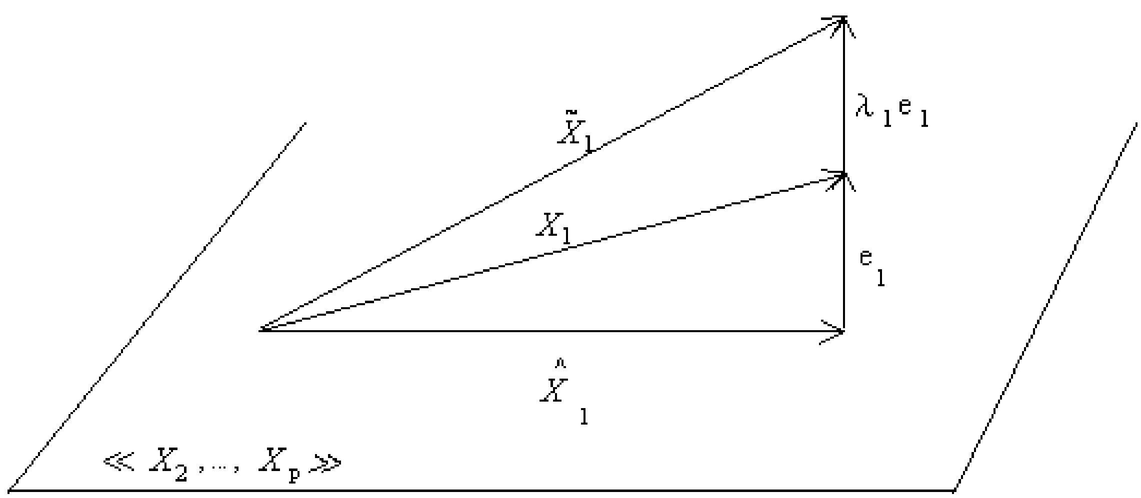

Figura 2: Alzado de la variable X1

siendo $\boldsymbol{e}_1 = \boldsymbol{X}_1 - \hat{\boldsymbol{X}}_1$ el vector de residuos[3] que aparece al hacer la regresión por MCO de $\boldsymbol{X}_1$ sobre $\{\boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p\}$. Esta variable $\tilde{\boldsymbol{X}}_1$, que denominaremos *alzada* de $\boldsymbol{X}_1$, tiene menos problemas de cuasi-colinealidad con $\{\boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p\}$ que $\boldsymbol{X}_1$, pues

$$\begin{aligned} \mathrm{d}(\tilde{\boldsymbol{X}}_1, \ll \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p \gg) = & \|\boldsymbol{e}_1 + \lambda_1 \boldsymbol{e}_1\| = \\ & (1+\lambda_1)\|\boldsymbol{e}_1\| > \|\boldsymbol{e}_1\| = \mathrm{d}(\boldsymbol{X}_1, \ll \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p \gg) \end{aligned}$$

(Véase Figura 2).

Las proyecciones de $\boldsymbol{X}_1$ (el vector de datos originales) y $\tilde{\boldsymbol{X}}_1$ (el vector de datos corregidos o alzado) sobre el hiperplano $\ll \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots, \boldsymbol{X}_p \gg$ son la misma: $\hat{\boldsymbol{X}}_1$.

Entre la variable original $\boldsymbol{X}_1$, su alzada $\tilde{\boldsymbol{X}}_1$ y la proyección de ambas $\hat{\boldsymbol{X}}_1$ sobre el citado subespacio se establecen las siguientes relaciones:

$$\begin{aligned} \boldsymbol{X}_1 &= \hat{\boldsymbol{X}}_1 + \boldsymbol{e}_1 \\ \boldsymbol{X}_1^2 &= \hat{\boldsymbol{X}}_1^2 + \boldsymbol{e}_1^2 \\ \tilde{\boldsymbol{X}}_1^2 &= \hat{\boldsymbol{X}}_1^2 + (1+\lambda_1)^2 \boldsymbol{e}_1^2 \end{aligned}$$

Lema 3.1 *Los espacios vectoriales* $\ll \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p \gg$ *y* $\ll \tilde{\boldsymbol{X}}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p \gg$ *son idénticos.*

Demostración:
En efecto, $\tilde{\boldsymbol{X}}_1 = \boldsymbol{X}_1 + \lambda_1 \boldsymbol{e}_1$ donde $\boldsymbol{e}_1$ es el vector de residuos en la regresión por MCO de $\boldsymbol{X}_1$ sobre $\{\boldsymbol{X}_2, \ldots, \boldsymbol{X}_p\}$. Por tanto, $\boldsymbol{e}_1 = \boldsymbol{X}_1 - \gamma_2 \boldsymbol{X}_2 - \ldots - \gamma_p \boldsymbol{X}_p$. Luego, $\tilde{\boldsymbol{X}}_1$ es combinación lineal de $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p\}$ y $\ll \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p \gg = \ll \tilde{\boldsymbol{X}}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p \gg$

□

Lema 3.2 *Sean* $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\hat{\boldsymbol{\beta}}$ *el estimador de MCO de* $\boldsymbol{\beta}$, $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ *la proyección de* $\boldsymbol{Y}$ *sobre el subespacio* $\ll \boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p \gg$ *y* $\boldsymbol{e}_{\mathrm{o}}$ *el vector de residuos,* $\boldsymbol{e}_{\mathrm{o}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$. *Supóngase que se hace la regresión auxiliar por MCO de* $\boldsymbol{X}_k$, $1 \leq k \leq p$ *sobre*

[3] Se representa por $\boldsymbol{e}_{\mathrm{o}}$ el vector de residuos asociado a regresión de $\boldsymbol{Y}$ sobre $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p\}$ y $\boldsymbol{e}_i$ al asociado a la regresión de $\boldsymbol{X}_i$ sobre $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1}, \boldsymbol{X}_{i+1}, \ldots, \boldsymbol{X}_p\}$.

$\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k-1}, \boldsymbol{X}_{k+1}, \ldots, \boldsymbol{X}_p\}$ siendo $\hat{\boldsymbol{X}}_k = \sum_{i=1, i\neq k}^{p} \gamma_i \boldsymbol{X}_i$ y $\boldsymbol{e}_k = \boldsymbol{X}_k - \hat{\boldsymbol{X}}_k$ el correspondiente vector de residuos. Sea $\tilde{\boldsymbol{X}}$ la matriz $\tilde{\boldsymbol{X}} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{k-1}, \tilde{\boldsymbol{X}}_k, \boldsymbol{X}_{k+1}, \ldots, \boldsymbol{X}_p]$, donde $\tilde{\boldsymbol{X}}_k = \boldsymbol{X}_k + \lambda_k \boldsymbol{e}_k$, y el modelo $\boldsymbol{Y} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ se resuelve por MCO notando por $\tilde{\boldsymbol{\beta}}$ el correspondiente estimador de $\boldsymbol{\beta}$ e $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}$, se verifica que $\tilde{\boldsymbol{Y}} = \hat{\boldsymbol{Y}}$.

Demostración:
En efecto, $\boldsymbol{e}_k = \boldsymbol{X}_k - \hat{\boldsymbol{X}}_k \in \ll \boldsymbol{X}_1, \ldots, \boldsymbol{X}_p \gg$. Al hacer la regresión por MCO de $\boldsymbol{Y}$ sobre $\{\boldsymbol{X}_1, \ldots, \tilde{\boldsymbol{X}}_k, \ldots, \boldsymbol{X}_p\}$ se obtiene $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$; $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}$ que es la proyección de $\boldsymbol{Y}$ sobre $\ll \boldsymbol{X}_1, \ldots, \tilde{\boldsymbol{X}}_k, \ldots, \boldsymbol{X}_p \gg = \ll \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, \ldots, \boldsymbol{X}_p \gg$ por el lema 3.1. Como la proyección sobre un subespacio es única, se tiene que $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}$.

□

Teorema 3.3 *Dado el modelo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, sea $\tilde{\boldsymbol{X}}$ la matriz $\boldsymbol{X}$ donde una de sus columnas se ha modificado por el método de alzado. Sean $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ y $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$ los estimadores por MCO en ambos casos, respectivamente. Se verifica:*
(1) Sus correspondientes coeficientes de determinación R^2 y $\tilde{R}^2$, respectivamente, son iguales.
(2) Los estadísticos F y $\hat{F}$ son idénticos.

Demostración:
En efecto, se probará la igualdad de numeradores y denominadores en los coeficientes de determinación

$$R^2 = \frac{SCE}{SCT} \quad y \quad \tilde{R}^2 = \frac{S\tilde{C}E}{S\tilde{C}T}$$

Por una parte, $SCT = \sum(Y_i - \overline{Y})^2 = S\tilde{C}T$, puesto que la variable $\boldsymbol{Y}$ no se ha modificado.

