

# Micronumerosidad aproximada y regresión lineal múltiple

Salmerón Gómez, Román (romansg@ugr.es)

*Departamento de Métodos Cuantitativos para la Economía y la Empresa  
Universidad de Granada*

Blanco Izquierdo, Víctor (vblanco@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*

## RESUMEN

Una de las suposiciones realizadas en la regresión lineal múltiple es que el número de observaciones ha de ser mayor que el número de regresores considerados. Cuando el número de observaciones apenas supera el número de parámetros a estimar pueden surgir problemas de multicolinealidad como la estimación de signos inesperados o sensibilidad a cambios en la muestra. Para resaltar el problema del tamaño de la muestra, Goldberger definió el término de micronumerosidad aproximada de manera homóloga a la multicolinealidad aproximada. En este ambiente, en el que la relación lineal de las variables no se debe tanto a su naturaleza si no al bajo tamaño de la muestra, proponemos identificar aquellas observaciones que puedan estar provocando el problema de multicolinealidad y valorar su eliminación del análisis. El presente trabajo supone una primera aproximación al problema, por lo que se propone una primera solución y se establecen posibles pautas a seguir para obtener una solución óptima del mismo.

***Palabras clave:*** micronumerosidad; multicolinealidad; regresión lineal.

***Área temática:*** Aspectos Cuantitativos de Problemas Económicos y Empresariales con certeza.

## ABSTRACT

One of the main assumptions when analyzing multiple linear regression models is the one that asserts that the number of observations must be greater than the number of regressors. When the number of observations is slightly greater than the number of parameters to be estimated, a multicollinearity problem may appear (unexpected signs in the parameters, high sensibility to sample changes, etc). Goldberger introduced the notion of approximated micronumerosity analogously to the concept of approximated multicollinearity. In this framework, where the linear relationship between variables is mainly due to the size of the sample, we propose to identify those observations that induce the multicollinearity and evaluate whether those should be removed from the sample. We present a first approach to this problem and future steps to be taken in this line.

## 1 INTRODUCCIÓN

Se entiende que en el modelo de regresión lineal general hay multicolinealidad aproximada cuando existen dos o más variables exógenas relacionadas linealmente de forma aproximada. La existencia de un alto grado de multicolinealidad no es una cuestión baladí, ya que puede poner en entredicho las conclusiones obtenidas en el análisis realizado al ser erróneas. De ahí la importancia de detectarla y tratarla de forma adecuada.

Las causas que producen multicolinealidad en un modelo son diversas, sin embargo, podrían dividirse en los dos siguientes bloques (ver Spanos y McGuirk 2002):

**Multicolinealidad sistemática:** debida a un problema estructural, es decir, a la alta correlación lineal de las variables exógenas consideradas.

**Multicolinelidad errática:** debida a un problema puramente numérico, es decir,

a un mal condicionamiento de los datos considerados.

Dentro del segundo bloque quedaría encuadrado el término de micronumerosidad acuñado por Goldberger 1991. Este término hace referencia a la situación en la que se presenta multicolinealidad aproximada debido al bajo número de observaciones disponibles en el estudio. Es decir, los signos de la multicolinealidad aproximada no se deben a que las variables exógenas estén realmente relacionadas linealmente sino a la poca variabilidad existente en las observaciones debido al reducido tamaño de la muestra. En esta situación, la eliminación de algunas observaciones podría mitigar el problema de multicolinealidad aproximada.

En el presente trabajo se realiza una primera aproximación a esta posible solución: tratar de identificar aquellas observaciones que puedan estar provocando el problema de multicolinealidad y valorar su eliminación del análisis.

En cuanto al segundo aspecto, hay que tener en cuenta que los datos candidatos a ser eliminados podrían:

- tener tal importancia dentro de la muestra (correspondan a los países, comunidades, empresas, etc más relevantes dentro de la muestra o sean las observaciones más recientes en una serie temporal) que se desaconsejase su eliminación, o
- que el número de observaciones sea elevado en relación al tamaño de la muestra por lo que no sería adecuado eliminarlas, o
- simplemente que no exista un conjunto de observaciones red de un tamaño razonable a eliminar que mitigue la multicolinealidad aproximada.

