Regresión alzada y el número de condición: algunos problemas

Salmerón Gómez, Román (romansg@ugr.es)
Departamento de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada
García García, Catalina (cbgarcia@ugr.es)
Departamento de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada
García Pérez, José (jgarcia@ual.es)
Departamento de Economía y Empresa
Universidad de Almería
López Martín, María del Mar (mariadelmarlopez@ugr.es)
Departamento de Didáctica de las Matemáticas
Universidad de Granada

RESUMEN

La regresión lineal múltiple es ampliamente usada cuando se desean establecer relaciones entre variables. Cuando las variables independientes usadas presentan una alta relación lineal el análisis realizado es inestable por lo que las conclusiones obtenidas quedan en entredicho. Este problema es conocido como multicolinealidad aproximada. La literatura recoge diversas opciones para abordar este problema, siendo la más usada la eliminación del análisis de las variables que se consideran causantes de la multicolinealidad. Como alternativa, la Regresión Alzada es una técnica cuantitativa que mitiga el problema de la multicolinealidad desde un punto de vista geométrico aumentando el ángulo existente entre las variables independientes del modelo. Tras aplicar dicha técnica, es conveniente comprobar si el problema ha sido o no mitigado, por lo que se hace necesario usar algunas de las herramientas existentes para detectar la multicolinealidad. En el presente trabajo se presenta el uso del Número de Condición en la Regresión Alzada focalizando en dos problemas que surgen: la dificultad de la obtención de una expresión algebraica cerrada y la sensibilidad de esta medida a las transformaciones en los datos.

Palabras clave: multicolinealidad, número de condición, regresión alzada.
Área temática: Aspectos Cuantitativos de Problemas Económicos y Empresariales con certeza.

ABSTRACT

Multiple linear regression analysis is a methodology widely applied in many different fields to establish relations between variables. When the independent variables present a high linear relation between them, the analysis is unstable and the conclusions may be questionable. This problem is known as approximate multicollinearity. The literature offers several options to solve this question being traditionally the most applied to eliminate the variables that are considered to cause multicollinearity. The raised regression is a quantitative technique that mitigates the problem of multicollinearity from a geometrical point of view. After applying this technique, it is recommendable to check if the collinearity has been mitigated or not. In this paper, it is presented the use of the condition number in the raise regression focusing on two problems that arise when this extension is addressed: the difficulty of obtaining a closed algebraic expression and the sensitivity of this measure to transformations in the original data.

1 INTRODUCCIÓN

Dado el siguiente modelo con n observaciones y p variables:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

donde se asume que se verifican las hipótesis básicas del modelo de regresión lineal múltiple, cuando se verifica que las variables independientes están relacionadas linealmente de forma aproximada, se dice que hay multicolinealidad aproximada en el modelo. Cuando el grado de multicolinealidad es alto, la estimación por Mínimos Cuadrados Ordinarios (MCO) es inestable y puede proporcionar resultados erróneos. Por tanto, un diagnóstico adecuado de este problema es esencial para una correcta interpretación del modelo (1). En este sentido, adviértase que en la práctica siempre hay multicolinealidad, la clave reside en establecer cuándo el grado de multicolinealidad existente es preocupante. En la literatura especializada existen distintas medidas para diagnosticar la presencia de multicolinealidad, siendo las más extendidas el Factor de Inflación de la Varianza, FIV (Marquardt 1970, Theil 1971 y Fox y Monette 1992) y el Número de Condición, NC (Belsley 1982, Belsley y otros 1980 y Belsley 1991).

Aparte del diagnóstico de la multicolinealidad, otro tema interesante es la estimación del modelo (1) cuando existe multicolinealidad grave. Igualmente existen distintos métodos que abordan esta cuestión, por ejemplo, la Regresión Cresta, RC (Hoerl y Kennard 1970a, 1970b) y la Regresión Alzada, RA (García y otros 2010). Una vez aplicados estos métodos es conveniente comprobar si la multicolinealidad ha sido mitigada. Por tal motivo, se hace necesario extender las medidas de detección a estas técnicas, es decir, establecer cómo se han de calcular el FIV o el NC en la RC o en la RA. En García y otros 2015a y Salmerón y otros 2016a, 2016b se aborda la adaptación del FIV y NC a RC, siendo el objetivo del presente trabajo la extensión del NC a RA.