Por otro lado, $SCE = \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\overline{Y}$ y $S\tilde{C}E = \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{X}}'\boldsymbol{Y} - n\overline{Y}$ expresiones que serán iguales si $\hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} = \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{X}}'\boldsymbol{Y}$. Pero esto es inmediato a partir del lema 3.2.

Además $SCR = S\tilde{C}R$ por lo que los estadísticos

$$F = \frac{SCE/p}{SCR/(n-p+1)} \quad y \quad \tilde{F} = \frac{S\tilde{C}E/p}{S\tilde{C}R/(n-p+1)}$$

también son iguales.

$\square$

En los epígrafes siguientes se estudia cuál es la mejor elección del "factor de alzamiento" λ_1.

4 Relación entre el factor de alzamiento de una variable y su FIV

Recordemos que en un modelo de regresión lineal ordinario se tiene:

$$
\begin{aligned}
SCT &= \sum_i (Y_i - \overline{Y})^2 = \|\boldsymbol{Y} - \overline{\boldsymbol{Y}}\|^2 \\
SCE &= \sum_i (\hat{Y}_i - \overline{Y})^2 = |\hat{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|^2 \\
SCR &= \sum_i e_{i\mathrm{o}}^2 = \|\boldsymbol{e}_{\mathrm{o}}\|^2
\end{aligned}
$$

donde $\overline{\boldsymbol{Y}} = (\overline{Y}, \overline{Y}, \ldots, \overline{Y})$.
Además, se sabe que $\|\boldsymbol{Y} - \overline{\boldsymbol{Y}}\|^2 = \|\hat{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|^2 + \|\boldsymbol{e}_{\mathrm{o}}\|^2$.

Por tanto, $SCT = SCE + SCR$ y, el coeficiente de determinación puede expresarse:

$$
R^2 = \frac{SCE}{SCT} = \frac{\|\hat{\boldsymbol{Y}} - \overline{\boldsymbol{Y}}\|^2}{\|\boldsymbol{Y} - \overline{\boldsymbol{Y}}\|^2} = 1 - \frac{SCR}{SCT} = 1 - \frac{\|\boldsymbol{e}_{\mathrm{o}}\|^2}{\|\boldsymbol{Y} - \overline{\boldsymbol{Y}}\|^2}
$$

Si nota por R_i^2 al coeficiente de determinación en la regresión por MCO de $\boldsymbol{X}_i$ sobre las restantes $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1}, \boldsymbol{X}_{i+1}, \ldots, \boldsymbol{X}_p\}$, los factores inflactores de la varianza, a partir de estas regresiones auxiliares, se definen como

$$
\mathrm{FIV}_i = \frac{1}{1 - R_i^2} = \frac{\|\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i\|^2}{\|\boldsymbol{e}_i\|^2}. \tag{1}
$$

A partir de la definición, y teniendo en cuenta que $\bar{\boldsymbol{e}}_i = (0, 0, \ldots, 0)'$ se pueden establecer las relaciones,

$$\begin{aligned}\overline{\tilde{\boldsymbol{X}}}_1 =& \overline{\boldsymbol{X}}_1 + \lambda_1 \overline{\boldsymbol{e}}_1 = \overline{\boldsymbol{X}}_1 \\ \overline{\hat{\boldsymbol{X}}}_1 =& \overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{e}}_1 = \overline{\boldsymbol{X}}_1\end{aligned}$$

Los módulos verifican

$$\boldsymbol{X}_1^2 = \hat{\boldsymbol{X}}_1^2 + \boldsymbol{e}_1^2 \quad \text{y} \quad \tilde{\boldsymbol{X}}_1^2 = \hat{\boldsymbol{X}}_1^2 + (\lambda_1 + 1)^2 \boldsymbol{e}_1^2 = \hat{\boldsymbol{X}}_1^2 + \delta_1^2 \boldsymbol{e}_1^2,$$

siendo $\delta_1 = \lambda_1 + 1$.

Proposición 4.1 *Entre las variables $\tilde{\boldsymbol{X}}_1$, $\boldsymbol{X}_1$, $\hat{\boldsymbol{X}}_1$ y $\overline{\boldsymbol{X}}_1$ se verifican las siguientes relaciones:*

$$\begin{aligned}(i) \qquad & (\tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 = (\hat{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 + \delta_1^2 \boldsymbol{e}_1^2 \\ (ii) \qquad & (\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^2 = (\hat{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 + \boldsymbol{e}_1^2 \\ (iii) \qquad & (\tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 - (\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^2 = (\delta_1^2 - 1) \boldsymbol{e}_1^2\end{aligned}$$

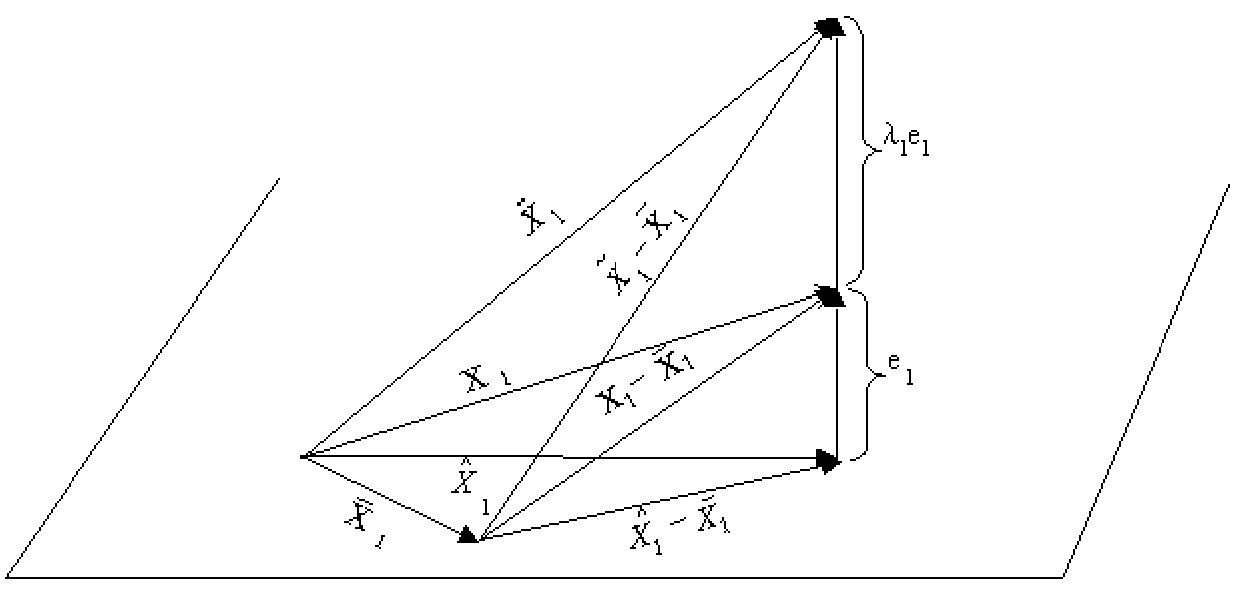

Figura 3: Relación entre una variable y su alzada

Demostración:
$\tilde{\boldsymbol{X}}_1 = \hat{\boldsymbol{X}}_1 + \delta_1 \boldsymbol{e}_1 \implies \tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1 = \hat{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1 + \delta_1 \boldsymbol{e}_1 \implies (\tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 = (\hat{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)^2 + \delta_1^2 \boldsymbol{e}_1^2 + 2\delta_1(\hat{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1)\boldsymbol{e}_1$. Pero $\hat{\boldsymbol{X}}_1 \perp \boldsymbol{e}_1 \implies \hat{\boldsymbol{X}}_1 \boldsymbol{e}_1 = 0$. Como $\overline{\boldsymbol{X}}_1 = (\overline{X}_1, \overline{X}_1, \ldots, \overline{X}_1)$ $\implies \overline{\boldsymbol{X}}_1 \boldsymbol{e}_1 = \sum_i X_1 e_{i1} = X_1 \sum_i e_{i1} = 0$ por ser $\overline{\boldsymbol{e}}_1 = 0$. Por tanto, la relación (i) es cierta.

De modo similar se puede probar la veracidad de (ii).

Siendo ciertas (i) e (ii) es inmediata (iii).

