Anuario Jurídico y Económico Escurialense, XXXVIII (2005) 315-332 / ISSN: 1133-3677

Errores frecuentes en la interpretación del coeficiente de determinación lineal

Elena Martínez Rodríguez
Real Centro Universitario
«Escorial-María Cristina»
San Lorenzo del Escorial

Resumen: El objetivo de este trabajo es evidenciar, de forma sencilla a través de ejemplos numéricos, algunos de los graves errores que se cometen en el análisis de regresión, al abusar de la interpretación del coeficiente de determinación como única medida de la bondad del ajuste del modelo lineal estimado a un conjunto de datos.

Abstract: The aim of this project is to show, in an easy way and using numerical examples, some of the important mistakes committed in the interpretation of the regression analysis, due to the overuse of the determination coefficient as the one an only tool to measure the goodness fit of linear model estimated for a set of data values.

Palabras clave: Coeficiente de determinación lineal, Regresión, Bondad del ajuste, Error de Interpretación.

Keywords: linear determination coefficient, linear, Regression, Measure the goodness, Misunderstanding errors.

Sumario:

I. INTRODUCCIÓN

Una de las características de la realidad, sobre todo de la económica, es la relación que existe entre las distintas magnitudes que la definen. El análisis de la covariación entre variables, una *Y*, variable dependiente o endógena, y una o varias variables *X*, independientes o exógenas, supone obtener, en el caso de la regresión lineal, una ecuación lineal (o conjunto de ecuaciones lineales) que exprese la relación entre la variable endógena *Y* y las variables exógenas *X*. Se trata de encontrar la línea media que resuma o sintetice la dependencia entre la variable *Y* y las *X*, con la doble finalidad práctica de explicación o descripción causal de la variable dependiente y previsión de los valores futuros de *Y* para valores dados de *X*. Como línea media o medida de posición, debe acompañarse siempre de alguna medida de dispersión, que demuestre el grado en el que el promedio puede sustituir a las observaciones individuales de las que se obtuvo, esto es, que permita medir la bondad del ajuste realizado.

El desarrollo de la informática, la accesibilidad a ordenadores de gran potencia y a programas estadísticos y econométricos que facilitan los cálculos complejos han propiciado la generalización de los estudios de correlación y de regresión, incluso fuera del propio ámbito de la economía. De hecho, podemos encontrar Tesis Doctorales en las que el doctorando propone modelos de regresión para avalar las conclusiones de sus investigaciones, trabajos en los que los autores se valen de modelos de regresión para expresar la preferencia de los votantes o estudios clínicos en los que se intenta explicar la variación en la calidad de vida de los pacientes en función de las dosis tomadas de ciertos medicamentos.

El inconveniente de este uso generalizado lo encontramos cuando el investigador hace (generalmente por falta de un conocimiento más profundo) un mal uso de las medidas y técnicas de regresión. En este artículo pretendo poner de manifiesto de una manera sencilla, a través de ejemplos numéricos, algunos de los errores graves en el análisis de regresión a los que conduce la sola consideración del coeficiente de determinación, denominado R^2, como medida del grado de fiabilidad o bondad del ajuste del modelo ajustado a un conjunto de datos.

En el capítulo segundo se hará una breve presentación de este coeficiente y de cuál es su interpretación. En los capítulos siguientes se abordan distintas situaciones en las que claramente una inadecuada interpretación de R^2 puede llevarnos a situaciones como mínimo paradójicas. En concreto, en el capítulo 3 se analizan los efectos que estructuras determinadas del conjunto de observaciones, no detectadas por R^2, pueden tener sobre las aplicaciones empíricas de las técnicas de regresión. El capítulo 4 recoge la importancia que tiene trabajar con un número adecuado de grados de libertad del modelo ajustado, separando los problemas derivados del tamaño muestral de los derivados del número de variables exógenas incluidas en el modelo. El objetivo del capítulo 5 es poner de manifiesto la inconsistencia de una práctica cada vez más generalizada: buscar modelos de regresión con valores de R^2 elevados. Por último, el capítulo 6 se dedica a conclusiones.