La estructura del trabajo es como sigue: en la sección 2 se presenta la Regresión Alzada, la cual se basa en una visión geométrica de la multicolinealidad ya que consiste en aumentar el ángulo existente entre las variables independientes de un modelo disminuyendo así el grado de multicolinealidad. En la sección 3 se recuerda brevemente cómo calcular el Número de Condición y se aborda su extensión a la Regresión Alzada mostrando la dificultad existente en la obtención de una expresión algebraica cerrada. En la sección 4 se ponen en práctica los desarrollos anteriores mostrando al mismo tiempo la sensibilidad de esta medida a las transformaciones en los datos. Finalmente, en la sección 5 se destacan las principales conclusiones del trabajo.

2 REGRESIÓN ALZADA

Dado el modelo (1) la regresión alzada consiste en estimar por MCO el modelo:

$$\mathbf{Y} = \widetilde{\mathbf{X}}\boldsymbol{\alpha} + \mathbf{v}, \tag{2}$$

donde:

$$\widetilde{\mathbf{X}} = \mathbf{X} + \lambda \cdot (0 \dots \mathbf{e}_i \dots 0) = \left(\mathbf{X}_1 \dots \widetilde{\mathbf{X}}_i \dots \mathbf{X}_p\right),$$

siendo $\widetilde{\mathbf{X}}_i = \mathbf{X}_i + \lambda \mathbf{e}_i$ con $\lambda \geq 0$ y siendo $\mathbf{e}_i$ los residuos obtenidos al estimar por MCO la regresión auxiliar:

$$\mathbf{X}_i = \mathbf{X}_{-i}\boldsymbol{\delta} + \mathbf{w}, \tag{3}$$

siendo $\mathbf{X}_{-i}$ igual a la matriz $\mathbf{X}$ tras eliminar la variable $\mathbf{X}_i$. En este caso se dirá que se ha alzado la variable $\mathbf{X}_i$. Adviértase que $\mathbf{e}_i$ y $\mathbf{X}_{-i}$ son ortogonales, esto es, $\mathbf{X}_j^t \mathbf{e}_i = 0$ para $j = 1, \dots, p$ con $j \neq i$.

Por tanto, para cada valor de λ se tendrá un estimador distinto de los coeficientes de las variables del modelo (2), es decir, la estimación dependerá de dicho parámetro (que llamaremos parámetro de alzado). En García y otros 2016 se desarrolla el caso en el que $p = 2$, mostrándose que el coeficiente de determinación, la estimación de la varianza de la perturbación aleatoria, el valor de la F experimental asociada al contraste de significación conjunta y la t experimental asociada al contraste de significación individual de la variable alzada no dependen de λ, es decir, permanecen constantes. Por otro lado, la t experimental asociada al contraste de significación individual de la variable intacta si depende de λ, por lo que variará con los valores de este parámetro. También se obtiene el FIV, el cual también depende del parámetro de alzado y verifica las siguientes propiedades:

1. Cuando $\lambda = 0$ coincide con el FIV obtenido para el modelo (1), es decir, es continuo en cero.

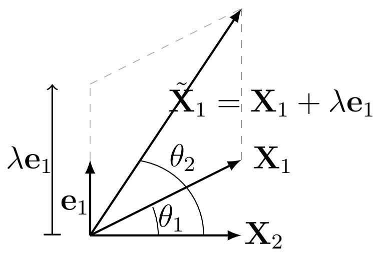

Figura 1: Representación del método de alzado para $p = 2$

2. Es decreciente en λ, es decir, conforme aumenta el valor de este parámetro disminuye el grado de multicolinealidad en el modelo.

3. Es siempre mayor que 1, que es el mínimo valor que puede tomar el FIV.

Estas propiedas se verifican también en la extensión realizada del FIV a RC por García y otros 2015c. De igual forma sería deseable que se verificaran también en la extensión que se plantea en la siguiente sección del NC a la RA.

Finalmente, en la Figura 1 se muestra la representación del método de alzado cuando $p = 2$ desde un punto de vista geométrico. Se puede observar claramente que conforme aumenta λ el ángulo existente entre la variable alzada y la variable intacta aumenta, por lo que el grado de multicolinealidad disminuye. De ahí la importancia de que se verifique la segunda propiedad comentada anteriormente.