El trabajo se estructura como sigue: en la sección 2 se realiza una primera aproximación al problema mostrando en un modelo con dos variables independientes el efecto que tiene eliminar de forma reiterativa una de las observaciones y estimar

posteriormente el modelo. Puesto que eliminar observaciones y estimar el modelo para comprobar los efectos de la eliminación no supone una estrategia eficiente, en la sección 3 se plantea una metodología para detectar aquellas posibles observaciones que tras ser eliminadas hagan disminuir el grado de multicolinealidad en el modelo. Dicha metodología será aplicada en un modelo con tres variables independientes en la sección 4. Finalmente, en la sección 5 se destacan las principales conclusiones del trabajo.

## 2 PRIMERA APROXIMACIÓN

Supongamos que en el modelo lineal general con  $n$  observaciones y  $p$  variables independientes dado por:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

donde la perturbación aleatoria,  $\mathbf{u}$ , es esférica, se verifica que el grado de multicolinealidad presente en el modelo es grave. Para detectar este último aspecto usaremos el Factor de Inflación de la Varianza (FIV) dado por:

$$FIV(i) = \frac{1}{1 - R_i^2}, \quad i = 2, \dots, p, \quad (2)$$

donde  $R_i^2$  es el coeficiente de determinación de la regresión de  $\mathbf{X}_i$  sobre el resto de variables independientes,  $\mathbf{X}_{-i}$ . Es comunmente adoptado el criterio de que si alguno de estos  $p - 1$  índices es superior a 10 se considera que el grado de multicolinealidad presente en el modelo es preocupante.

Más concretamente, consideremos el modelo presentado en Pasha y Ali Shah 2004 sobre el número de personas empleadas en Pakistán:

$$PE_t = \beta_0 + \beta_1 P_t + \beta_2 A_t + u_t, \quad t = 1974, \dots, 2002, \quad (3)$$

donde:

**PE:** número de personas empleadas (en millones).

**P:** población de Pakistán (en millones).

**A:** índice de alfabetismo (en porcentaje).

Al estimarlo por Mínimos Cuadrados Ordinarios (MCO) se obtienen los siguientes resultados:

$$\widehat{PE}_t = 2.5988 + 0.0949 \cdot P_t + 0.2335 \cdot A_t \quad R^2 = 0.9878 \quad F_{2,25} = 1015$$

$$(0.7123) \quad (0.0487) \quad (0.01925)$$

Calculado el FIV se obtiene que es igual a 10.368, de forma que al ser superior a 10 se puede pensar en que en el modelo hay multicolinealidad grave.

Considerando que la relación lineal<sup>1</sup> existente entre las variables se debe a la muestra concreta con la que se está trabajando y no a la relación estructural de las variables (mayor población no tiene por qué implicar mayor índice de alfabetismo) se puede pensar que la eliminación de algunas observaciones puede mitigar el problema de la multicolinealidad.

| Observación eliminada | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $R^2$   | FIV     |
|-----------------------|-----------------|-----------------|-----------------|---------|---------|
| 1                     | 2.8524          | 0.104671        | 0.22824         | 0.98689 | 10.4718 |
| 2                     | 3.1597          | 0.114735        | 0.22244         | 0.98858 | 10.3161 |
| 3                     | 2.5856          | 0.094489        | 0.23375         | 0.98697 | 10.1922 |
| 4                     | 2.4882          | 0.091862        | 0.23540         | 0.98727 | 10.1289 |
| 5                     | 2.3840          | 0.090496        | 0.23668         | 0.98791 | 10.0530 |
| 6                     | 2.4489          | 0.093475        | 0.23518         | 0.98782 | 10.0200 |

Sigue en la página siguiente...

