Revista Colombiana de Estadística
N° 17-18, 1988

INFERENCIA EN REGRESION NO LINEAL

Sergio Yañez Canal

Universidad Nacional de Colombia
Seccional Medellín

Resumen. Bajo el supuesto de errores i.i.d. $N(0,\sigma^2)$ los estimadores mínimo-cuadráticos de regresión lineal son los mejores estimadores lineales insesgados. Bajo idénticos supuestos estos resultados son ciertos en regresión no lineal, pero *asintóticamente*. Ahora bien, en muestras pequeñas, que es el caso común en la práctica, ninguna de dichas propiedades se cumple. Se presenta en este trabajo el porcentaje de sesgo de las estimaciones como medida de validez de las inferencias asintóticas. Se ilustra el método con un modelo de demanda residencial de energía eléctrica para Medellín.

Medellín, abril 1990.

1. Introducción. Sea,

$$y = X\beta + \varepsilon \tag{1.1}$$

un modelo de regresión lineal donde los ε_i, $i=1,\ldots,n$ son independientes e idénticamente distribuídos (i.i.d) $N(0,\sigma^2)$. Bajo estas condiciones se sabe que los estimadores mínimos cuadráticos (M.C.) de los parámetros son los mejores estimadores lineales insesgados, normalmente distribuídos y de mínima varianza, cualquiera sea el tamaño muestral "n".

Así se tienen los siguientes resultados:

$$\beta_M = (X'X)^{-1}X'Y \tag{1.2}$$

$$\beta_M \sim N(\beta,\sigma^2(X'X)^{-1}) \tag{1.3}$$

donde β_M es el vector de estimadores M.C. de β.

A partir de (1.3) se pueden construír todas las inferencias usuales sobre los parámetros de regresión.

Para el caso de la predicción de una nueva observación Y_o correspondiente a x_o, esto es $Y_o = x_o\beta$ se sabe que

$$\sigma^2(Y_o) = \sigma^2(1+x_o'(X'X)^{-1}x_o) \tag{1.4}$$

y de aquí se puede construir un intervalo de confianza para la predicción utilizando un estadístico "t".

Ahora bien, para modelos no lineales de regresión, los resultados obtenidos bajo la condición de errores i.i.d. $N(0,\sigma^2)$ son análogos al caso lineal pero de carácter asintótico.

Seber y Wild (1989) señalan que el carácter asintótico de dicha aproximación obliga a analizar qué pasa en el caso de muestras pequeñas, pues se conoce, que inclusive bajo la condición de que los errores sean i.i.d. $N(0,\sigma^2)$, los estimadores M.C. son sesgados, no se distribuyen normalmente y no son de mínima varianza.

Siguiendo a Ratkowsky (1983) se mostrará en este trabajo como el sesgo se puede utilizar como una medida del grado de no linealidad del modelo, para determinar la validez de las inferencias, cuya justificación reposa en los supuestos de linealidad.

2. Los Mínimos Cuadrado en Regresión no Lineal.

Sea,

$$y_1 = f(x_1;\beta)+\varepsilon_1 \qquad (i = 1,2,\ldots,n) \tag{2.1}$$

donde y_1 es la variable dependiente, ε_1 es el error aleatorio con $E(\varepsilon_1) = 0, x_1$ es un vector $N\times 1$ de variables independientes y β es un vector $K\times 1$ de parámetros.

Se considerarán modelos de regresión intrínsicamente no lineales, es decir que no pueden ser transformados en modelos lineales.

El siguiente modelo de demanda de energía eléctrica en el sector residencial (que se usará en la sección 4) es un ejemplo

$$Q = k\left[\alpha C^{\beta/D} Y^{\tau}\right]^{D/(D-\beta)} + \varepsilon \tag{2.2}$$

Obsérvese que en (2.2) el número de parámetros no coincide con el número de variables independientes y por lo tanto para los modelos no lineales, en general, no existe un forma matricial equivalente a (1.1). Los estimadores M.C. de β, escritos β_M, minimizan la suma de cuadrados de los errores

$$S(\beta) = \sum_i [y_i - f(x_i;\beta)]^2 \tag{2.3}$$

pero en el caso no lineal se obtienen por medio de una aproximación lineal, que se presenta en la sección 2.1. Los resultados númericos se obtienen por medio de métodos interativos, variantes del método de Gauss-Newton.