3 EL NÚMERO DE CONDICIÓN EN LA REGRESIÓN ALZADA

Dado el modelo de regresión lineal múltiple (1), el Número de Condición (ver, por ejemplo, Belsley 1982) se define como:

$$K(\mathbf{X}) = \frac{\mu_{max}}{\mu_{min}}, \tag{4}$$

donde μ_{max} y μ_{min} son, respectivamente, los valores singulares máximo y mínimo de la matriz $\mathbf{X}$.

Puesto que la descomposición en valores singulares de $\mathbf{X}$ corresponde a $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^t$ donde $\mathbf{U}^t\mathbf{U} = \mathbf{I}$, $\mathbf{V}^t\mathbf{V} = \mathbf{I}$ siendo $\mathbf{I}$ la matriz identidad (de dimensiones adecuadas en cada caso) y $\mathbf{D} = diag(\mu_1 \dots \mu_p)$, con μ_i los valores singulares de $\mathbf{X}$, $i = 1, \dots, p$, se tiene que:

$$\mathbf{X}^t\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}^t\mathbf{U}\mathbf{D}\mathbf{V}^t = \mathbf{V}\mathbf{D}^2\mathbf{V}^t.$$

En tal caso se verifica que los autovalores de la matriz $\mathbf{X}^t\mathbf{X}$ coinciden con el cuadrado de los valores singulares de $\mathbf{X}$, es decir, $\xi_i = \mu_i^2$ para $i = 1, \dots, p$. Entonces, la expresión (4) es equivalente a:

$$K(\mathbf{X}) = \sqrt{\frac{\xi_{max}}{\xi_{min}}}, \tag{5}$$

donde ξ_{max} y ξ_{min} son, respectivamente, los autovalores máximo y mínimo de la matriz $\mathbf{X}^t\mathbf{X}$.

Según Belsley 1982 valores de $K(\mathbf{X})$ entre 20 y 30 suponen multicolinealidad moderada y valores superiores a 30 multicolinealidad grave. Además, los datos deben tener longitud unidad, es decir, han de ser divididos por su norma.

Atendiendo a lo expuesto, para calcular el NC en la RA, modelo (2), hay que calcular los autovalores de la matriz:

$$\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}} = \mathbf{X}^t\mathbf{X} + A, \tag{6}$$

donde:

$$A = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & \lambda(\lambda+2)\mathbf{e}_i^t\mathbf{e}_i & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}.$$

Justo aquí aparece el primero de los problemas referenciados en el resumen: ¿es posible expresar explicitamente los autovalores de $\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}}$ en función de los autovalores de $\mathbf{X}^t\mathbf{X}$ y A?

Evidentemente, desde un punto de vista computacional no tiene mayor complicación calcular los autovalores de la matriz:

$$\widetilde{\mathbf{X}}^t\widetilde{\mathbf{X}} = \begin{pmatrix} \sum\limits_{j=1}^{n} \mathbf{X}_{1j}^2 & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{1j}\mathbf{X}_{ij} & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{1j}\mathbf{X}_{kj} \\ \vdots & & \vdots & & \vdots \\ \sum\limits_{j=1}^{n} \mathbf{X}_{ij}\mathbf{X}_{1j} & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{ij}^2 + \lambda(\lambda+2)\mathbf{e}_i^t\mathbf{e}_i & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{ij}\mathbf{X}_{kj} \\ \vdots & & \vdots & & \vdots \\ \sum\limits_{j=1}^{n} \mathbf{X}_{kj}\mathbf{X}_{1j} & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{kj}\mathbf{X}_{ij} & \cdots & \sum\limits_{j=1}^{n} \mathbf{X}_{kj}^2 \end{pmatrix}.$$

Ahora bien, este cálculo puramente numérico no permite comprobar si la extensión del NC a la RA, que claramente depende del parámetro de alzado, verifica las tres propiedades comentadas anteriormente.

4 EJEMPLO NUMÉRICO

Para ilustrar los desarrollos anteriores usaremos los datos recogidos por Wissell 2009. El modelo (1) está formado por la información disponible para las siguientes variables desde el año 1995 a 2011: deuda pendiente de hipoteca, $\mathbf{Y}$, consumo personal, $\mathbf{X}_1$, ingreso personal, $\mathbf{X}_2$, y crédito al consumo pendiente, $\mathbf{X}_3$. Todas las variables están medidas en millones de dólares. En la Tabla 1 se muestran los datos de cada variable.