□

Al sustituir $\boldsymbol{X}_1$ por su alzada $\tilde{\boldsymbol{X}}_1$ y hacer la regresión de $\tilde{\boldsymbol{X}}_1$ sobre $\{\boldsymbol{X}_2, \ldots, \boldsymbol{X}_i, \ldots, \boldsymbol{X}_p\}$ se obtiene un nuevo coeficiente de determinación que notamos $\tilde{R}_1^2$ y da lugar a un FIV que se representa por $\tilde{\text{FIV}}_1$

$$\tilde{\text{FIV}}_1 = \frac{1}{1 - \tilde{R}_1^2} = \frac{\|\tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1\|^2}{\tilde{\boldsymbol{e}}_1^2} \tag{2}$$

donde $\tilde{\boldsymbol{e}}_1 = \tilde{\boldsymbol{X}}_1 - \hat{\boldsymbol{X}}_1 = \boldsymbol{e}_1 + \lambda_1 \boldsymbol{e}_1 = \delta_1 \boldsymbol{e}_1$.

Teorema 4.2 *Al alzar una variable disminuye su FVI*

Demostración:
Aplicando el apartado (iii) de la proposición 4.1, se verifica que:

$$\tilde{\text{FIV}}_1 = \frac{\|\tilde{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_1\|^2}{\tilde{\boldsymbol{e}}_1^2} = \frac{(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^2 + (\delta_1^2 - 1)\boldsymbol{e}_1^2}{\delta_1^2 \boldsymbol{e}_1^2} =$$

$$\frac{(\boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1)^2}{\delta_1^2 \boldsymbol{e}_1^2} + \frac{\delta_1^2 - 1}{\delta_1^2} = \frac{1}{\delta_1^2}\text{FIV}_1 + 1 - \frac{1}{\delta_1^2}$$

Luego,

$$\tilde{\text{FIV}}_1 - 1 = \frac{1}{\delta_1^2}(\text{FIV}_1 - 1) \tag{3}$$

y, dado que $\delta_1 > 1$, se deduce $\tilde{\text{FIV}}_1 \leq \text{FIV}_1$.

$\square$

Si K_1 es una cota para el $\tilde{\text{FIV}}_1$. Entonces,

$$\tilde{\text{FIV}}_1 = 1 + \frac{1}{\delta_1^2}(\text{FIV}_1 - 1) \ < \ K_1$$

Operando

$$\frac{1}{\delta_1^2} \ < \ \frac{K_1 - 1}{\text{FIV}_1 - 1}$$

$$\delta_1 > +\sqrt{\frac{FIV_1 - 1}{K_1 - 1}}, \, \delta_1 > 1; \quad \lambda_1 = \delta_1 - 1, \, \lambda_1 > 0 \tag{4}$$

El haber hecho el razonamiento sobre $\boldsymbol{X}_1$ no supone restricción alguna puesto que si en lugar de actuar sobre la variable $\boldsymbol{X}_1$ se actúa sobre $\boldsymbol{X}_j$, $j = 2, 3, \ldots, p$ la ecuación (4) quedaría:

$$\delta_j > \sqrt{\frac{FIV_j - 1}{K_j - 1}}, \, \delta_j > 1; \quad \lambda_j = \delta_j - 1, \, \lambda_j > 0 \tag{5}$$

5 Sesgo, varianza y ECM del estimador

Supóngase, para simplificar y sin que ello signifique restricción, que el modelo a estimar es $\boldsymbol{Y} = \beta_o + \beta_1 \boldsymbol{X}_1 + \beta_2 \boldsymbol{X}_2 + \beta_3 \boldsymbol{X}_3 + \boldsymbol{\varepsilon}$ con una muestra de tamaño 4. Su estimador por MCO es $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$. Sea $\boldsymbol{X}_1$ la variable elegida para alzar; el nuevo modelo es $\boldsymbol{Y} = \beta_o + \beta_1 \tilde{\boldsymbol{X}}_1 + \beta_2 \boldsymbol{X}_2 + \beta_3 \boldsymbol{X}_3 + \boldsymbol{\varepsilon}$ donde $\tilde{\boldsymbol{X}}_1 = \boldsymbol{X}_1 + \lambda_1 \boldsymbol{e}_1$ siendo $\boldsymbol{e}_1$ es el vector de residuos de la regresión por MCO de $\boldsymbol{X}_1$ sobre $\{\boldsymbol{X}_2, \boldsymbol{X}_3\}$. El estimador de MCO del modelo modificado es $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$.

Como consecuencia del lema 3.2, se puede asegurar que los vectores $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ y $\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}$ son iguales por ser la proyección del vector $\boldsymbol{Y}$ sobre el mismo espacio euclídeo que es única. Por consiguiente,

$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}} \tag{6}$$

Desarrollando (6), se obtiene:

$$\begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} \\ 1 & X_{21} & X_{22} & X_{23} \\ 1 & X_{31} & X_{32} & X_{33} \\ 1 & X_{41} & X_{42} & X_{43} \end{bmatrix} \begin{bmatrix} \hat{\beta}_o \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 1 & X_{11}+\lambda_1 e_{11} & X_{12} & X_{13} \\ 1 & X_{21}+\lambda_1 e_{21} & X_{22} & X_{23} \\ 1 & X_{31}+\lambda_1 e_{31} & X_{32} & X_{33} \\ 1 & X_{41}+\lambda_1 e_{41} & X_{42} & X_{43} \end{bmatrix} \begin{bmatrix} \tilde{\beta}_o \\ \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \tilde{\beta}_3 \end{bmatrix};$$

$$\left.\begin{array}{rcl} \hat{\beta}_o + X_{11}\hat{\beta}_1 + X_{12}\hat{\beta}_2 + X_{13}\hat{\beta}_3 & = & \tilde{\beta}_o + X_{11}\tilde{\beta}_1 + \lambda_1 e_{11} + X_{12}\tilde{\beta}_2 + X_{13}\tilde{\beta}_3 \\ \hat{\beta}_o + X_{21}\hat{\beta}_1 + X_{22}\hat{\beta}_2 + X_{23}\hat{\beta}_3 & = & \tilde{\beta}_o + X_{21}\tilde{\beta}_1 + \lambda_1 e_{21} + X_{22}\tilde{\beta}_2 + X_{23}\tilde{\beta}_3 \\ \hat{\beta}_o + X_{31}\hat{\beta}_1 + X_{32}\hat{\beta}_2 + X_{33}\hat{\beta}_3 & = & \tilde{\beta}_o + X_{31}\tilde{\beta}_1 + \lambda_1 e_{31} + X_{32}\tilde{\beta}_2 + X_{33}\tilde{\beta}_3 \\ \hat{\beta}_o + X_{41}\hat{\beta}_1 + X_{42}\hat{\beta}_2 + X_{43}\hat{\beta}_3 & = & \tilde{\beta}_o + X_{41}\tilde{\beta}_1 + \lambda_1 e_{41} + X_{42}\tilde{\beta}_2 + X_{43}\tilde{\beta}_3 \end{array}\right\}$$

Operando se obtiene

$$\left.\begin{array}{rcl} (\tilde{\beta}_o - \hat{\beta}_o) + X_{11}(\tilde{\beta}_1 - \hat{\beta}_1) + X_{12}(\tilde{\beta}_2 - \hat{\beta}_2) + X_{13}(\tilde{\beta}_3 - \hat{\beta}_3) & = & -\lambda_1 e_{11}\tilde{\beta}_1 \\ (\tilde{\beta}_o - \hat{\beta}_o) + X_{21}(\tilde{\beta}_1 - \hat{\beta}_1) + X_{22}(\tilde{\beta}_2 - \hat{\beta}_2) + X_{23}(\tilde{\beta}_3 - \hat{\beta}_3) & = & -\lambda_1 e_{21}\tilde{\beta}_1 \\ (\tilde{\beta}_o - \hat{\beta}_o) + X_{31}(\tilde{\beta}_1 - \hat{\beta}_1) + X_{32}(\tilde{\beta}_2 - \hat{\beta}_2) + X_{33}(\tilde{\beta}_3 - \hat{\beta}_3) & = & -\lambda_1 e_{31}\tilde{\beta}_1 \\ (\tilde{\beta}_o - \hat{\beta}_o) + X_{41}(\tilde{\beta}_1 - \hat{\beta}_1) + X_{42}(\tilde{\beta}_2 - \hat{\beta}_2) + X_{43}(\tilde{\beta}_3 - \hat{\beta}_3) & = & -\lambda_1 e_{41}\tilde{\beta}_1 \end{array}\right\}$$

que, expresado matricialmente es:

$$\boldsymbol{X}[\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}] = -\lambda_1 \tilde{\beta}_1 \boldsymbol{e}_1$$

Multiplicando por la izquierda por $\boldsymbol{X'}$

$$\boldsymbol{X'X}[\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}] = -\lambda_1 \tilde{\beta}_1 \boldsymbol{X'}\boldsymbol{e}_1$$