II. Coeficiente de determinación: definición e interpretación

Si establecemos la hipótesis de que la mejor forma de describir la relación entre *X* e *Y* es mediante una línea recta, esto es:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

el problema inmediato que surge es el obtener los valores numéricos de los parámetros $\square_1$ y $\square_2$, que determinan la ecuación lineal concreta que expresa la relación de *Y* con *X*:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i = b_1 + b_2 X_i$$

Para ello acudimos a métodos de ajuste, básicamente el método de mínimos cuadrados [1], obteniendo un sistema de dos ecuaciones

$$\sum_i Y_i = nb_1 + b_2 \sum_i X_i$$

1. Novales, A., *Econometría*, Mc Graw-Hill, Madrid 1998.

$$\sum_i Y_i X_i = b_1 \sum_i X_i + b_2 \sum_i X_i^2$$

que permiten estimar los parámetros de la relación.

Ahora bien, el carácter de línea «media», que discurre entre las observaciones y que trata de sintetizarlas, que adquiere esta ecuación de regresión, obliga a que se acompañe, como cualquier promedio, de medidas de dispersión que permitan conocer el grado en que la misma puede sustituir a las observaciones de las que se obtuvo.

Así, podemos definir una primera medida de la dispersión de las $\hat{Y}_i$ observadas respecto a las «medias» Y_i calculada como la suma media de desviaciones cuadráticas entre ambas variables:

$$S_e^2 = \frac{\sum_i \left(Y_i - \hat{Y}_i \right)^2}{n}$$

expresión que recibe el nombre de varianza residual, ya que la diferencia

$$\left(Y_i - \hat{Y}_i \right)$$

mide el error (e_i) que cometemos al «sustituir» el valor observado por el valor estimado o ajustado mediante la regresión. A este error se le denomina también residuo.

Valores elevados de esta varianza indican que los residuos son grandes, lo que significa que la línea de regresión estimada se aleja mucho de los valores observados y, por tanto, la ecuación es poco representativa. Cuando es pequeña, dicha representatividad es elevada.

Por definición, se trata de una cantidad positiva (como cualquier varianza) acotada superiormente por el valor de la varianza de la variable observada Y, esto es:

$$0 \leq S_e^2 \leq S_Y^2$$

La cota superior es fácil de demostrar [2], ya que en el modelo de regresión lineal con ordenada se verifica la siguiente relación entre varianzas:

$$S_Y^2 = S_R^2 + S_e^2$$

---

2. LÓPEZ URQUÍA, J., y CASA ARUTA, E., *Estadística intermedia*, Vicens-Vives, Madrid 1969.

siendo S_R^2 la varianza explicada por la regresión, y cuya expresión matemática es:

$$S_R^2 = \frac{\sum_i \left(\hat{Y}_i - \bar{Y}_i \right)^2}{n}$$

A partir de esta varianza podemos definir una medida de dispersión relativa para la ecuación de regresión, comparando la misma con la varianza total de Y. Así lo que conocemos como coeficiente de determinación lineal se define por la expresión:

$$R^2 = \frac{S_R^2}{S_Y^2}$$

También podemos definir las relaciones anteriores mediante sumas de cuadrados, de forma que

$$STC = \sum_i \left(Y_i - \bar{Y} \right)^2$$

representa la variación total de los valores reales de Y respecto de su media muestral, recibiendo el nombre de suma total de cuadrados.

$$SEC = \sum_i \left(\hat{Y}_i - \bar{\hat{Y}} \right)^2 = \sum_i \left(\hat{Y}_i - \bar{Y} \right)^2$$

es la variación de los valores estimados de Y alrededor de su media, que se denomina suma de cuadrados debida a la regresión o explicada por la regresión. Y, por último,

$$SRC = \sum_i \left(Y_i - \hat{Y}_i \right)^2 = \sum e_i^2$$

es la variación residual o no explicada de los valores de Y alrededor de la recta de regresión, y que se conoce como suma de residuos al cuadrado. Así el coeficiente R^2 se puede definir como

$$R^2 = \frac{SCE}{SCT}$$

Cualquiera de estas dos expresiones permiten interpretar el coeficiente de determinación como la proporción o porcentaje de variación total en Y respecto a su media, que es explicada por el modelo de regresión. Es usual expresar esta medida en tanto por ciento, multiplicándola por cien.