La matriz de correlaciones de las variables independientes es la siguiente:

$$\begin{pmatrix} 1 & 0.998 & 0.997 \\ 0.998 & 1 & 0.994 \\ 0.997 & 0.994 & 1 \end{pmatrix}.$$

Las altas correlaciones entre las variables sugieren que en el modelo considerado hay un alto grado de multicolinealidad.

Si bien Belsley 1982 indica que los datos han de tener longitud unidad, existen otras transformaciones que deben ser consideradas. Para analizar la sensibilidad del NC a la transformación en los datos se consideraran los siguientes siete escenarios:

S.1 Datos originales (sin transformar).

S.2 Datos no centrados con varianza igual a uno.

S.3 Datos no entrados con varianza igual a uno entre el número de observaciones.

S.4 Datos centrados.

S.5 Datos centrados con varianza igual a uno (tipificación).

S.6 Datos entrados con varianza igual a uno entre el número de observaciones (estandarización).

S.7 Datos no centrados divididos por la raíz cuadrada de la suma de sus cuadrados (longitud unidad).

Así, las Tablas 2 a 5 muestran los números de condición obtenidos para los modelos (1) y (2), en este segundo caso, para cada valor del parámetro de alzado, λ, suponiendo que se ha alzado cada una de las variables[1]. Se observa que:

- En todos los casos se verifican las propiedades de continuidad en cero, es decir, el NC asociado a la RA para $\lambda = 0$ coincide con el obtenido en la regresión original.
- Es decreciente en λ, es decir, conforme aumenta el parámetro de alzado el NC disminuye.

[1]En García y otros 2016 se sugieren algunos criterios para seleccionar la variable a alzar. Puesto que este no es el objetivo del presente trabajo se alzan todas y cada una de las variables.

Año	$\mathbf{Y}$	$\mathbf{X}_1$	$\mathbf{X}_2$	$\mathbf{X}_3$
1995	3.8051	4.7703	4.8786	8.0823
1996	3.9458	4.7784	5.051	7.9803
1997	4.0579	4.9348	5.362	8.0612
1998	4.1913	5.0998	5.5585	8.6565
1999	4.3585	5.2907	5.8425	9.973
2000	4.5453	5.4335	6.1523	11.407
2001	4.8149	5.6194	6.5206	12.534
2002	5.1286	5.8318	6.9151	13.248
2003	5.6151	6.1258	7.423	14.205
2004	6.2249	6.4386	7.8024	15.321
2005	6.7864	6.7394	8.4297	17.175
2006	7.4944	6.9104	8.7241	18.672
2007	8.3993	7.0993	8.8819	19.741
2008	9.3951	7.2953	9.1636	20.78
2009	10.68	7.5614	9.7272	21.913
2010	12.071	7.8036	10.301	22.849
2011	13.44821	8.0441	10.983	23.875

Tabla 1: Valores de la deuda pendiente de hipoteca, consumo personal, ingreso personal y crédito al consumo pendiente

S.1	S.2	S.3	S.4	S.5	S.6	S.7
12592.757	113.286	113.286	12947.962	51.728	51.728	105.019

Tabla 2: Número de Condición para el modelo original dado en (1)

- Siempre es mayor a uno.
- Sin embargo, al no tener una expresión algebraica no se puede estudiar si estas tres propiedades se verifican siempre.
- El NC para las tranformaciones S.2 y S.3 coincide. Al igual que para las transformaciones S.5 y S.6.
- Los resultados obtenidos para las transformaciones S.1 y S.4 no son razonables.
- Alzando la primera variable, para las transformaciones S.1, S.2, S.3, S.4 y S.7 se tiene que todos los valores obtenidos para el NC son superiores a 30, por lo que la RA no habría mitigado la multicolinealidad grave del modelo (1). Para las transformaciones S.5 y S.6 se tendría que para $\lambda = 0.8, 0.9, 1$ los valores obtenidos para el NC son inferiores a 30 pero superiores a 20, por lo que la multicolinealidad presente en el análisis sería moderada.
- Alzando la segunda y tercera variable, se tiene para todas las transformaciones que todos los valores obtenidos para el NC son superiores a 30, por lo que la RA no habría mitigado la multicolinealidad grave del modelo (1).