---

<sup>1</sup>Puesto que la regresión auxiliar es una regresión simple, su coeficiente de determinación coincide con el coeficiente de correlación al cuadrado. Por tanto, si el FIV es igual a 10.368 se tiene que el coeficiente de correlación entre las dos variables independientes es 0.95055. Por tanto, ambas variables están altamente relacionadas linealmente.

| Observación eliminada | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $R^2$   | FIV            |
|-----------------------|-----------------|-----------------|-----------------|---------|----------------|
| 7                     | 2.4869          | 0.095448        | 0.23422         | 0.98783 | 10.0370        |
| 8                     | 2.4697          | 0.090992        | 0.23578         | 0.98787 | 10.3029        |
| 9                     | 2.5546          | 0.094412        | 0.23400         | 0.98770 | 10.2199        |
| 10                    | 2.5827          | 0.095065        | 0.23357         | 0.98770 | 10.2378        |
| 11                    | 2.5991          | 0.094885        | 0.23351         | 0.98773 | 10.3052        |
| 12                    | 2.5907          | 0.097445        | 0.23270         | 0.98787 | 10.4422        |
| 13                    | 2.6051          | 0.099009        | 0.23206         | 0.98792 | 10.6376        |
| 14                    | 2.5865          | 0.092111        | 0.23454         | 0.98785 | 11.0275        |
| 15                    | 2.6174          | 0.097978        | 0.23233         | 0.98785 | 11.6259        |
| 16                    | 2.6396          | 0.100659        | 0.23126         | 0.98785 | 12.4304        |
| <b>17</b>             | <b>2.7165</b>   | <b>0.109977</b> | <b>0.22755</b>  | 0.98793 | <b>13.4359</b> |
| 18                    | 2.5864          | 0.094469        | 0.23386         | 0.98786 | 10.3264        |
| 19                    | 2.5729          | 0.095456        | 0.23380         | 0.98817 | 10.2818        |
| 20                    | 2.5442          | 0.096306        | 0.23392         | 0.98859 | 10.2256        |
| 21                    | 2.5293          | 0.096715        | 0.23392         | 0.98844 | 10.1547        |
| 22                    | 2.4574          | 0.098442        | 0.23427         | 0.98965 | 10.0688        |
| <b>23</b>             | <b>2.4282</b>   | <b>0.099727</b> | <b>0.23417</b>  | 0.98962 | <b>9.9642</b>  |
| <b>24</b>             | <b>2.6360</b>   | <b>0.093773</b> | <b>0.23340</b>  | 0.98720 | <b>9.8444</b>  |
| <b>25</b>             | <b>2.6471</b>   | <b>0.093198</b> | <b>0.23346</b>  | 0.98703 | <b>9.7060</b>  |
| <b>26</b>             | <b>2.6577</b>   | <b>0.092509</b> | <b>0.23355</b>  | 0.98683 | <b>9.5580</b>  |
| <b>27</b>             | <b>2.8245</b>   | <b>0.077942</b> | <b>0.23607</b>  | 0.98822 | <b>9.4388</b>  |
| 28                    | 2.6610          | 0.054594        | 0.24492         | 0.98717 | 10.5674        |

Tabla 1: Estimación de los coeficientes del modelo 3, su coeficiente de determinación y el FIV

Con tal objetivo se elimina cada una de las observaciones y se estima el modelo

y calcula el FIV. Los resultados se muestran en la Tabla 1, se puede observar que eliminando las observaciones 23, 24, 25, 26 y 27 se obtienen FIVs menores que 10, obteniéndose el mínimo valor cuando se elimina la observación 27. En tal caso se obtienen los siguientes resultados:

$$\widehat{PE}_t = 2.8245 + 0.07794 \cdot P_t + 0.236 \cdot A_t \quad R^2 = 0.9882 \quad F_{2,24} = 1007$$

$$(0.6784) \quad (0.0465) \quad (0.01814)$$

### 3 SELECCIÓN DE OBSERVACIONES CANDIDATAS A SER ELIMINADAS

Parece evidente que eliminar observaciones de 1 en 1 (ó de 2 en 2 ó ...) y estimar posteriormente el modelo y comprobar sí el problema de multicolinealidad se ha mitigado puede ser una buena primera aproximación para dar solución al problema propuesto pero sin duda no es la metodología de actuación más eficiente.