Por su definición, es una medida acotada, siendo sus límites

$$0 \leq R^2 \leq 1$$

Un R^2 igual a 1 significa un ajuste lineal perfecto, ya que *STC=SEC,* esto es, la variación total de la variable *Y* es explicada por el modelo de regresión. El valor cero indica la no representatividad del modelo lineal, ya que $SEC = 0$, lo que supone que el modelo no explica nada de la variación total de la variable *Y*.

De las dos medidas de la bondad del ajuste del modelo lineal presentadas, la varianza residual y el coeficiente de determinación, es preferible este coeficiente. ya que la primera es una medida de carácter absoluto, por lo que su cuantía depende de la propia magnitud de la variable endógena. En cambio, R^2 es una medida adimensional, de fácil cálculo e interpretación, debido a su recorrido acotado entre cero y uno, lo que conduce a una profusa utilización de la misma, con interpretaciones abusivas en unos casos y erróneas en otros. Sin tratar de mermar la importancia de este coeficiente, R^2 debe tomarse, como veremos a lo largo de este artículo, como una primera medida, a completar con otras, para evaluar el modelo lineal de regresión ajustado y obtener conclusiones válidas sobre su grado de ajuste al conjunto de observaciones. Su exclusiva consideración puede, en muchas ocasiones, conducirnos a errores importantes en los análisis de regresión.

III. ESTRUCTURA DE LOS DATOS

Supongamos que deseamos conocer la relación que existe entre dos variables *X* e *Y*, que creemos es lineal, basándonos en la información proporcionada por una muestra de once observaciones conjuntas. Pero en lugar de trabajar con una única muestra, vamos a realizar, para valores prefijados de la variable exógena *X*, tres mediciones de la respuesta de la variable endógena *Y*, es decir, vamos a generar tres muestras diferentes [3]. La tabla I muestra los valores prefijados de *X,* así como los valores obtenidos de *Y,* en cada muestra.

3. Ejemplo basado en un ejemplo propuesto por Anscombe.

Dato	Variable *X* (valor prefijado)	Variable *Y* (muestra 1)	Variable *Y* (muestra 2)	Variable *Y* (muestra 3)
1	4	4,84	3,96	5,28
2	5	5,99	5,21	5,73
3	6	6,67	6,28	6,19
4	7	5,92	7,21	6,68
5	8	7,88	7,93	7,17
6	9	6,84	8,55	7,67
7	10	8,26	9,03	8,17
8	11	8,95	9,39	8,62
9	12	10,71	9,62	9,11
10	13	9,83	9,73	11,9
11	14	10,52	9,76	10,13

TABLA I

Realizando el ajuste lineal por el método de los mínimos cuadrados, para cada una de las tres muestras obtenemos la misma ecuación y el mismo valor para el coeficiente de determinación:

$$\hat{Y}_i = 2{,}5665 + 2{,}7565X_i \qquad\qquad R^2 = 0{,}8998$$

A la vista del resultado analítica podemos afirmar que el ajuste del modelo es bueno, ya que el valor de $R^2 = 0{,}8998$ es cercano a 1, en concreto, el 89,98% de la variabilidad de la variable *Y* a su promedio es explicado por el modelo de regresión ajustado. Podemos concluir que el modelo lineal es adecuado para describir la relación que existe entre estas variables.

Sin embargo, si añadimos a esta información cuantitativa sobre la que basamos nuestro análisis, la representación gráfica de los datos y la recta de regresión estimada para cada muestra veremos que la realidad es bien distinta.

La figura I muestra la representación gráfica de la nube de puntos correspondiente a la primera muestra. Resulta evidente que en este caso el modelo lineal ajustado describe de forma satisfactoria la relación entre ambas variables: todos los puntos se localizan de forma relativamente homogénea alrededor de la ecuación de regresión.

En el caso de trabajar con la segunda muestra, la representación gráfica de la nube de puntos (fig. II) pone en evidencia que el modelo lineal no es el más adecuado. Hemos cometido un error en la especificación del modelo, ya que sería más acertado proponer una ecuación parabólica y no una lineal.