Si se considera que el parámetro de alzado varia entre 0 y 100 tomando valores de 0.1 en 0.1 se tienen los valores del NC mostrados en las representaciones gráficas de la Figura 2 alzando la primera (azul), la segunda (verde) y tercera (rojo) variable considerando datos no centrados con varianza igual a 1 entre el número de observaciones (arriba), estandarizados (centro) y de longitud unidad (abajo). Se puede observar que:

λ	S.1	S.2	S.3	S.4	S.5	S.6	S.7
0	12592.757	113.286	113.286	12947.962	51.728	51.728	105.019
0.1	11465.895	103.802	103.802	11776.511	47.042	47.0422	95.777
0.2	10528.444	95.943	95.943	10800.8908	43.139	43.139	88.102
0.3	9736.703	89.332	89.332	9975.935	39.841	39.841	81.634
0.4	9059.457	83.702	83.702	9269.387	37.017	37.017	76.114
0.5	8473.815	78.856	78.856	8657.594	34.574	34.574	71.353
0.6	7962.609	74.646	74.646	8122.822	32.442	32.442	67.209
0.7	7512.711	70.959	70.959	7651.5103	30.569	30.569	63.573
0.8	7113.911	67.707	67.707	7233.117	28.915	28.915	60.3608
0.9	6758.147	64.821	64.821	6859.323	27.453	27.453	57.505
1	6438.972	62.245	62.245	6523.478	26.164	26.164	54.952

Tabla 3: Número de Condición en los siete escenarios considerados alzando la primera variable

λ	S.1	S.2	S.3	S.4	S.5	S.6	S.7
0	12592.7578	113.286	113.286	12947.962	51.728	51.728	105.019
0.1	11453.503	103.057	103.057	12065.517	48.315	48.315	95.477
0.2	10504.627	94.537	94.537	11351.738	45.629	45.629	87.525
0.3	9702.199	87.332	87.332	10766.5309	43.497	43.497	80.796
0.4	9014.842	81.1602	81.1602	10281.206	41.794	41.794	75.0302
0.5	8419.547	75.814	75.814	9874.706	40.423	40.423	70.032
0.6	7899.056	71.1401	71.1401	9531.244	39.312	39.312	65.6604
0.7	7440.174	67.018	67.018	9238.773	38.404	38.404	61.802
0.8	7032.637	63.358	63.358	8987.957	37.657	37.657	58.373
0.9	6668.343	60.085	60.085	8771.474	37.037	37.037	55.306
1	6340.81	57.143	57.143	8583.511	36.517	36.517	52.545

Tabla 4: Número de Condición en los siete escenarios considerados alzando la segunda variable

λ	S.1	S.2	S.3	S.4	S.5	S.6	S.7
0	12592.757	113.286	113.286	12947.962	51.728	51.728	105.019
0.1	11587.236	103.611	103.611	12397.191	49.447	49.447	96.29804
0.2	10760.2652	95.602	95.602	11966.181	47.7202	47.7202	89.099
0.3	10070.357	88.872	88.872	11623.796	46.392	46.392	83.072
0.4	9487.8701	83.149	83.149	11348.284	45.354	45.354	77.965
0.5	8991.051	78.229	78.229	11124.107	44.5308	44.5308	73.592
0.6	8563.584	73.962	73.962	10939.952	43.867	43.867	69.815
0.7	8192.982	70.233	70.233	10787.435	43.326	43.326	66.528
0.8	7869.536	66.951	66.951	10660.255	42.879	42.879	63.648
0.9	7585.579	64.045	64.045	10553.601	42.505	42.505	61.109
1	7334.976	61.458	61.458	10463.761	42.188	42.188	58.861

Tabla 5: Número de Condición en los siete escenarios considerados alzando la tercera variable

- En todos los casos los valores obtenidos para el NC se estabilizan para valores altos de λ.

- Si se alza la primera variable, los valores de obtenidos se estabilizan entorno a 19 para S.3, 18 para S.6 y 18 para S.7. Es decir, en todos los casos se obtienen valores inferiores a 20 por lo que la multicolinealidad presente en el modelo no sería preocupante.