Los estimadores M.C. en regresión no lineal, son utilizados para hacer inferencias aproximadas aplicando resultados asintóticos a muestras finitas. Se puede probar que bajo ciertas condiciones (no se exige normalidad de los errores) son estimadores consistentes y asintóticamente normales, ver por ejemplo Judge et al (1982). Si se asume normalidad de los errores entonces los estimadores M.C. son también estimadores máximo-verosímiles.

2.1. La aproximación lineal.

Se usará la siguiente notación

$$f_i(\beta) = f(x_i;\beta), \tag{2.4}$$

$$f(\beta) = (f_1(\beta), f_2(\beta), \dots, f_n(\beta)) \tag{2.5}$$

$$F(\beta) = \frac{\delta f(\beta)}{\delta\beta'} = \left[\left(\frac{\delta f_i(\beta)}{\delta\beta_j}\right)\right]_{n\times k} \tag{2.6}$$

$$F = F(\beta^*); F_M = F(\beta_M) \tag{2.7}$$

donde β^* es el verdadero valor de β y β_M es el estimativo mínimo cuadrático.

La idea es aproximar $f_i(\beta)$ por el término lineal de la expansión en series de Taylor, que es la mejor aproximación lineal:

$$f_i(\beta) \approx f_i(\beta^*) + \sum_{j=1}^{k} \left[\frac{\delta f_i(\beta)}{\delta\beta}\right](\beta_j - \beta_j^*) \tag{2.8}$$

$$f(\beta) \approx f(\beta^*) + F(\beta - \beta^*) \tag{2.9}$$

y se puede obtener

$$y^* = F\beta + \varepsilon \tag{2.10}$$

donde

$$y^* = y - f(\beta^*) + F\beta^*$$

Malinvaud (1970) llama a (2.10) el seudo-modelo lineal, que es el análogo de (1.1) en regresión lineal y así se obtiene el análogo de (1.2)

$$\beta_M = (F'F)^{-1}F'y^* \tag{2.11}$$

El valor de β_M suficientemente cercano a β^* se obtiene por métodos interativos a partir de (2.9) minimizando la suma de cuadrados (2.3). Una completa descripción de los distintos métodos computacionales para M.C. no lineales, así como una discusión sobre los valores iniciales se puede encontrar en Seber y Wild (1989).

Es claro de esta aproximación, que existe una discrepancia entre β_M y β^*, de la cual podemos obtener el sesgo de los estimativos de los parámetros en el método M.C. en regresión no lineal. En la sección 3 se examinará dicho sesgo.

2.2. Inferencia asintótica.

Dado que los ε_i sean i.i.d. $N(0,\sigma^2)$ en (2.1) se puede probar para n grande

$$\beta_M \sim N_k(\beta, \sigma^2(F'F)^{-1}) \qquad (2.12)$$

resultado análogo a (1.3) y a partir del cual se pueden construír todas las inferencias usuales sobre los parámetros de regresión. F se estima por F_M y σ^2 por $s^2 = S(\beta_M)/(n-k)$.

Para el caso de la predicción de una nueva observación y_o correspondiente a x_o, Seber y Wild (1989) partiendo de la expansión en series de Taylor

$$f(x_o;\beta_M) \approx f(x_o;\beta) + f_o'(\beta_M - \beta) \qquad (2.13)$$

donde f_o' es el vector $1\times K$ de primeras derivadas de $f(x_o;\beta)$ con respecto a cada uno de los elementos de β, prueba que

$$\sigma^2(y_o) = \sigma^2(1 + f_o'(F'F)^{-1}f_o) \qquad (2.14)$$

resultado análogo a (1.4) y del cual se puede construír un intervalo de confianza para la predicción utilizando un estadístico "t".

3. El Sesgo en Regresión no Lineal.

Box (1971) partiendo de una expansión del modelo en series de Taylor hasta el término de segundo orden, encontró la siguiente fórmula para el sesgo de β_M.

$$\text{sesgo}(\beta_M) = E(\beta_M - \beta^*) = -\frac{\sigma^2}{2}\Big(\sum_u f_u f_u'\Big)^{-1}\sum_t f_t \,\text{tr}\left[\Big(\sum_u f_u f_u'\Big)^{-1} H_t\right] \qquad (3.1)$$

donde $f_u (= f_t)$ es el vector $k\times 1$ de primeras derivadas de $f(X_t;\beta)$ y H_t es la matriz $k\times k$ de segundas derivadas con respecto a cada uno de los elementos de β, evaluados en X_t, donde

$t = 1,2,\ldots,n$. En la práctica s^2 y β_M son usados en lugar de σ^2 y β.