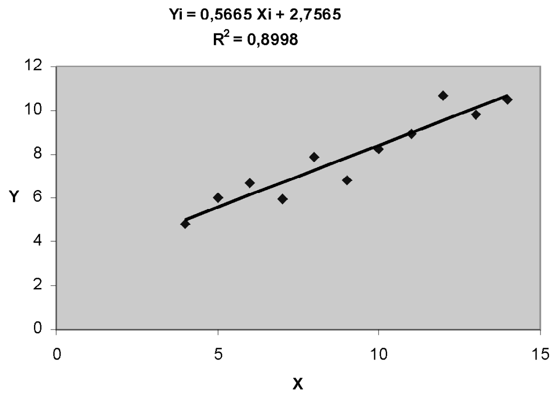

FIGURA I

Asimismo, en el caso de la tercera muestra, el gráfico (fig. III) revela la existencia de un dato anómalo que se aleja de la pauta seguida por el resto de las observaciones y que condiciona el resultado de la estimación. Es probable que el origen de esta anomalía esté en un simple error de medición, lo que debería ser comprobado.

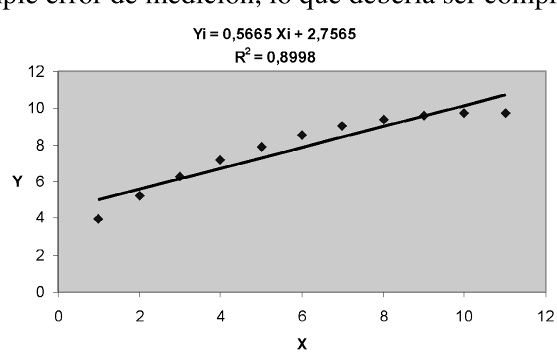

FIGURA II

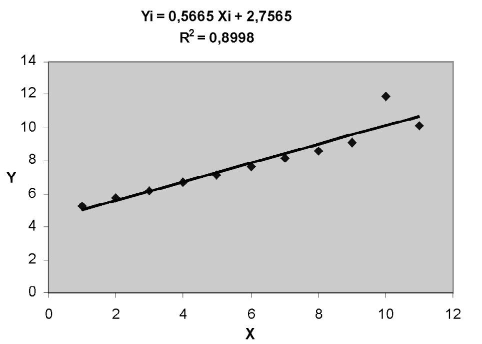

FIGURA III

Como vemos en este sencillo ejemplo, la estructura particular del conjunto de observaciones en las que nos basamos para estimar el modelo de regresión lineal condiciona el resultado de la estimación, de forma que tres situaciones bien diferentes nos han proporcionado las mismas estimaciones mínimas cuadráticas y el mismo valor de R^2.

La única consideración del coeficiente de determinación como medida de la bondad del ajuste nos lleva, en dos de las tres situaciones planteadas, a elaborar conclusiones equivocadas y a la afirmación de que es necesario contar con medidas adicionales a R^2 que completen el análisis.

IV. GRADOS DE LIBERTAD DEL MODELO

Un error que se comete frecuentemente en la interpretación de R^2 es el de no reparar en el tamaño muestral. El coeficiente de determinación lineal y el número de datos suelen variar de forma inversa, de tal manera que bastaría con considerar un número pequeño de observaciones para que R^2 alcance un valor próximo a la unidad, sin que ello evidencie la existencia de una marcada relación lineal entre dos variables.

Para resaltar esta idea recurro de nuevo a un sencillo ejemplo numérico, en el que queremos analizar la relación entre la variable exógena X y la endógena Y, aunque en este caso sabemos *a priori* que se encuentran relacionadas, casi exactamente, por una función parabólica. La tabla II recoge quince valores observados de la variable (X, Y).

Dato	**Variable *X***	**Variable *Y***
1	40	12
2	50	25,5
3	60	47,8
4	70	64,8
5	80	80,0
6	90	90,5
7	100	99,1
8	110	106,2
9	120	111,4
10	130	116,5
11	140	119,3
12	150	121,5
13	160	121,9
14	170	119,5
15	180	114,9

TABLA II

Vamos a empezar trabajando sólo con los seis primeros datos, para los que estimamos el modelo lineal a través del método de mínimos cuadrados. La figura IV recoge la representación gráfica del conjunto de observaciones y la recta ajustada. Podemos observar que el valor de R^2 igual a un 99% nos permite afirmar, apoyados en este caso por la gráfica, que el modelo lineal es adecuado.