Con el objetivo de mejorar lo expuesto en la sección anterior sería deseable saber a priori las observaciones *problemáticas*, es decir, aquellas que puedan ser causantes del grado de multicolinealidad existente en el modelo. ¿Es posible limitar la eliminación de observaciones a un subconjunto (reducido) de toda la muestra?

Para responder a dicha pregunta se considera la regresión auxiliar usada para el cálculo del FIV,  $\mathbf{X}_i = \mathbf{X}_{-i}\boldsymbol{\alpha} + \mathbf{v}$ . Los valores responsables de un buen ajuste en dicha regresión y, por tanto, de que  $R_i^2$  sea alto (suponiendo por tanto un *FIV* alto) serán aquellos puntos que estén más cerca de la recta ajustada. En definitiva, las observaciones *responsables* del grado de multicolinealidad del modelo original serán aquellas con menor residuo en la regresión auxiliar. De esta forma se habrían identificado las observaciones candidatas a ser eliminadas.

En las gráficas de la Figura 1 se tiene para el modelo (3) la representación de las observaciones de las regresiones auxiliares y de la recta ajustada. Puesto que no



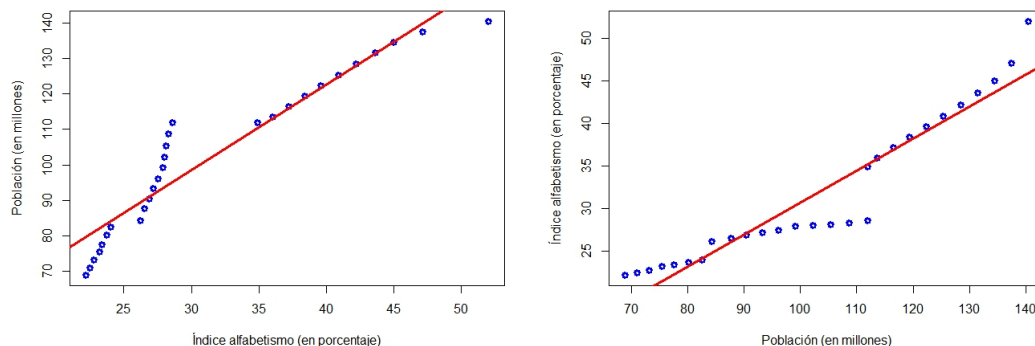


Figura 1: Representación de las observaciones de las regresiones auxiliares y de la recta ajustada

es fácil identificar las observaciones con menor residuo en dichas representaciones se hace conveniente mostrarlos en la Tabla 2. Puesto que el signo es irrelevante, ha sido ordenado su valor absoluto de menor a mayor. Así por ejemplo, a partir de la regresión de  $P$  sobre  $A$  se tiene que:

- Las observaciones 26, 25, 24 y 23 (que conducen tras su eliminación a un modelo con FIV menor que 10) corresponden a las observaciones 1, 2, 3 y 5 con menor residuo.
- La observación 27 que corresponde a un modelo con menor FIV tras su eliminación corresponde a la observación 13 con menor residuo.
- La observación con mayor residuo, la 17, conduce a un modelo donde se obtiene el mayor FIV tras eliminar la observación (ver Tabla 1).