- Si se alza la segunda variable, los valores de obtenidos se estabilizan entorno a 14 para S.3, 26 para S.6 y 11 para S.7. Es decir, sólo en el segundo caso se obtienen valores superiores a 20, por lo que la multicolinealidad presente en el modelo no sería preocupante en los escenarios S.3 y S.7.

- Si se alza la tercera variable, los valores de obtenidos se estabilizan entorno a 28 para S.3, 32 para S.6 y 26 para S.7. Es decir, en todos los casos se obtienen valores no inferiores a 20 por lo que la multicolinealidad presente en el modelo sería preocupante.

Adviértase que dependiendo de la transformación realizada, se llegan a conclusiones distintas. Por tanto, es evidente que es necesario determinar qué transformación realizada sobre los datos es la más adecuada para calcular el NC en la RA.

Para finalizar, en las representaciones gráficas de la Figura 3 se muestran los autovalores mínimo (izquierda) y máximo (derecha) alzando la primera (azul), la segunda (verde) y tercera (rojo) variable considerando datos no centrados con varianza igual a 1 entre el número de observaciones (arriba), estandarizados (centro) y de longitud unidad (abajo). Se observa que el comportamiento del autovalor mínimo es crecer conforme aumenta el valor del parámetro de alzado y que el del autovalor máximo es decrecer. Por tal motivo el número de condición decrece conforme aumenta dicho parámetro.

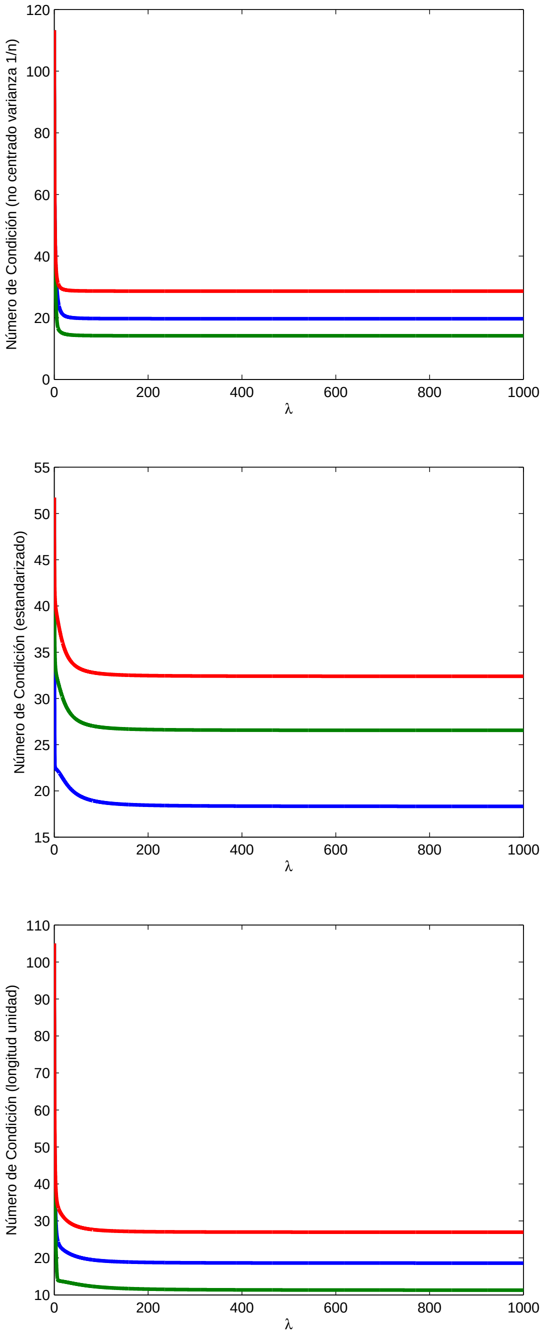

Figura 2: Valores del Número Condición alzando la primera (azul), la segunda (verde) y tercera (rojo) variable considerando datos en el escenario S.3 (arriba), S.6 (centro) y S.7 (abajo)