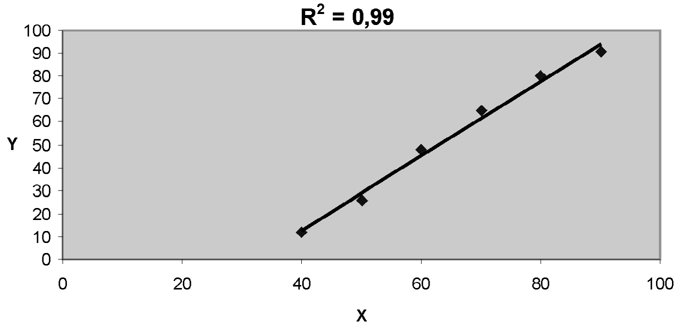

FIGURA IV

Igual conclusión podemos obtener (fig. V) si trabajamos en lugar de con seis datos con las doce primeras observaciones recogidas en la tabla II.

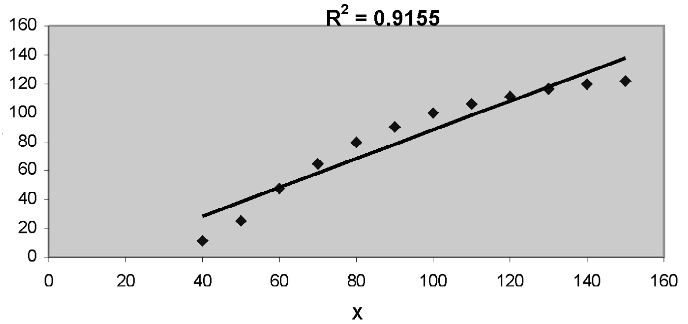

FIGURA V

Por último, repetimos el trabajo para las 15 observaciones muestrales. Sólo en este caso (fig. VI) el valor de R^2, que empieza a alejarse de 1 y la gráfica, nos permite dudar que el ajuste lineal sea adecuado.

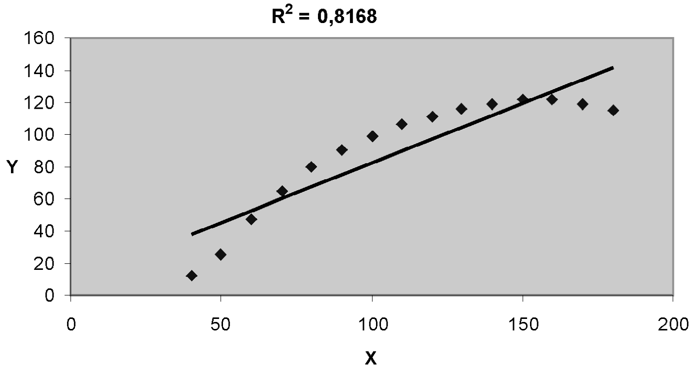

FIGURA VI

La consideración únicamente de R^2 para medir el grado de ajuste cuando trabajamos con muestras pequeñas nos conduce, de nuevo, a errores graves al aceptar la dependencia lineal entre las variables X e Y, cuando realmente mantienen una relación bien distinta.

Ahora bien, el tamaño muestral no es la única magnitud que influye en el valor de R^2. El número de variables explicativas consideradas en el modelo también condiciona el valor de este coeficiente, ya que R^2 es una función no decreciente del número de variables exógenas o regresoras presentes en el modelo, de forma que a medida que aumenta el número de variables regresoras R^2 aumenta. Su justificación es inmediata con tan sólo recordar su definición.

Según esto, R^2 mide la capacidad explicativa de la variable X sobre la variable Y. Al introducir en el modelo otra variable regresora el nivel explicativo será mayor entre las dos que sólo con la primera o, en todo caso, no disminuirá, pues la primera variable continúa como explicativa.

Así pues, en la interpretación de R^2 no sólo es preciso considerar el tamaño de la muestra, sino también el número de variables explicativas incluidas en el modelo de regresión. En una palabra, hay que tener en cuenta los grados de libertad del modelo, definidos como la diferencia entre el número de datos y el número de coeficientes de la ecuación.

En la literatura econométrica podemos encontrar varias soluciones al problema del incremento artificial del valor de R^2. Una de ellas es proponer un coeficiente de determinación corregido o ajustado, denotado por $\overline{R}^2$, y definido como

$$\overline{R}^2 = 1 - \frac{\dfrac{S_e^2}{n-k}}{\dfrac{S_Y^2}{n-1}}$$

donde k es el número de parámetros en el modelo, incluyendo el término independiente.