Por tanto, eliminar las observaciones con menor residuo no garantiza obtener el menor FIV tras su eliminación pero si se observa que elegir las candidatas a ser eliminadas entre las observaciones con menor residuo en la regresión auxiliar puede conducir a buenos resultados.

| Observación | Residuo<br>regresión<br>$P$ sobre $A$ | Residuo<br>regresión<br>$A$ sobre $P$ | Orden | Res. ordenado<br>regresión<br>$P$ sobre $A$ | Res. ordenado<br>regresión<br>$A$ sobre $P$ |
|-------------|---------------------------------------|---------------------------------------|-------|---------------------------------------------|---------------------------------------------|
| 1           | -10.8704754                           | 3.1419341                             | 1     | <b>0.1730050</b>                            | 0.1704902                                   |
| 2           | -9.4837061                            | 2.6503892                             | 2     | <b>0.1964047</b>                            | 0.1834035                                   |
| 3           | -8.0339367                            | 2.1351993                             | 3     | <b>0.4778144</b>                            | 0.2484129                                   |
| 4           | -6.7579109                            | 1.6948631                             | 4     | 0.5639920                                   | 0.2970818                                   |
| 5           | -5.1673980                            | 1.1172050                             | 5     | <b>0.5734806</b>                            | 0.3019787                                   |
| 6           | -3.2756287                            | 0.4361247                             | 6     | 0.5929146                                   | 0.3732017                                   |
| 7           | -1.5478593                            | -0.1834035                            | 7     | 0.5950694                                   | 0.4011543                                   |
| 8           | -5.1702174                            | 1.3883147                             | 8     | 0.6294223                                   | 0.4361247                                   |
| 9           | -2.3884481                            | 0.3732017                             | 9     | 0.6771468                                   | 0.4860856                                   |
| 10          | -0.6294223                            | -0.2484129                            | 10    | 1.4543471                                   | 0.6503744                                   |
| 11          | 1.4543471                             | -1.0015541                            | 11    | 1.5478593                                   | 0.8116607                                   |
| 12          | 3.6261164                             | -1.7877234                            | 12    | 1.5690936                                   | 1.0015541                                   |
| 13          | 5.6451422                             | -2.5069206                            | 13    | <b>2.2276195</b>                            | 1.0523045                                   |
| 14          | 8.4723987                             | -3.5583951                            | 14    | 2.3884481                                   | 1.1172050                                   |
| 15          | 11.4106551                            | -4.6515299                            | 15    | 3.2756287                                   | 1.3259763                                   |
| 16          | 14.1981680                            | -5.6784432                            | 16    | 3.6261164                                   | 1.3883147                                   |
| <b>17</b>   | <b>16.7359373</b>                     | -6.6019787                            | 17    | 5.1673980                                   | 1.6948631                                   |
| 18          | 1.5690936                             | -0.3019787                            | 18    | 5.1702174                                   | 1.7877234                                   |
| 19          | 0.5929146                             | 0.1704902                             | 19    | 5.6451422                                   | 2.1351993                                   |
| 20          | 0.5639920                             | 0.2970818                             | 20    | 6.7579109                                   | 2.2996480                                   |
| 21          | 0.5950694                             | 0.4011543                             | 21    | 8.0339367                                   | 2.5069206                                   |
| 22          | 0.6771468                             | 0.4860856                             | 22    | 8.4723987                                   | 2.6503892                                   |
| <b>23</b>   | <b>0.5734806</b>                      | 0.6503744                             | 23    | 9.4837061                                   | 3.1419341                                   |

Sigue en la página siguiente...

| Observación | Residuo<br>regresión | Residuo<br>regresión | Orden | Res. ordenado<br>regresión | Res. ordenado<br>regresión |
|-------------|----------------------|----------------------|-------|----------------------------|----------------------------|
|             | $P$ sobre $A$        | $A$ sobre $P$        |       | $P$ sobre $A$              | $A$ sobre $P$              |
| <b>24</b>   | <b>0.4778144</b>     | 0.8116607            | 24    | 10.8704754                 | 3.5583951                  |
| <b>25</b>   | <b>0.1964047</b>     | 1.0523045            | 25    | 11.0630535                 | 4.6515299                  |
| <b>26</b>   | <b>-0.1730050</b>    | 1.3259763            | 26    | 11.4106551                 | 5.6784432                  |
| <b>27</b>   | <b>-2.2276195</b>    | 2.2996480            | 27    | 14.1981680                 | 6.0883325                  |
| 28          | -11.0630535          | 6.0883325            | 28    | <b>16.7359373</b>          | 6.6019787                  |