Premultiplicando por $(\boldsymbol{X'X})^{-1}$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}[\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}] = -\lambda_1 \tilde{\beta}_1 (\boldsymbol{X'X})^{-1}\boldsymbol{X'}\boldsymbol{e}_1$$

de donde

$$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} = -\lambda_1\tilde{\beta}_1(\boldsymbol{X'X})^{-1}\boldsymbol{X'e}_1$$

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \lambda_1\tilde{\beta}_1(\boldsymbol{X'X})^{-1}\boldsymbol{X'e}_1$$

Sea $\boldsymbol{a}_1$ el vector columna $(\boldsymbol{X'X})^{-1}\boldsymbol{X'e}_1$, $\boldsymbol{a}_1 = (a_{01}, a_{11}, a_{21}, a_{31})'$

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \lambda_1\tilde{\beta}_1\boldsymbol{a}_1 \tag{7}$$

Por lo tanto,

$$\left.\begin{array}{rcl}
\tilde{\beta}_0 & = & \hat{\beta}_0 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}}\hat{\beta}_1 a_{01} \\
\tilde{\beta}_1 & = & \hat{\beta}_1 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}}\hat{\beta}_1 a_{11} \\
\tilde{\beta}_2 & = & \hat{\beta}_2 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}}\hat{\beta}_1 a_{21} \\
\tilde{\beta}_3 & = & \hat{\beta}_3 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}}\hat{\beta}_1 a_{31}
\end{array}\right\} \tag{8}$$

Tomando valores esperados en (7)

$$\mathrm{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta} - \lambda_1\mathrm{E}[\tilde{\beta}_1]\boldsymbol{a}_1 \tag{9}$$

lo que prueba que $\tilde{\boldsymbol{\beta}}$ es un estimador sesgado.

$$\mathrm{E}[\tilde{\boldsymbol{\beta}}] = \boldsymbol{\beta} - \lambda_1\mathrm{E}[\tilde{\beta}_1]\begin{bmatrix} a_{01} \\ a_{11} \\ a_{21} \\ a_{31} \end{bmatrix}$$

Desarrollando

$$\left.\begin{array}{lcl}
\mathrm{E}[\tilde{\beta}_0] & = & \beta_0 - \lambda_1 \mathrm{E}[\tilde{\beta}_1] a_{01} \\
\mathrm{E}[\tilde{\beta}_1] & = & \beta_1 - \lambda_1 \mathrm{E}[\tilde{\beta}_1] a_{11} \\
\mathrm{E}[\tilde{\beta}_2] & = & \beta_2 - \lambda_1 \mathrm{E}[\tilde{\beta}_1] a_{21} \\
\mathrm{E}[\tilde{\beta}_3] & = & \beta_3 - \lambda_1 \mathrm{E}[\tilde{\beta}_1] a_{31}
\end{array}\right\}$$

de donde

$$\left.\begin{array}{lcl}
\mathrm{E}[\tilde{\beta}_0] & = & \beta_0 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}} \beta_1 a_{01} \\
\mathrm{E}[\tilde{\beta}_1] & = & \beta_1 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}} \beta_1 a_{11} \\
\mathrm{E}[\tilde{\beta}_2] & = & \beta_2 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}} \beta_1 a_{21} \\
\mathrm{E}[\tilde{\beta}_3] & = & \beta_3 - \dfrac{\lambda_1}{1+\lambda_1 a_{11}} \beta_1 a_{31}
\end{array}\right\} \tag{10}$$

Como $\tilde{\boldsymbol{\beta}}$ es un estimador de mínimos cuadrados,

$$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$$

$$\mathrm{var}(\tilde{\boldsymbol{\beta}}) = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\sigma^2 \tag{11}$$

que, evidentemente, dependen del factor de alzamiento.

La generalización a p variables explicativas es obvia. Si en lugar del modelo $\boldsymbol{Y} = \beta_o + \beta_1\boldsymbol{X}_1 + \beta_2\boldsymbol{X}_2 + \beta_3\boldsymbol{X}_3 + \boldsymbol{\varepsilon}$ se considera uno más general, $\boldsymbol{Y} = \beta_o + \beta_1\boldsymbol{X}_1 + \cdots + \beta_p\boldsymbol{X}_p + \boldsymbol{\varepsilon}$, y la variable alzada es $\boldsymbol{X}_k$, $1 \leq k \leq p$, su vector de residuos asociado $\boldsymbol{e}_k$ y $\boldsymbol{a}_k = (\boldsymbol{X'X})^{-1}\boldsymbol{X'e}_k$, la forma del estimador equivalente a (8) sería

$$\tilde{\beta}_j = \hat{\beta}_j - \frac{\lambda_k}{1+\lambda_k a_{kk}} \hat{\beta}_k a_{jk}, \quad j = 0, \ldots, p \tag{12}$$

y la correspondiente a (9)

$$\mathrm{E}[\tilde{\beta}_j] = \beta_j - \frac{\lambda_k}{1+\lambda_k a_{kk}} \beta_k a_{jk}, \quad j = 0, \ldots, p \tag{13}$$

Teorema 5.1 *Si las variables del modelo están centradas y normalizadas de modo que $\boldsymbol{X'X}$ sea la matriz de correlación, la componente a_{kk} del vector $\boldsymbol{a}_k$ es igual a uno.*

Demostración:
Si $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p]$ y $\boldsymbol{C} = (c_{ij}) = (\boldsymbol{X'X})^{-1}$ se sabe que $c_{ii} = 1/(1 - R_i^2)$ y que $R_i^2 = 1 - \boldsymbol{e}_k^2/\|\boldsymbol{X}_k - \overline{\boldsymbol{X}}_k\|^2 = 1 - \boldsymbol{e}_k^2$ puesto que $\|\boldsymbol{X}_k - \overline{\boldsymbol{X}}_k\|^2 = 1$. Por consiguiente, $c_{ii} = 1/\boldsymbol{e}_i^2$ y $\boldsymbol{X'e}_k$ tiene cero todas sus componentes excepto la k-ésima que vale $\boldsymbol{e}_k^2$. Luego,

$$\begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{kk} \\ \vdots \\ a_{pk} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2k} & \cdots & c_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{k1} & c_{k2} & \cdots & c_{kk} & \cdots & c_{kp} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{p1} & c_{p2} & \cdots & c_{pk} & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \boldsymbol{e}_k^2 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} c_{1k}\boldsymbol{e}_k^2 \\ c_{2k}\boldsymbol{e}_k^2 \\ \vdots \\ c_{kk}\boldsymbol{e}_k^2 \\ \vdots \\ c_{pk}\boldsymbol{e}_k^2 \end{bmatrix}$$

Por tanto, $a_{kk} = c_{kk}\boldsymbol{e}_k^2 = \dfrac{1}{\boldsymbol{e}_k^2}\boldsymbol{e}_k^2 = 1$.

□

Sesgo

Se puede suponer, sin pérdida de generalidad, que las columnas de la matriz $\boldsymbol{X}$ están centradas y normalizadas de modo que $\boldsymbol{X'X}$ es la matriz de correlación.

Si la variable alzada es $\boldsymbol{X}_k$, por el teorema 5.1, $a_{kk} = 1$ y de (13)[4] se sigue que

$$\mathrm{E}(\tilde{\beta}_i) - \beta_i = -\frac{\lambda}{1+\lambda}\beta_k a_{ik}$$

Por tanto, elevando al cuadrado y sumando,

$$\sum_i (\mathrm{E}[\tilde{\beta}_i] - \beta_i)^2 = \frac{\lambda^2}{(1+\lambda)^2}\beta_k^2 \sum_i a_{ik}^2 \tag{14}$$

[4]Para facilitar los cálculos se ha prescindido del subíndice de λ_k

que es el *cuadrado del sesgo.*

$$(\text{sesgo})^2 = \sum_i (\mathrm{E}[\tilde{\beta}_i] - \beta_i)^2 = \frac{\lambda^2}{(1+\lambda)^2}\beta_k^2 \|\boldsymbol{a}_k\|^2 \tag{15}$$

Varianza total

El concepto de varianza total está asociado a la expresión

$$\mathrm{var}(\tilde{\beta}) = \mathrm{E}[(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])'(\tilde{\boldsymbol{\beta}} - \mathrm{E}[\tilde{\boldsymbol{\beta}}])] = \sum_i (\tilde{\beta}_i - \mathrm{E}[\tilde{\beta}_i])^2$$