Para la predicción y_o, Box (1971) derivó la fórmula

$$\text{sesgo}(y_o) = E(y_o - f(x_o;\beta)) = f_o'\,\text{sesgo}(\beta_M) + \tfrac{1}{2}\,\text{tr}\left[H_o \,\text{Cov}(\beta_M)\right] \tag{3.2}$$

donde $\text{Cov}(\beta_M) = s^2(F_M'F_M)^{-1}$ es la matriz de varianza-covarianza de β_M.

Ratkowsky (1983 , muestra por medio de estudios de simulación que el porcentaje de sesgo

$$\%\text{sesgo}(\beta_{iM}) = \frac{\text{sesgo}(\beta_{iM})(100)}{\beta_{iM}} \tag{3.3}$$

es una cantidad útil en la medida en que un valor absoluto en exceso del 1%, es una buena regla práctica para determinar el grado de no linealidad del modelo. Retkowsky (1983) muestra, también, que en el caso de la proyección

$$\%\text{sesgo}(y_o) = \frac{\text{sesgo}(y_o)(100)}{y_o} \tag{3.4}$$

se puede utilizar en el mismo sentido de (3.3) para dicidir si las inferencias asintóticas de la sección 3 son válidas.

4. Aplicación.

El modelo que se usará es tomado de Vélez et al (1987). La demanda residencial de Energía Eléctrica en dos ciudades colombianas: un modelo económico. Artículo basado en el segundo capítulo de Botero et al (1986).

El modelo para el caso de Medellín es el siguiente:

$$Q_t = k\left[\alpha C_t^{\beta/D_t} Y_t^{\tau}\right]^{D_t/(D_t-\beta)} + \varepsilon_t \tag{4.1}$$

donde t = 1970, 1971,...1983; n = 14 datos; Q_t = consumo del subscritor medio; C_t = representa el intercepto de la función de oferta cuando el precio es uno; D_t = elasticidad de la oferta respecto al precio; Y_t = ingreso per capita; k = parámetro constante; α = parámetro que representa el efecto de las preferencias y necesidades de los subscriptores sobre la demanda; β = parámetro que representa la elasticidad de la demanda con respecto al precio; τ = parámetro que representa la elastididad de la demanda con respecto del ingreso; ε_t = término de error del modelo.

El modelo se ajustó utilizando el modulo NONLIN del paquete estadístico SYSTAT que utiliza algoritmos cuasi-Newton para encontrar los estimadores M.C., al respecto ver Wilkinson (1988).

El ajuste fue satisfactorio razonando por analogía al caso lineal. Con base en los resultados de consumo de energía eléctrica en el sector residencial para Medellín de 1984 y 1985 que ya se conocían, se observaba el buen comportamiento del modelo respecto a predicciones. La interpretación de los parámetros desde el punto de vista económico concluía que los estimativos eran perfectamente aceptables. Así utilizando criterios estadísticos y económicos se concluyó que el modelo estaba bien especificado.

Para el caso de interés de este trabajo, calcularemos el porcentaje de sesgo para los parámetros y la predicción para determinar la validez de las inferencias que se hicieron siguiendo los criterios asintóticos expuestos en la sección 3, siendo el tamaño muestral en este caso pequeño, n = 14 datos. Ratkowsky (1983) observa que para un parámetro que represente un término constante en el modelo, el porcentaje de sesgo puede ser arbitrariamente grande o pequeño y por ello recomienda que este tipo de análisis no se haga con dichos parámetros.

Así, utilizaremos el siguiente modelo:

$$Q = 2.8\left[\alpha C^{\beta/D} Y^{\tau}\right]^{D/(D-\beta)} + \varepsilon \tag{4.2}$$

donde k = 2.8 es el valor obtenido al ajustar (4.1) Se ajustó (4.2) utilizando el modulo NONLIN del SYSTAT y se observa que el ajuste es bueno, razonando por analogía al caso lineal.