Tabla 2: Residuos de las regresiones auxiliares y ordenación de menor a mayor de su valor absoluto

## 4 EJEMPLO PRÁCTICO

En la presente sección se pretende ilustrar globalmente la metodología presentada en el trabajo mediante un ejemplo práctico. Por tanto, se considerará un modelo de regresión lineal múltiple en el que se seguirán los siguientes pasos:

- Establecer que el grado de multicolinealidad es preocupante mediante el cálculo de los FIVs.
- Seleccionar una de las regresiones auxiliares y usar sus residuos para detectar las posibles observaciones *responsables* del grado de multicolinealidad.
- Eliminar la observación u observaciones seleccionadas en el punto anterior y volver a estimar el modelo, comprobando si se ha mitigado el problema de multicolinealidad y discutiendo los nuevos resultados.

Más concretamente, se considerará el siguiente modelo en el que se persigue el objetivo de predecir el número de horas mensuales por hombre con fines de dotación

en 17 hospitales navales de Estados Unidos:

$$H_i = \beta_0 + \beta_1 L_i + \beta_2 X_i + \beta_3 S_i + u_i, \quad i = 1, \dots, 17, \quad (4)$$

donde:

**H:** horas mensuales por hombre.

**L:** carga media de pacientes diaria.

**X:** exposición mensual a rayos X.

**S:** duración media (en días) de permanencia de los pacientes.

Este modelo es una versión reducida del usado por Myers 1990, el cual consideraba 5 variables independientes.

La estimación del modelo (4) por MCO conduce a los siguientes resultados:

$$\begin{aligned} \hat{H}_i &= 1475.024 + 29.731 \cdot L_i + 0.0534 \cdot X_i - 318.14 \cdot S_i & R^2 &= 0.9894 \\ &(811.002) & (3.3169) & (0.0208) & (158.4065) & F_{3,13} &= 404.6 \end{aligned}$$

Además, los FIVs asociados a cada una de las variables son:

$$FIV(1) = 11.32138, \quad FIV(2) = 7.771393, \quad FIV(3) = 2.498503.$$

Puesto que el mayor de los FIVs es superior a 10 se puede considerar que el grado de multicolinealidad es grave.

Puesto que hay tres regresiones auxiliares posibles, se trabajará con aquella que ha conducido a un mayor FIV. En este caso se trata de la regresión de  $L$  sobre  $X$  y  $S$ . Los residuos de esta regresión se muestran en la Tabla 3. Se observa que el menor residuo corresponde a la observación 15. Tras eliminar esta observación<sup>2</sup> y volver a estimar el modelo por MCO se obtiene que:

---

<sup>2</sup>Esta observación corresponde al hospital número 15. El investigador debe valorar si dicha observación puede ser eliminada del análisis. Es posible que dicho hospital tenga una especial importancia dentro de la muestra y no se deba prescindir de la información que aporta al estudio.

$$\begin{array}{ccccccc} \hat{H}_i = & 751.1309 & + & 29.757 \cdot L_i & + & 0.05095 \cdot X_i & - & 177.97 \cdot S_i & R^2 = 0.989 \\ & (1131.33) & & (3.3359) & & (0.02108) & & (220.058) & F_{3,12} = 360.7 \end{array}$$

Además, los FIVs son ahora iguales a:

$$FIV(1) = 9.348342, \quad FIV(2) = 7.579579, \quad FIV(3) = 1.754767.$$

Se tiene que los tres FIVs han disminuido y que el mayor de ellos es menor que 10. Por tanto, el problema de la multicolinealidad ha sido mitigado.