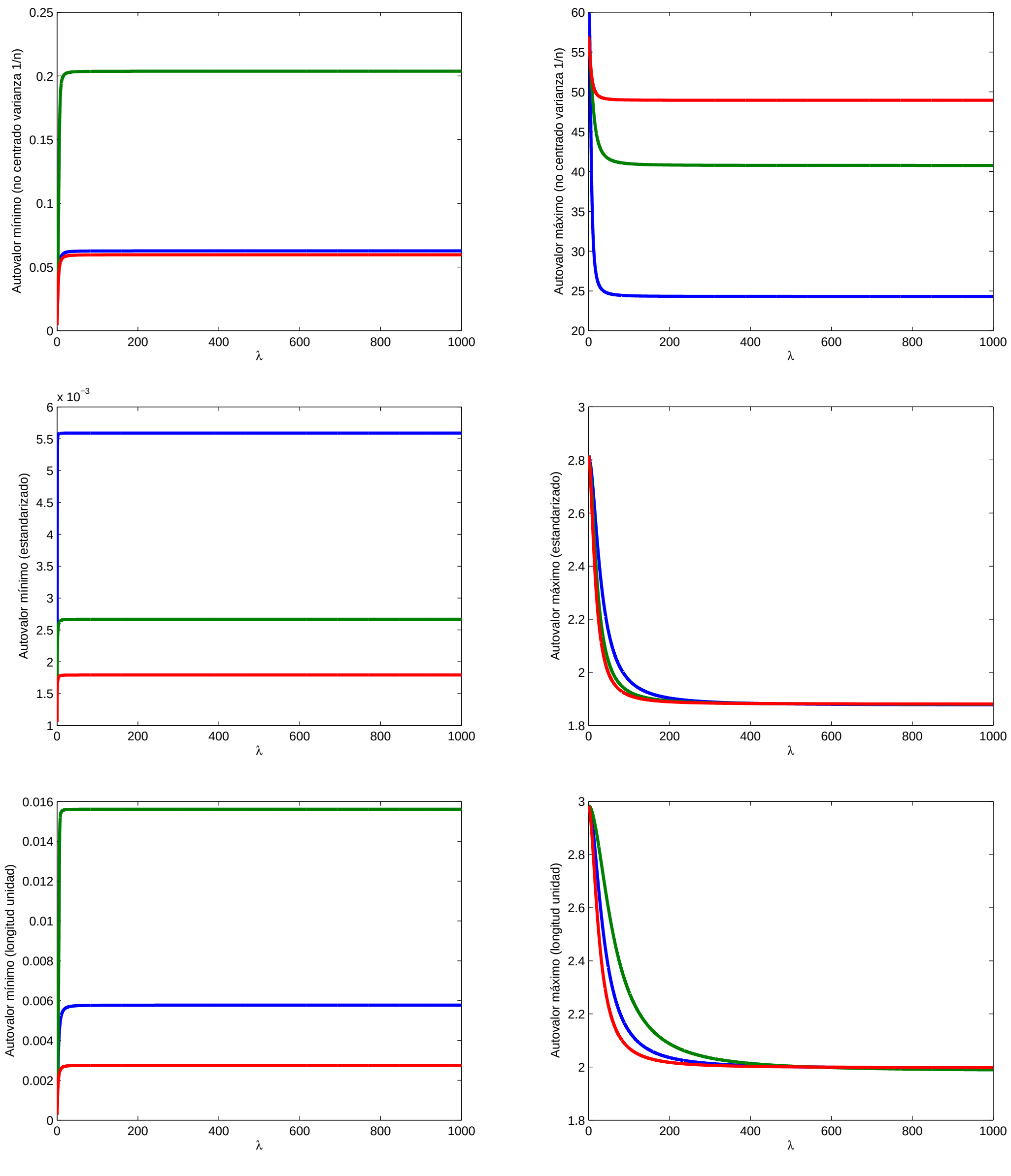

Figura 3: Autovalores mínimo (izquierda) y máximo (derecha) alzando la primera (azul), la segunda (verde) y tercera (rojo) variable considerando datos en el escenario S.3 (arriba), S.6 (centro) y S.7 (abajo)

5 CONCLUSIONES

Cuando en un modelo de regresión lineal general se detecta un grado alto de multicolinealidad, una de las posibilidades es aplicar la Regresión Alzada. Esta técnica depende de un parámetro de alzado, de forma que conforme aumenta el valor de éste parámetro el grado de multicolinealidad en el modelo disminuye. Una vez seleccionado el valor idóneo del parámetro de alzado (ver García y otros 2015b para más detalles) es conveniente comprobar si la multicolinealidad ha sido realmente mitigada. Por tanto, se hace necesaria la extensión de algunas de las medidas de detección de multicolinealidad a la Regresión Alzada.