En el cuadro 1 se puede observar que el porcentaje de sesgo para β y τ superan el 1% lo cual indica un grodo de no-linealidad alto en el modelo. El caso de τ es 18.8%, (además comparando dicho sesgo con el error estandar se tiene un resultado de 4.4) lo cual indica que la influencia de τ sobre la no-linealidad es muy grande y sugiere que el modelo no está bien especificado respecto a τ y a la variable asociada Y. Obsérvese, también, en el cuadro 2, la alta influencia de τ sobre las proyecciones: si corregimos los sesgos de los parámetros, el sesgo obtenido al comparar las proyecciones con los parámetros originales y las proyecciones con los parámetros corregidos es del orden del 36% para los años desde 1984 a 1990. También en el cuadro 2 se ve que la sobre-estimación de las proyecciones es debida exclusivamente a τ.

CUADRO 1.

PARAM.	VALOR	E.S.	SESGO	%SESGO	ABS(SES/E.S.)
α	168.040	16.661	-0.606872550	-0.361%	0.03642473
β	-0.048	0.010	-0.001989626	4.145%	0.19896262
τ	0.306	0.013	0.057778291	18.882%	4.44448398

CUADRO 2

AÑO	PROYEC.	τ=0.248222	β=.046011	τ=0.248222 α=168.6468 β=.046011	SESGO	%SESGO
1984	5294.056	3392.065	5267.760	3389.161	1904.895	35.982%
1985	5468.177	3494.303	5433.661	3487.030	1981.147	36.230%
1986	5522.694	3522.537	5487.769	3515.171	2007.524	36.350%
1987	5591.336	3558.010	5555.893	3550.527	2040.809	36.499%
1988	5652.247	3589.420	5616.345	3581.832	2070.415	36.630%
1989	5705.114	3616.630	5668.812	3608.951	2096.162	36.742%
1990	5758.472	3644.044	5721.766	3636.274	2122.198	36.853%

Ahora bien, respecto al sesgo de las proyecciones, se puede decir que el efecto del sesgo de τ, sobreestima las proyecciones, como ya de observó en el cuadro 2. En la fórmula (3.2) el efecto del sesgo de los parámetros se ve en el cuadro 3 en la columna (δ*SPA) y el efecto de la aproximación de la columna (TRAZA/2), es claro que el porcentaje de sesgo se debe completamente a τ y es de una magnitud similar a la señalada en el cuadro 2. (La columna E.S(PRO) del cuadro 3, muestra el error estandar de las proyecciones calculadas de acuerdo a (2.14) y son resultados aparentemente aceptables).

Desde el punto de vista estadístico se concluye que el modelo debe ser revisado, concretamente la especificación respec to al ingreso per-cápita (Y) la variable asociada al parámetro τ. Tal como está, el modelo es altamente no-lineal y las inferencias respecto a los parámetros y a la predicción no tienen ninguna validez estadística, a pesar de que los diagnósticos realizados sobre (4.2) eran "buenos", utilizando criterios análogos a los utilizados en regresión lineal.

CUADRO 3

AÑO	PROYEC.	E.S(PRO)	δ*SPA	TRAZA/2	SESGO
1984	5294.056	397.239	2310.671	-45.175	2265.496
1985	5468.177	443.637	2393.719	-44.586	2349.133
1986	5522.694	447.376	2427.863	-45.224	2382.639
1987	5591.336	452.118	2470.996	-46.030	2424.966
1988	5652.247	456.357	2509.404	-46.748	2462.657
1989	5705.114	460.059	2542.839	-47.372	2495.467
1990	5758.472	463.817	2576.679	-48.004	2528.675

AÑO	%(δ*SPA)	%(TR/2)	%SESGO
1984	43.647%	-0.853%	42.793%
1985	43.775%	-0.815%	42.960%
1986	43.962%	-0.819%	43.143%
1987	44.193%	-0.823%	43.370%
1988	44.397%	-0.827%	43.570%
1989	44.571%	-0.830%	43.741%
1990	44.746%	-0.834%	43.912%

Este ejemplo ilustra con claridad, como a pesar de que el trabajo en regresión lineal se hace por analogía a la regresión lineal, sus estimadores se comportan de manera *completamente* diferente, dependiendo del grado de no linealidad en muestras pequeñas, de forma que las inferencias asintóticas pueden carecer de toda validez.