Es fácil comprobar la relación que mantienen R^2 y $\overline{R}^2$

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{k-1}$$

Observemos que para $k > 1$, $\overline{R}^2 < R^2$, lo cual implica que a medida que el número de variables exógenas aumenta, $\overline{R}^2$ ajustado aumenta menos que R^2 no ajustado. Observemos que, en este caso, ($k > 1$), si

$\overline{R}^2 = 0$ R^2 puede ser negativa, a pesar de que R^2 sea una magnitud no negativa. Si esto ocurre $\overline{R}^2$, se interpreta como si su valor fuese 0.

Establecida esta relación entre los dos coeficientes, podemos afirmar que R^2 corregido tiene la propiedad de ser neutral frente a la introducción de variables adicionales. En opinión de algunos autores [4], es mejor utilizar $\overline{R}^2$ en lugar de R^2, porque R^2 tiende a dar una imagen demasiado optimista del ajuste de la regresión, particularmente cuando el número de variables explicativas no es muy pequeño comparado con el número de observaciones, lo que podríamos considerar como un grado de libertad del modelo inadecuado. No obstante, esta opinión no es compartida totalmente, ya que se pueden proponer otras formas de corregir el aumento indeseado de R^2, como, por ejemplo, el coeficiente modificado que define Goldberger [5]:

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{k-1}$$

Ahora bien, debemos tener en cuenta que la utilización de cualquiera de los coeficientes alternativos propuestos, R^2 ajustado o R^2 corregido tiene problemas propios de interpretación y no resuelve siempre las deficiencias de R^2.

V. MAXIMIZACIÓN DE R^2

En ocasiones los investigadores tratan de maximizar R^2, es decir, escogen el modelo para el cual la R^2 es más elevada. Pero esto puede ser peligroso por varios motivos. En primer lugar, en el análisis de regresión el objetivo no es obtener un valor elevado de R^2, sino obtener estimadores precisos de los verdaderos coeficientes de regresión poblacional. En el análisis empírico no es raro encontrarnos con valores altos de R^2, pero tampoco que encontremos que alguno de los coeficientes de regresión no son estadísticamente significativos o muestran signos contrarios a los esperados *a priori*. El investigador

---

4. THEIL, H., *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, New York 1978.

5. GOLDBERGER, A. S., *A Course in Econometrics.Harvard* , University Press, Cambridge 1991.

debe preocuparse por la relevancia lógica o teórica que tienen las variables explicativas para la variable endógena y por su significación estadística. Si en este proceso se obtiene un valor de R^2 elevado, muy bien, aunque ello no es evidencia a favor del modelo, y si este valor es pequeño, esto no significa que el modelo sea necesariamente malo. Respecto a esta cuestión, señalar, sin entrar en más detalles, que la práctica de seleccionar un modelo con base en R^2 más elevada puede tener como consecuencia la introducción en el modelo de lo que se conoce como *sesgo preprueba*, que puede destruir algunas de las propiedades de los estimadores mínimos cuadrados del modelo de regresión lineal.

En segundo lugar, es importante señalar, al comparar modelos de regresión sobre la base del coeficiente de determinación, el tamaño muestral, y la variable dependiente deben ser los mismos. Por ejemplo, para los modelos

$$\ln Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i$$

los coeficientes R^2 no son comparables. La razón la encontramos en la propia definición de R^2, que mide la proporción de variación en la variable dependiente explicada por las variables exógenas. En el primer modelo considerado R^2 mide la proporción de la variación en Y explicada por X, mientras que en el segundo modelo mide la proporción de la variación de $\ln Y$.

Pero incluso cuando la variable endógena es la misma, podemos encontrarnos con problemas al comparar los valores de R^2 si, por ejemplo, el número de regresores es distinto. Pude probarse que cuando se añade una variable al modelo la suma residual siempre disminuye. Por tanto, si uno de los dos modelos tiene las mismas variables que el otro y alguna más (modelos anidados), como, por ejemplo, los modelos:

$$Y_i = \beta_1 + \beta_2 X_{1i} + u_i$$

$$\beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + u_i$$

el modelo amplio tendrá siempre un valor mayor de R^2, siendo por ello más preferido. El problema se vuelve más delicado en el caso en el que los modelos no sean anidados, siendo necesario recurrir a otros criterios y extremar la precaución en la interpretación que vamos a dar sobre cuál de los modelos es, en este caso, preferido.