Si se resta a (12) la expresión (13), se eleva al cuadrado y se suma, resulta

$$\sum_i (\tilde{\beta}_i - \mathrm{E}[\tilde{\beta}_i])^2 =$$

$$\sum_i (\hat{\beta}_i - \beta_i)^2 + \frac{\lambda^2(\hat{\beta}_k - \beta_k)^2}{(1+\lambda)^2}\sum_i a_{ik}^2 - 2(\hat{\beta}_k - \beta_k)\frac{\lambda}{1+\lambda}\sum_i a_{ik}(\hat{\beta}_i - \beta_i)$$

Aplicando el operador esperanza se obtiene

$$\mathrm{var}(\tilde{\beta}) = \mathrm{var}(\hat{\beta}) + \frac{\lambda^2 \mathrm{var}(\hat{\beta}_k)}{(1+\lambda)^2}\sum_i a_{ik}^2 - 2\frac{\lambda}{1+\lambda}\sum_i a_{ik}\mathrm{cov}(\hat{\beta}_i, \hat{\beta}_k) \tag{16}$$

Error Cuadrático Medio

El ECM de $\tilde{\beta}$ viene dado por la expresión

$$\mathrm{ECM}(\tilde{\boldsymbol{\beta}}) = \mathrm{E}[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})] = \sum_i \mathrm{E}(\tilde{\beta}_i - \beta_i)^2$$

Si de (8) se resta el vector $\boldsymbol{\beta}$ se sigue que

$$\tilde{\beta}_i - \beta_i = \hat{\beta}_i - \beta_i - \frac{\lambda}{1+\lambda}a_{ik}\hat{\beta}_k = \hat{\beta}_i - \beta_i - \frac{\lambda}{1+\lambda}a_{ik}(\hat{\beta}_k - \beta_k) - \frac{\lambda}{1+\lambda}a_{ik}\beta_k$$

Elevando al cuadrado y sumando, resulta

$$\sum_i (\tilde{\beta}_i - \beta_i)^2 = \sum_i (\hat{\beta}_i - \beta_i)^2 + \frac{\lambda^2}{(1+\lambda)^2}\hat{\beta}_k^2 \sum_i a_{ik}^2 +$$

$$\frac{\lambda^2}{(1+\lambda)^2}(\hat{\beta}_k - \beta_k)^2 \sum_i a_{ik}^2 - 2\frac{\lambda}{1+\lambda}\sum_i a_{ik}\beta_k(\hat{\beta}_i - \beta_i)$$

$$-2\frac{\lambda}{1+\lambda}\sum_i a_{ik}(\hat{\beta}_k - \beta_k)(\hat{\beta}_i - \beta_i) + 2\frac{\lambda^2}{(1+\lambda)^2}\sum_i a_{ik}^2 \beta_k(\hat{\beta}_k - \beta_k)$$

Aplicando el operador esperanza matemática y teniendo en cuenta que $\hat{\boldsymbol{\beta}}$ es un estimador centrado, resulta

$$\sum_i \mathrm{E}(\tilde{\beta}_i - \beta_i)^2 = \sum_i \mathrm{E}(\hat{\beta}_i - \beta_i)^2 + \frac{\lambda^2}{(1+\lambda)^2}\mathrm{var}(\hat{\beta}_k)\sum_i a_{ik}^2$$

$$-2\frac{\lambda}{1+\lambda}\sum_i a_{ik}\mathrm{cov}(\hat{\beta}_k, \hat{\beta}_i) + \frac{\lambda^2}{(1+\lambda)^2}\beta_k^2 \|\boldsymbol{a}_k\|^2 \qquad (17)$$

que según (15) y (16) es la varianza más el cuadrado del sesgo, es decir,

$$\mathrm{ECM}(\tilde{\boldsymbol{\beta}}) = \mathrm{var}(\tilde{\boldsymbol{\beta}}) + (\mathrm{sesgo}(\tilde{\boldsymbol{\beta}}))^2 \qquad (18)$$

Teorema 5.2 *Existe un conjunto de valores del factor de alzamiento para los que el estimador de alzado es mejor que el de MCO en el sentido de que* $d(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}) < d(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$.

Demostración:

$$\sum_i (\tilde{\beta}_i - \beta_i)^2 - \sum_i (\hat{\beta}_i - \beta_i)^2 = \lambda_1^2 \Big(\frac{\hat{\beta}_1}{1+\lambda_1 a_{11}}\Big)^2 \sum_i a_{i1}^2 - 2\lambda_1 \frac{\hat{\beta}_1}{1+\lambda_1 a_{11}}\sum_i a_{i1}(\hat{\beta}_i - \beta_i).$$

Por otra parte, se tiene que:

$$\sum_i (\tilde{\beta}_i - \beta_i)^2 - \sum_i (\hat{\beta}_i - \beta_i)^2 < 0 \iff$$

$$\iff 0 < \lambda_1^2 \Big(\frac{\hat{\beta}_1}{1+\lambda_1 a_{11}} \Big)^2 \sum_i a_{i1}^2 \ < \ 2\lambda_1 \frac{\hat{\beta}_1}{1+\lambda_1 a_{11}} \sum_i a_{i1}(\hat{\beta}_i - \beta_i) \iff$$

$$\iff 0 < \lambda_1^2 (\frac{\hat{\beta}_1}{1+\lambda_1 a_{11}})^2 \sum_i a_{i1}^2 \ < \ 2\lambda_1 \frac{\hat{\beta}_1(1+\lambda_1 a_{11})}{(1+\lambda_1 a_{11})^2} \sum_i a_{i1}(\hat{\beta}_i - \beta_i)$$

$$\lambda_1^2 \hat{\beta}_1^2 \sum_i a_{i1}^2 \ - 2\lambda_1 \hat{\beta}_1 (1+\lambda_1 a_{11}) \sum_i a_{i1}(\hat{\beta}_i - \beta_i) \ < \ 0$$

$$\lambda_1^2 \Big(\hat{\beta}_1^2 \sum_i a_{i1}^2 \ - 2\hat{\beta}_1 a_{11} \sum_i a_{i1}(\hat{\beta}_i - \beta_i) \Big) - 2\lambda_1 \hat{\beta}_1 \sum_i a_{i1}(\hat{\beta}_i - \beta_i) \ < \ 0 \qquad (19)$$

La expresión (19) es un polinomio de segundo grado en λ_1 cuyas raíces son:

$$\lambda_{11} = 0; \quad \lambda_{12} = \frac{2b}{\hat{\beta}_1 \|a_1\|^2 \ - 2a_{11}b} \qquad (20)$$

Si $\hat{\beta}_1 \|a_1\|^2 \ - 2a_{11}b > 0$ la parábola (19) está abierta hacia arriba y los valores de λ_1 que hacen que $\mathrm{d}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}) \ < \ \mathrm{d}(\hat{\beta}, \boldsymbol{\beta})$ son los del intervalo $(0, \lambda_{12})$ si $\lambda_{12} > 0$.

En el caso de que $\hat{\beta}_1 \|\boldsymbol{a}_1\|^2 \ - 2a_{11}b < 0$, la parábola está abierta hacia abajo y los valores de λ_1 que hacen que $\mathrm{d}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta}) \ < \ \mathrm{d}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ son los del intervalo $(\lambda_{12}, +\infty)$ con $\lambda_{12} > 0$.

$\square$

Teorema 5.3 *Existe un factor de alzamiento único que hace mínima la distancia entre el vector de parámetros estimado y su estimador*

Demostración:
Restando $\boldsymbol{\beta}$ a ambos miembros de (7), desarrollando, elevando al cuadrado y sumando, resulta

$$\sum_i (\tilde{\beta}_i - \beta_i)^2 = \sum_i (\hat{\beta}_i - \beta_i)^2 + \lambda_1^2 \tilde{\beta}_1^2 \sum_i a_{i1}^2 - 2\lambda_1 \tilde{\beta}_1 \sum_i a_{i1}(\hat{\beta}_i - \beta_i) =$$

$$= \sum_i (\hat{\beta}_i - \beta_i)^2 + \lambda_1^2 (\frac{\hat{\beta}_1}{1+\lambda_1 a_{11}})^2 \sum_i a_{i1}^2 - 2\lambda_1 \frac{\hat{\beta}_1}{1+\lambda_1 a_{11}} \sum_i a_{i1}(\hat{\beta}_i - \beta_i)$$

Llamando $\tilde{L}^2 = \sum_i (\tilde{\beta}_i - \beta_i)^2$ y derivando ambos miembros respecto a λ_1, resulta

$$(\tilde{L}^2)' = 0 + \frac{2\lambda_1(1+\lambda_1 a_{11})^2 - 2\lambda_1^2 a_{11}(1+\lambda_1 a_{11})}{(1+\lambda_1 a_{11})^4}\hat{\beta}_1^2 \sum_i a_{i1}^2$$

$$-2\hat{\beta}_1 \frac{1}{(1+\lambda_1 a_{11})^2} \sum_i a_{i1}(\hat{\beta}_i - \beta_i)$$

Igualando la derivada a 0, sustituyendo $\sum_i a_{i1}^2$ por $\|\boldsymbol{a}_1\|^2$ y despejando λ_1, resulta

$$\lambda_1 = \frac{\sum_i a_{i1}(\hat{\beta}_i - \beta_i)}{\hat{\beta}_1 \|\boldsymbol{a}_1\|^2 - \sum_i a_{i1}(\hat{\beta}_i - \beta_i)} \tag{21}$$

Haciendo $b = \sum_i a_{i1}(\hat{\beta}_i - \beta_i)$, la expresión (21) da lugar a

$$\lambda_1 = \frac{b}{\hat{\beta}_1 \|\boldsymbol{a}_1\|^2 - b} \tag{22}$$

que es el valor del factor de alzamiento de la variable, en este caso $\boldsymbol{X}_1$, que minimiza $\mathrm{d}(\tilde{\boldsymbol{\beta}}, \boldsymbol{\beta})$.