5. Consideraciones Finales.

El estudio de los modelos no lineales de regresión es un campo relativamente nuevo. Su desarrollo importante es de la decada de los 80's, a partir de los trabajos de Bates y Watts (1980) y Ratkowsky (1983) quienes utilizando los trabajos pioneros de Beale (1960) y Box (1971) presentan reglas prácticas para la determinación del grado de no-linealidad del modelo. El método expuesto en este artículo es desarrollado por Ratkowsky (1983) y se recomienda se utilice conjuntamente con las medidas de curvatura de Bates y Watts (1980) así como también, con estudios de simulación sobre las propiedades muestrales de los estimadores M.C.. Se escogió, en este trabajo, la medida de porcentaje de sesgo por aparecer como más natural desde el punto de vista estadístico para ilustrar el hoy complejo campo de la regresión no-lineal. Es pertinente anotar que Bates y Watts (1980) muestran la relación del sesgo aquí utilizado, con sus medidas de curvatura.

En el momento de la implementación del modelo de la sección 4, la bibliografía que conocíamos escasamente citaba el artículo de Bates y Watts (1980), pero no destacaba la importancia central de dicho artículo en el desarrollo de la regresión no-lineal, hoy claramente reconocida. Las referencias más utilizadas fueron: Draper y Smith (1981) y Amemiya (1983), este último especialista en modelos no lineales en econometría. Miller (1974) señalaba la no validez del jackknife en modelos no lineales como reductores de sesgo. Simonoff y Tsai (1986) desarrollan métodos basados en jackknife para regresión no lineal, teniendo en cuenta los efectos de no linealidad. Los aná lisis de residuales son revisados por Cook y Tsai (1985) al incluir las medidas de no linealidad. En fin, se puede afirmar con Seber y Wild (1989) que hasta hace pocos años la situación de los modelos no lineales era en general deficiente y que las medidas de no linealidad son una de las principales razones de su reciente desarrollo.

Para terminar se puede citar el epígrafe, al capítulo 7º sobre Medidas de Curvatura de no linealidad, del libro de Bates y Watts (1988): "La gran tragedia de la Ciencia: la muerte violenta de una bella hipótesis por una fea realidad". Thomas Huxley.

BIBLIOGRAFIA

Amemiya, T., (1983). "Non-Linear Regression Models" capítulo 6º en Griliches, Z. e Intrilligator, M.D. Handbook of Econometric, volumen 1. North Holland: Amsterdan.

Bates, D.M. y Watts, D.G., (1980). Relative curvature measures of nonlinearity. J.R. Stat. Soc. B, 42, 1-25.

Bates, D.M. y Watts, D.G., (1988). Nonlinear Regression Analysis & its aplications. Wiley: New York.

Beale, E.M.L., (1960). Confidence regions in non-linear estimation. J.R. Stat. Soc. B, 22, 41-88.

Botero, J., Velez, C.E., Garcia, G., Castaño, E. y Yañez, S., (1986). Revisión y reestimación del submodelo de demanda de energía eléctrica en Colombia. Copia a máquina. Medellín: Centro de Investigaciones Económicas. Universidad de Antioquia.

Box, M.J., (1971). Bias in nolinear estimation. J.R. Stat. Soc. B, 33, 171-201.

Cook, R.D. y Tsai, C.L., (1985). Residuals in nolinear regression. Biometrika, 72, 23-29.

Draper, N.R. y Smith, H., (1981). Applied Regression Analysis, 2a. Ed. Wiley: New York.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., y Lee, T.C., (1982). Introduction to the theory and practice of econometrics. Wiley: New York.

Malinvaud, E., (1970). Statistical Methods of Econometrics. North Holland: Amsterdam.

Miller, R.G., (1974). An Unbalanced jackknife. The Annals of statistics, 2, 880-891.

Ratkowsky, D.A., (1983). Nonlinear Regression Modeling. Marcel Dekker: New York.

Simonoff, J.S. y Tsai, C.L., (1986). Jackknife - Bosed estimators and confidence regions in Nonlinear Regression. Technometrics, 28, 103-112.

Seber, G.A.F., Wild, C.J. (1989). Nonlinear Regression. Wiley New York.

*