En el presente trabajo se plantea la extensión del Número de Condición, encontrando dificultad en la obtención de una expresión algebraica cerrada lo cual imposibilita comprobar si la extensión planteada verifica una serie de propiedades consideradas como deseables. Por otro lado, también se aborda la sensibilidad que presenta el Número de Condición a las transformaciones en los datos. En este sentido, los resultados más razonables se obtienen cuando se trabaja con variables con varianza igual a 1 (que coinciden con los obtenidos cuando la varianza es igual a 1 dividido entre el número de observaciones) ya sea con datos centrados o no y con datos con longitud unidad. Para determinar qué transformación es la más adecuada en la Regresión Alzada se hace necesario un análisis más profundo. Quizás sería conveniente comparar los resultados obtenidos a partir del Número de Condición con los obtenidos con el Factor de Agrandamiento de la Varianza y ver en qué casos se obtienen conclusiones congruentes.

6 REFERENCIAS BIBLIOGRÁFICAS

- MARQUARDT, D.W. (1970). "Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation". Technometrics, 12, 591–612.

- THEIL, H. (1971). “Principles of econometrics”. New York: Wiley.
- FOX, J. y MONETTE, G. (1992). “Generalized collinearity diagnostics”. Journal of the American Statistical Association, 87, 178–183.
- BELSLEY, D.A. (1982). “Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise”. Journal of Econometrics, 20, 211–253.
- BELSLEY, D.A., KUH, E. y WELSCH, R.E. (1980). “Regression diagnostics”. New York, NY: John Wiley.
- BELSLEY, D.A. (1991). “Conditioning diagnostics: Collinearity and weak data in regression”. New York: John Wiley.
- HOERL, A.E. y KENNARD, R.W. (1970a). “Ridge regression: Applications to nonorthogonal problems”. Technometrics, 12, 69–82.
- HOERL, A.E. y KENNARD, R.W. (1970b). “Ridge regression: Biased estimation for nonorthogonal problems”. Technometrics, 12, 55–67.
- GARCÍA, C. B., GARCÍA, J. y SOTO, J. (2010). “The raise method: An alternative procedure to estimate the parameters in presence of collinearity”. Quality and Quantity, 45 (2), 403–423.
- GARCÍA, C. B., GARCÍA, J., LÓPEZ MARTÍN, M. M. y SALMERÓN, R. (2015a). “Collinearity: Revisiting the variance inflation factor in ridge regression”. Journal of Applied Statistics, 42 (3), 648–661.
- GARCÍA, C. B., GARCÍA, J., SALMERÓN, R. y LÓPEZ MARTÍN, M. M. (2015b). “Raise regression: selection of the raise parameter”. International Conference on Data Mining, 185–190.

- GARCÍA, J., SALMERÓN, R., GARCÍA, C.B. y LÓPEZ MARTÍN, M.M. (2015c). "Standardization of Variables and Collinearity Diagnostic in Ridge Regression". International Statistical Review. DOI: 10.1111/insr.12099.

- GARCÍA, C., GARCÍA, J., LÓPEZ MARTÍN, M. M. y SALMERÓN, R. (2016). "Raise estimator: inference and properties". Communications in Statistics-Theory and Methods. Pendiente de publicación.

- SALMERÓN, R., GARCÍA, J., LÓPEZ MARTÍN, M.M. y GARCÍA, C.B. (2016a). "Collinearity diagnostic applied in ridge estimation through the variance inflation factor". Journal of Applied Statistics. DOI: 10.1080/02664763. 2015.1120712.

- SALMERÓN, R., GARCÍA, J., GARCÍA, C.B. y LÓPEZ MARTÍN, M.M. (2016b). "Transformation of variables and the condition number in ridge estimation". Computational Statistics and Data Analysis. En revisión.

- WISSEL, J. (2009). "A new biased estimator for multivariate regression models with highly collinear variables". Ph.D. thesis, Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-UniversitätWürzburg.