Este resultado es fácilmente generalizable. Si en lugar de actuar sobre la variable $\boldsymbol{X}_1$ se elige $\boldsymbol{X}_k$, el correspondiente valor del factor de alzamiento es

$$\lambda_k = \frac{b}{\hat{\beta}_k \|\boldsymbol{a}_k\|^2 - b} \tag{23}$$

□

6 Alzado de vectores y Regresión Cresta

Supóngase que se está interesado en resolver un problema de regresión lineal del tipo $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (con las variables centradas y normalizadas). Así mismo y para simplificar, se acepta que sólo se dispone de tres variables independientes $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}$. Se sabe que la solución por MCO es $\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$, y a través del método de RC $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X'X} + k\boldsymbol{I})^{-1}\boldsymbol{X'Y}$. Si $\tilde{\boldsymbol{X}}$ es la transformada de la matriz $\boldsymbol{X}$ en el sentido de que se ha sustituido el vector $\boldsymbol{X}_1$ de $\boldsymbol{X}$ por su alzado, se puede probar que $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} = \boldsymbol{X'X} + \boldsymbol{K}$, donde $\boldsymbol{K} = \mathrm{diag}(k_1, 0, 0)$.

En efecto,

$$\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{\boldsymbol{X}}'_1 \\ \boldsymbol{X}'_2 \\ \boldsymbol{X}'_3 \end{bmatrix} [\tilde{\boldsymbol{X}}_1, \boldsymbol{X}_2, \boldsymbol{X}_3] = \begin{bmatrix} \boldsymbol{X}'_1 + \lambda_1\boldsymbol{e}'_1 \\ \boldsymbol{X}'_2 \\ \boldsymbol{X}'_3 \end{bmatrix} [\boldsymbol{X}_1 + \lambda_1\boldsymbol{e}_1, \boldsymbol{X}_2, \boldsymbol{X}_3] \tag{24}$$

Los elementos de la diagonal principal son

$$\tilde{\boldsymbol{X}}'_1\tilde{\boldsymbol{X}}_1 = (\boldsymbol{X}'_1 + \lambda_1\boldsymbol{e}'_1)(\boldsymbol{X}_1 + \lambda_1\boldsymbol{e}_1) = \boldsymbol{X}'_1\boldsymbol{X}_1 + \boldsymbol{X}'_1\lambda_1\boldsymbol{e}_1 + \lambda_1\boldsymbol{e}'_1\boldsymbol{X}_1 + \lambda_1\lambda_1\boldsymbol{e}'_1\boldsymbol{e}_1 =$$

$$\boldsymbol{X}'_1\boldsymbol{X}_1 + 2\lambda_1\boldsymbol{X}'_1\boldsymbol{e}_1 + \lambda_1^2\boldsymbol{e}'_1\boldsymbol{e}_1 = \boldsymbol{X}'_1\boldsymbol{X}_1 + k_1$$

donde $k_1 = 2\lambda_1\boldsymbol{X}'_1\boldsymbol{e}_1 + \lambda_1^2\boldsymbol{e}'_1\boldsymbol{e}_1$.

Por otra parte,

$$\tilde{\boldsymbol{X}}'_1\tilde{\boldsymbol{X}}_j = (\boldsymbol{X}'_1 + \lambda_1\boldsymbol{e}'_1)\boldsymbol{X}_j = \boldsymbol{X}'_1\boldsymbol{X}_j + \lambda_1\boldsymbol{e}'_1\boldsymbol{X}_j,\ j \neq 1$$

donde $\lambda_1\boldsymbol{e}'_1\boldsymbol{X}_j = 0$ porque $\boldsymbol{e}_1$ es ortogonal a $\boldsymbol{X}_j$, $1 \neq j$. Luego,

$$\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} = \boldsymbol{X'X} + \boldsymbol{K1}, \quad \text{donde } \boldsymbol{K1} = \mathrm{diag}(k_1, 0, 0) \tag{25}$$

Ahora el modelo a resolver es

$$\boldsymbol{Y} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{26}$$

y su solución por MCO

$$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$$

que conserva el coeficiente de determinación según el teorema 3.3.

Si en (26), se alza la variable $\boldsymbol{X}_2$ y se llama $\tilde{\tilde{\boldsymbol{X}}}$ a la matriz $\tilde{\tilde{\boldsymbol{X}}} = [\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2, \boldsymbol{X}_3]$, por similitud con (24), se verifica:

$$\tilde{\tilde{\boldsymbol{X}}}'\tilde{\tilde{\boldsymbol{X}}} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{K2}, \ \text{ donde } \boldsymbol{K2} = \text{diag}(0, k_2, 0) \tag{27}$$

Sustituyendo (25) en (27), resulta

$$\tilde{\tilde{\boldsymbol{X}}}'\tilde{\tilde{\boldsymbol{X}}} = \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{K2} = \boldsymbol{X'X} + \boldsymbol{K2} + \boldsymbol{K1} = \boldsymbol{X'X} + \text{diag}(k_1, k_2, 0) \tag{28}$$

Suponiendo que tenemos p variables, repitiendo el proceso reiteradamente, y llamando $\tilde{\boldsymbol{X}}^*$ a la matriz con todas las variables alzadas, se llega a que

$$\tilde{\boldsymbol{X}}'^*\tilde{\boldsymbol{X}}^* = \boldsymbol{X'X} + \boldsymbol{K}, \ \text{ donde } \boldsymbol{K} = \text{diag}(k_1, k_2, \ldots, k_p) \tag{29}$$

De este modo se pone de manifiesto la estrecha relación existente entre el Método de Alzado y la Regresión Cresta.

7 Ejemplo

En Rawlings [Raw88, p. 276] se propone el siguiente ejemplo para ilustrar la medida de la colinealidad. Se considera una matriz 20×4 que consta de una variable $\boldsymbol{X}_{\text{o}}$ de unos (correspondiente al término independiente), y tres variables independientes construidas de la siguiente forma:

$\boldsymbol{X}_1$ es la secuencia de números de 20 a 29 repetida.

$\boldsymbol{X}_2$ es $\boldsymbol{X}_1$ -25, con la 1^a y la 11^a observaciones cambiadas de -5 a -4 para evitar la dependencia lineal exacta.

$\boldsymbol{X}_3$ es una secuencia periódica de los números $5, 4, 3, 2, 1, 2, 3, 4, 5, 6$.
De este modo las variables quedan como sigue:

$\boldsymbol{X}_o$	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
1	20	-4	5
1	21	-4	4
1	22	-3	3
1	23	-2	2
1	24	-1	1
1	25	0	2
1	26	1	3
1	27	2	4
1	28	3	5
1	29	4	6
1	20	-4	5
1	21	-4	4
1	22	-3	3
1	23	-2	2
1	24	-1	1
1	25	0	2
1	26	1	3
1	27	2	4
1	28	3	5
1	29	4	6

Más adelante, en la p. 341, se hace referencia a este ejemplo para mostrar la aplicación del método de regresión cresta y se propone el siguiente modelo

$$Y_i = 10 + 0.4X_{i1} - 0.2X_{i2} + 0.4X_{i3} + \varepsilon_i$$

donde ε_i es generada a partir de una distribución $N(0, 1)$.