Adviértase que una vez eliminada la observación 15 los coeficientes estimados para el término independiente y la duración media de permanencia de los pacientes han variado sustancialmente. Así, por ejemplo, suponiendo una carga diaria de 20 pacientes, una exposición mensual a rayos X de 4000 horas y una duración media de permanencia de los pacientes de 4 días se tiene que:

- el modelo original predice 1010.694 horas mensuales por hombre,
- mientras que el modelo resultante tras eliminar la observación 15 predice 838.1849 horas mensuales.

Como se puede observar las predicciones son sensiblemente distintas.

Para finalizar, en la Tabla 4 se tiene la estimación de los coeficientes del modelo (4) tras eliminar cada una de las observaciones y el FIV asociado a cada variable. Se puede observar que la eliminación de la observación 15 es la única que conduce a que todos los FIVs sean menores que 10. Además, la eliminación de la observación con mayor residuo, la 16, conduce a los mayores valores para cada uno de los FIVs. Curiosamente, la eliminación de estas dos observaciones hace que los coeficientes estimados para el término independiente y la duración media de permanencia de los pacientes varíen sustancialmente.

| Observación | Residuos    |
|-------------|-------------|
| 1           | 6.2309455   |
| 2           | -46.3481295 |
| 3           | 8.3374639   |
| 4           | 4.7558842   |
| 5           | -14.3247158 |
| 6           | -21.5115180 |
| 7           | -12.4188545 |
| 8           | 6.1618206   |
| 9           | -7.8267440  |
| 10          | -40.0769967 |
| 11          | -23.9573681 |
| 12          | 59.8353024  |
| 13          | 7.2754212   |
| 14          | -36.3175431 |
| 15          | 0.8950598   |
| 16          | 155.5033352 |
| 17          | -46.2133631 |

Tabla 3: Residuos de la regresión auxiliar de  $L$  sobre  $X$  y  $S$

| Obs. eliminada | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | FIV(1)    | FIV(2)    | FIV(3)   |
|----------------|-----------------|-----------------|-----------------|-----------------|-----------|-----------|----------|
| 1              | 1502.4141       | 29.74841        | 0.05318471      | -321.5524       | 10.823861 | 7.528513  | 2.411791 |
| 2              | 1477.4009       | 29.74004        | 0.05337894      | -318.7531       | 11.843232 | 7.581326  | 2.851086 |
| 3              | 1437.7499       | 29.70484        | 0.05364731      | -313.0715       | 10.871102 | 7.576064  | 2.406361 |
| 4              | 1269.3025       | 29.65988        | 0.05378483      | -288.1232       | 10.842887 | 7.620528  | 2.361825 |
| 5              | 1471.6652       | 29.78187        | 0.05333664      | -319.9200       | 11.104406 | 7.596984  | 2.496623 |
| 6              | 1538.3997       | 29.50050        | 0.05480504      | -323.4755       | 11.168502 | 7.824212  | 2.395818 |
| 7              | 1471.6975       | 29.80892        | 0.05338049      | -321.7352       | 11.122653 | 7.596645  | 2.504410 |
| 8              | 1429.6891       | 29.68701        | 0.05401112      | -313.8489       | 11.104316 | 7.642797  | 2.473277 |
| 9              | 1449.6946       | 29.62383        | 0.05302725      | -304.9426       | 11.257730 | 7.668404  | 2.523048 |
| 10             | 1328.0112       | 28.63955        | 0.05974437      | -275.3008       | 11.862832 | 8.093143  | 2.576496 |
| 11             | 1498.9094       | 30.01694        | 0.05212160      | -329.7848       | 11.433192 | 7.811943  | 2.529482 |
| 12             | 1845.8749       | 31.05665        | 0.04560651      | -382.2775       | 12.664822 | 8.519852  | 2.728872 |
| 13             | 1523.7299       | 29.79750        | 0.05300825      | -323.5343       | 11.325967 | 7.774426  | 2.499527 |
| 14             | 1906.0711       | 31.64269        | 0.03863464      | -413.1005       | 11.468678 | 7.875344  | 2.528047 |
| 15             | 751.1310        | 29.75708        | 0.05095229      | -177.9719       | 9.348342  | 7.579579  | 1.754767 |
| 16             | 641.1783        | 23.72388        | 0.08446886      | -135.3852       | 34.474956 | 22.519449 | 5.427786 |
| 17             | 1448.5947       | 26.07694        | 0.10719843      | -358.5753       | 10.773012 | 8.258941  | 2.542324 |