Como vemos, son muchos los aspectos que debemos considerar cuando tratamos de comparar los valores de los coeficientes de determinación de distintos modelos: que traten de explicar la misma variable endógena, que ambos tengan o no ordenada en el origen, que el tamaño muestral sea el mismo, que compartan el mismo número de regresores, cuestiones que, en ocasiones, no son tenidas en cuenta por el investigador, quien basándose exclusivamente en el valor de R^2 como única medida del grado de bondad del ajuste realizado y, por tanto, en una interpretación errónea de este coeficiente decide trabajar con el modelo que proporciona un máximo valor de R^2. De nuevo un mal uso de este coeficiente puede conducir a conclusiones no acertadas.

V. Conclusiones

A través de ejemplos numéricos sencillos y de las representaciones gráficas de sus ajustes se ha tratado de transmitir la idea de que R^2 no es la medida «mágica» que resuelve, en todos los casos, el problema de la medición del grado de bondad del ajuste realizado. La propia estructura de los datos, desconocida *a priori,* y unos grados de libertad del modelo inadecuados (número reducido de observaciones y/o un número elevado de variables exógenas en el modelo) son algunas de las situaciones que nos han permitido poner de manifiesto las deficiencias y limitaciones mostradas por R^2 en cuanto a medida de la bondad del ajuste, al tiempo que se evidencia la necesidad de profundizar en el análisis econométrico, proponiendo medidas complementarias al coeficiente de determinación, de forma que su utilización conjunta garantice una mayor confianza en las conclusiones obtenidas. De hecho, varios autores comparten la idea de reducir el énfasis en el uso de R^2 como medida de bondad del ajuste al igual que su uso para comparar dos o más valores de este coeficiente con el objetivo de decidir qué modelo de regresión es preferido.

Las consideraciones destacadas en los distintos apartados no invalidan la utilización de R^2 como medida de la bondad del ajuste, si nos atenemos a interpretarlo de acuerdo con su definición: medida que recoge cómo en términos generales la recta de regresión ajustada resume o describe los datos.

Los problemas de interpretación surgen cuando intentamos que el coeficiente de determinación avale la dependencia entre variables y, a partir de ella, predecir o extrapolar, en el tiempo o en el espacio, la recta de regresión. En este sentido, si R^2 es alto se considera que el ajuste es válido y que la ecuación obtenida representa adecuadamente la relación cuantitativa entre las variables, pudiendo, por tanto, aplicarse para determinar los valores de una de ellas, conocidas las demás.

Este razonamiento es el que desvirtúa el uso y la interpretación del coeficiente de determinación lineal. Medida que, a pesar de las deficiencias que presenta, no debe ser desechada como medida de evaluación complementaria. Por otra parte, siempre que su interpretación se ajuste a su definición, su utilización será correcta.

VI. Bibliografía

ACHEN, C. H., *Interpreting and Using Regression.*, Dage Publicaciones, California 1982, pp. 56-67.

ASCOMBE, T. W., «Graphs in Statistical Analisys», en *The American Statistician*, 27 (1973) 17-21.

GOLDBERGER, A. S., *A Course in Econometrics. Harvard,* University Press, Cambridge 1991.

GRANGER, C., y NEWBOLD, P., «R^2 and the Transformation of the Regression Variables», en *Journal of Econometrics*, vol. 4 (1976) 205-210.

GUJARATI, D. N., *Econometría*, Mc Graw-Hill, México 2003.

HERNÁNDEZ ALONSO, J., «Uso y abuso del coeficiente de determinación», en *ESIC Market*, 79 (1993) 77-92.

LÓPEZ URQUÍA, J., y CASA ARUTA , E., *Estadística Intermedia*, Vicens-Vives, Madrid 1969.

NOVALES, A., *Econometría*, Mc Graw-Hill, Madrid 1998.

– Estadística y Econometría, Mc Graw-Hill. Madrid. 1996.

RAYMOND, J. L.; ANGULO, J., y REPILADO, A., «Relaciones de causalidad en economía y criterios estadísticos para detectar su existencia», *Instituto de Estudios Fiscales*, 16 (1982).

THEIL, H., *Introduction to Econometrics*, Prentice-Hall, Englewood Cliffs, New York 1978.