El vector de observaciones obtenido a través del modelo de simulación propuesto es

$$\boldsymbol{Y} = \begin{bmatrix} 21.607 \\ 21.409 \\ 20.386 \\ 20.946 \\ 17.942 \\ 21.808 \\ 21.358 \\ 21.883 \\ 22.596 \\ 22.483 \\ 20.389 \\ 21.351 \\ 21.374 \\ 19.734 \\ 19.817 \\ 20.337 \\ 19.761 \\ 23.266 \\ 21.600 \\ 23.477 \end{bmatrix}$$

Si se expresa el modelo en términos de variables centradas y normalizadas, se tiene

$$Y_i = 21.28 + 5.138TX_{i1} - 2.44TX_{i2} + 2.683TX_{i3} + \varepsilon_i$$

siendo $\boldsymbol{TX}_j$, $j = 1, 2, 3$ la variable $\boldsymbol{X}_j$ centrada y normalizada para que tenga módulo 1.

La matriz de correlación de las variables independientes es

	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$
$\boldsymbol{X}_1$	1	0.995591	0.290129
$\boldsymbol{X}_2$	0.995591	1	0.342177
$\boldsymbol{X}_3$	0.290129	0.342177	1

La presencia de colinealidad se puede detectar a través de distintos métodos: índices de condición, índices métricos o factores inflactores de la varianza.

Índices de condición: 1 27.34 1.692

Índices métricos: 0.0768 0.0754 0.7698

Factores inflactores de la varianza: 169.35 175.65 1.7

Todos los indicadores detectan, como no podía ser de otra manera, la presencia de colinealidad entre las variables del modelo.

1. Solución por el método de Regresión Cresta

En [Raw88, p. 343] este ejercicio está resuelto a través del método de regresión cresta, obteniéndose los resultados de la Tabla 1

Valores estimados						
k	$\hat{\beta}_1^*(k)$	$\hat{\beta}_2^*(k)$	$\hat{\beta}_3^*(k)$	R^2	$\|\hat{\boldsymbol{\beta}}^*(k)\|$	$\mathrm{d}(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}^*(k))$
0.0000	14.477	-13.177	4.493	0.6261	29.19	14.351
0.0050	5.743	-4.279	3.962	0.6125	22.70	2.348
0.0058	5.278	-3.805	3.932	0.6110	22.50	1.889
0.0100	3.782	-2.279	3.828	0.6056	21.97	1.814
0.0200	2.431	-0.899	3.712	0.6000	21.65	3.296
0.0400	1.622	-0.069	3.593	0.5960	21.54	4.348
0.0600	1.331	0.234	3.507	0.5939	21.51	4.734
0.0800	1.182	0.391	3.433	0.5817	21.49	4.930
0.1000	1.092	0.487	3.364	0.5792	21.47	5.047
0.2000	0.913	0.678	3.073	0.5753	21.43	5.270

Tabla 1: Ejemplo 1: Regresión cresta

Los valores estimados de los coeficientes para $k = 0$ corresponden al método de MCO.

En la Figura 4 se muestra la traza de la cresta.

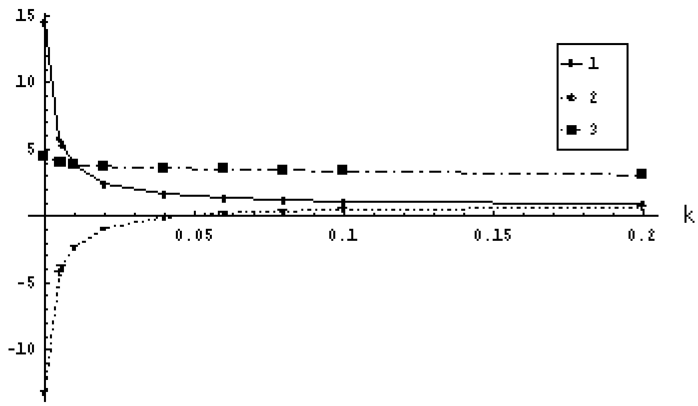

Figura 4: Traza de la cresta

En este trabajo se resuelve el mismo modelo utilizando dos métodos más: el de las componentes principales y el método de alzado.

2. Solución por el método de Componentes Principales.

Como es sabido, el método de componentes principales consiste en calcular, a través de la descomposición singular de la matriz $\boldsymbol{TX}$, tres nuevas variables $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$, $\boldsymbol{Z}_3$, cada una de ellas combinación lineal de $\boldsymbol{TX}_1$, $\boldsymbol{TX}_2$, $\boldsymbol{TX}_3$ y ortogonales entre sí. El prefijo "$\boldsymbol{T}$" que acompaña a los nombres de la matriz $\boldsymbol{X}$ y las variables $\boldsymbol{X}_i,\ i = 1, 2, 3$ significa "tipificada", es decir, centrada y normalizada de modo que cada columna tenga módulo uno.

$$\boldsymbol{TX} = \boldsymbol{U\Lambda V'} \quad \Longleftrightarrow \quad (\boldsymbol{TX})\boldsymbol{V} = \boldsymbol{U\Lambda}$$

donde $\boldsymbol{\Lambda}$ es una matriz diagonal cuyos elementos son los valores singulares de $\boldsymbol{TX}$ y, $\boldsymbol{U}$ y $\boldsymbol{V}$ son las matrices ortogonales de vectores singulares normalizados de $\boldsymbol{TX}$ por la izquierda y derecha respectivamente.

$$[\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{Z}_3] = [\boldsymbol{TX}_1, \boldsymbol{TX}_2, \boldsymbol{TX}_3] \cdot \boldsymbol{V}$$

donde

$$\begin{array}{lll} \overline{X}_1 = 24.5 & \Sigma_{i=1}^{20}(X_{i1} - \overline{X}_1)^2 = 165 & TX_{i1} = (X_{i1} - 24.5)/12.845 \\ \overline{X}_2 = -0.4 & \Sigma_{i=1}^{20}(X_{i2} - \overline{X}_2)^2 = 148.8 & TX_{i2} = (X_{i2} + 0.4)/12.198 \\ \overline{X}_3 = 3.5 & \Sigma_{i=1}^{20}(X_{i3} - \overline{X}_3)^2 = 45 & TX_{i3} = (X_{i3} - 3.5)/6.708 \end{array}$$

La matriz $\boldsymbol{V}$ es

$$\boldsymbol{V} = \begin{bmatrix} 0.655938 & -0.699977 & 0.282450 \\ 0.664552 & 0.712987 & 0.223649 \\ 0.357932 & -0.041003 & -0.932847 \end{bmatrix}$$

y las componentes principales

$$\begin{aligned} \boldsymbol{Z}_1 &= 0.655938 * \boldsymbol{TX}_1 + 0.664552 * \boldsymbol{TX}_2 + 0.357932 * \boldsymbol{TX}_3 \\ \boldsymbol{Z}_2 &= -0.699977 * \boldsymbol{TX}_1 + 0.712987 * \boldsymbol{TX}_2 - 0.041003 * \boldsymbol{TX}_3 \\ \boldsymbol{Z}_3 &= 0.282450 * \boldsymbol{TX}_1 + 0.223649 * \boldsymbol{TX}_2 - 0.932847 * \boldsymbol{TX}_3 \end{aligned}$$

donde $\boldsymbol{Z}_1$ explica el 72.23% de la variabilidad de $\boldsymbol{Y}$, $\boldsymbol{Z}_2$ el 0.1% y $\boldsymbol{Z}_3$ el 27.67%, de modo que $\boldsymbol{Z}_1 + \boldsymbol{Z}_3$ explican el 99.9%. Por tanto, haciendo la regresión de $\boldsymbol{Y}$ sobre $\boldsymbol{Z}_1$, $\boldsymbol{Z}_3$ se obtiene la Tabla 2 en la que se muestra que el modelo está globalmente bien estimado y los coeficientes de pendiente son significativos al 95%

que nos proporciona la ecuación

$$\boldsymbol{Y}^{cp} = 21.1762 + 2.347290 * \boldsymbol{Z}_1 - 3.049116 * \boldsymbol{Z}_3$$

en función de las variables $\boldsymbol{Z}_1$ y $\boldsymbol{Z}_3$, o bien

$$\boldsymbol{Y}^{cp} = 21.1762 + 0.678454 * \boldsymbol{TX}_1 + 0.877965 * \boldsymbol{TX}_2 + 3.68453 * \boldsymbol{TX}_3$$

en función de las variables tipificadas e

$$\boldsymbol{Y}^{cp} = 17.9885 + 0.052818 * \boldsymbol{X}_1 + 0.071974 * \boldsymbol{X}_2 + 0.549274 * \boldsymbol{X}_3$$

LS // Dependent Variable is $\boldsymbol{Y}$

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21.17620	0.199556	106.1167	0.0000
$\boldsymbol{Z}_1$	2.347290	0.606249	3.871824	0.0012
$\boldsymbol{Z}_3$	-3.049116	0.979511	-3.112896	0.0063

R-squared	0.592142	Mean dependent var	21.17620
Adjusted R-squared	0.544158	S.D. dependent var	1.321819
S.E. of regression	0.892440	F-statistic	12.34057
Sum squared resid	13.53964	Prob(F-statistic)	0.000489

Tabla 2: Regresión por MCO sobre CP

en función de las variables originales.