Tabla 4: Estimación de los coeficientes del modelo 4 tras eliminar una observación y el FIV asociado a cada variable

## 5 CONCLUSIONES

En el presente trabajo se presenta una metodología para detectar las posibles observaciones que puedan ser causantes de la presencia de multicolinealidad grave en un modelo de regresión lineal múltiple de forma que al eliminarlas del estudio se mitigue dicho problema.

Se trata de una primera aproximación por lo que presenta algunas limitaciones:

1. Cuando la relación lineal existente entre variables es estructural, es decir, se debe a la propia naturaleza de las mismas, parece lógico pensar que la eliminación de observaciones no suponga que el problema de multicolinealidad se mitigue. Por tanto, el ambiente donde se debe usar dicha metodología, llamado por Goldberger 1991 como micronumerosidad, es aquel en el que el reducido tamaño de la muestra pueda ser el causante del problema.
2. Una vez determinadas las observaciones candidatas a ser eliminadas, se hace necesario establecer algún procedimiento de actuación (algoritmo) que permita determinar la observación o conjunto de observaciones más adecuadas a eliminar para mitigar el problema.
3. El investigador debe ser consciente de que puede ocurrir que no exista solución al problema planteado, es decir, no exista un conjunto de observaciones que tras ser eliminadas mitiguen el problema de multicolinealidad. O incluso valorar si se deben eliminar del análisis las observaciones que lo mitigan ya que puedan tener cierta relevancia en la muestra o suponga una disminución preocupante en el tamaño de la misma.

Obsérvese que en general, si se pudiese prescindir de  $K$  observaciones en la muestra, la selección de éstas de forma óptima no es un problema trivial, pues supondría resolver un problema de programación matemática del siguiente tipo:



$$\min_{j_1, \dots, j_K} \max_i FIV^{j_1, \dots, j_K}(i)$$

sujeto a  $FIV^{j_1, \dots, j_K}(i) \leq 10$

donde  $FIV^{j_1, \dots, j_K}(i)$  es el factor de inflación de la varianza de la  $i$ -ésima regresión auxiliar cuando se eliminan las observaciones  $j_1, \dots, j_K$ . Es decir, el problema anterior pretende seleccionar las observaciones  $j_1, \dots, j_K$  de forma que al eliminarlas del modelo se obtiene el mínimo FIV. El problema anterior supone resolver un problema de programación matemática no lineal y entero.

Por otro lado, no es claro que desde el punto de vista del modelo interese obtener el FIV más bajo, pues esto podría producir un efecto no deseado sobre la calidad del modelo estimado final.

## 6 REFERENCIAS BIBLIOGRÁFICAS

- GOLDBERGER, A. (1991). “A course in Economics”. Harvard University Press, Cambridge.
- MYERS, R.H. (1990). “Classical and modern regression with applications”, 2nd ed., PWS-Kent, pp. 130–133.
- PASHA, G.R. y ALI SHAH, M.A. (2004). “Application of ridge regression to multicollinear data”. Journal of Research Science, 15 (1), pp. 97–106.
- SPANOS, A. y MCGUIRK, A. (2002). “The problem of near-multicollinearity revisited: erratic vs systematic volatility”. Journal of Econometrics, 108 (2), pp. 365–393.