En cualquiera de las ecuaciones, los coeficientes estimados se alejan mucho de los coeficientes del modelo lo que indica que, en este caso, este método no es el más adecuado.

3. Solución a través del Método de Alzado

En la detección de la colinealidad se ha visto que la variable "más colineal" con las demás es $\boldsymbol{X}_2$. Por tanto, se procederá al alzado de dicha variable .

Dado que al ser un modelo teórico se conocen los valores de los parámetros, se utilizarán varios factores de alzamiento λ, se calcularán los valores estimados $\tilde{\beta}_i(\lambda),\ i = 1, 2, 3$ ($\tilde{\beta}_o$ se estima como $\bar{\boldsymbol{Y}}$) y se determinará para que factor de alzamiento se obtiene la mejor predicción, en el sentido de que $\mathrm{d}(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}})$ sea mínima.

Tal y como se observa en la Tabla 3, la mejor estimación del vector de parámetros se obtiene para $\lambda = 3.40$.

Se puede comprobar que el valor del factor de alzamiento calculado por este método de tanteo coincide con el que se obtiene aplicando directamente la fórmula (23) de la

Valores originales de los coeficientes

	β_0	β_1	β_2	β_3	$\|\boldsymbol{\beta}\|$
VO	21.28	5.138	-2.44	2.683	22.190

Valores estimados

λ	$\tilde{\beta}_0$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	$\tilde{\beta}_3$	$\|\tilde{\boldsymbol{\beta}}\|$	$d(\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}})$
2.50	21.176	5.265	-3.765	3.945	22.49	1.837
2.75	21.176	5.020	-3.514	3.930	22.39	1.653
3.00	21.176	4.805	-3.294	3.917	22.31	1.541
3.30	21.176	4.580	-3.064	3.904	22.23	1.484
3.40	**21.176**	**4.512**	**-2.995**	**3.900**	**22.20**	**1.480**
3.50	21.176	4.447	-2.928	3.896	22.18	1.483
3.60	21.176	4.384	-2.865	3.892	22.15	1.491
3.75	21.176	4.295	-2.774	3.887	22.12	1.511
4.00	21.176	4.160	-2.635	3.897	22.08	1.561
4.50	21.176	3.925	-2.396	3.865	22.01	1.697

Tabla 3: Ejemplo 1: Método de Alzado

página 18. En este caso $\boldsymbol{a}_2 = (-0.978698, 1, 0.058228)$, $b = \sum_i a_{i2}(\hat{\beta}_i - \beta_i) = -19.98245$, $\|\boldsymbol{a}_2\|^2 = 1.9612$, $\lambda_2 = b/(\hat{\beta}_2 \|\boldsymbol{a}_2\|^2 - b) = 3.4097$.

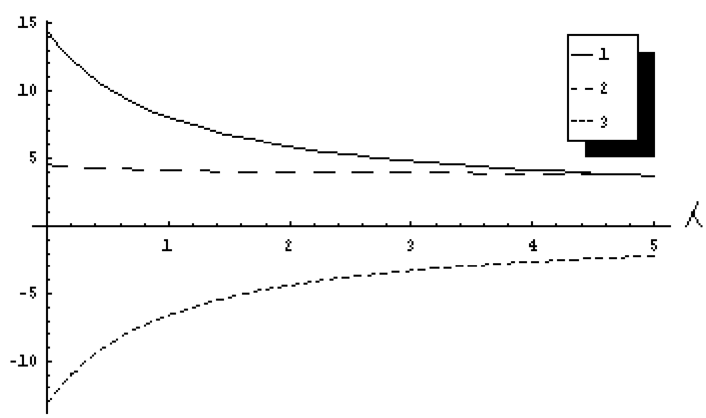

Figura 5: Variación de los coeficientes con el factor de alzamiento

En la Figura 5 se muestra la evolución de los estimadores de los parámetros en función del factor de alzamiento λ.

7.1 Comparación de los distintos métodos

VO: Valores originales.

MCO: Valores estimados por Mínimos Cuadrados Ordinarios.

RC: Valores estimados por Regresión Cresta, RC^{a}[5] y RC^{b}[6].

CP: Valores estimados por Componentes Principales.

AV: Valores estimados por el Método de Alzado.

Valores originales de los coeficientes

	β_0	β_1	β_2	β_3	$\|\boldsymbol{\beta}\|$
VO	21.28	5.138	-2.44	2.683	22.190

Valores estimados de los coeficientes

Método	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\|\hat{\boldsymbol{\beta}}\|$	$d(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})$
MCO	21.176	14.477	-13.177	4.493	29.186	14.345
RC^a	21.176	5.278	-3.805	3.932	22.499	1.858
RC^b	21.176	3.782	-2.279	3.828	21.967	1.785
CP	21.176	0.678	0.878	3.685	21.523	5.649
AV	**21.176**	**4.512**	**-2.995**	**3.900**	**22.203**	**1.480**

Tabla 4: Comparación de estimadores

Del análisis de los resultados que se muestran en la Tabla 4 se podría aceptar que el método que proporciona mejor estimación de los valores de los parámetros es el de *ALZADO DE VARIABLES*.

[5] $k = 0.0058$, [HKB75]

[6] $k = 0.01$

Referencias

[BB75] Brown, William G. y Beattie, Bruce R. Improving estimates of economics parameters by use of ridge regression with production function applications. *American Journal of Agricultural Economics*, 57(1):21–32, 1975.

[Cas85] Casella, George. Condition numbers and minimax ridge regression estimators. *Journal of the American Statistical Association*, 80(391):753–758, 1985.

[FG67] Farrar, D.E. y Glauber, R.R. Multicollinearity in regression analysis: The problem revisited. *Review of Economics and Statistics*, 49:92–107, 1967.

[FGP84] Fourgeaud, C., Gourieux, C., y Pradel, J. Some theoretical results for generalized ridge regression estimators. *Journal of Econometrics*, 25(2):191–203, 1984.

[Gar77] García Ferrer, A. El problema de la multicolinealidad en los modelos de regresión lineales: Algunas soluciones posibles. *Revista Española de Economía*, Mayo-Agosto,77:120–139, 1977.

[GAS99] García, José, Andújar, Antonio, y Soto, José. Acerca de la detección de la colinealidad en modelos de regresión lineal. XIII Reunión Asepelt España.- Burgos, 1999.

[Hem75] Hemmerle, W.J. An explicit solution for generalized ridge regression. *Technometrics*, 17(3):309–314, 1975.

[HK70a] Hoerl, A.E. y Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, a1970.

[HK70b] Hoerl, A.E. y Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12:69–82, b1970.

[HKB75] Hoerl, A.E., Kennard, R.W., y Baldwin, K.F. Ridge regression. Some simulations. *Communications in Statistics*, 4(2):105–123, 1975.

[HSL76] Hocking, R.R., Speed, F.M., y Lynn, M.J. A class of biased estimators in linear regression. *Technometrics*, 18:425–438, 1976.

[Joh89] Johnston, J. *Métodos de econometría*. Vicens-Vives, Barcelona, 1989.

[Ken57] Kendall, M.G. *A Course in Multivariate Analysis*. Griffin, London, 1957.

[Mar70] Marquardt, D.W. Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics*, 12(3):590–612, 1970.

[Mas65] Massy, William F. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*, 60:234–256, 1965.

[MG75] McDonald, Gary G. y Galarneau, Diane I. A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70(350):407–416, 1975.

[Pol78] Polverini, Francesco. Multicollinerita e stimatori "ridge"nel modello classico di regressione lineare. *Giornale degli Economisti e Annali di Economia*, 37(1-2):89–112, 1978.

[Rao62] Rao, C.R. A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *Journal of the Royal Statistical Society*, B(24):152–158, 1962.

[Raw88] Rawlings, John O. *Applied Regression Analysis: A Research Tool*. Cole Statistics and Probability Series. The Wadsworth and Brooks, California, 1988.

[SGA00] Soto, José, García, José, y Andújar, Antonio S. Acerca de la corrección de la multicolinealidad en modelos de regresión lineal. XIV Reunión de Asepelt España.- Oviedo, 2000.

[SGA01] Soto, José, García, José, y Andújar, Antonio S. Estimación de parámetros en presencia de cuasi-colinealidad. Método de alzado. XV Reunión de Asepelt España.- La Coruña, 2001.

[Sil69] Silvey, S.D. Multicollinearity and imprecise estimation. *Journal of Statistical Society*, 31(Ser. B):539–552, 1969.