LECTURAS
DE
ECONOMIA

Elkin Castaño Vélez

Sesgos de transformación en el ajuste de modelos no-lineales

Lecturas de Economía. No. 16. Medellín, enero-abril de 1985. pp. 207-218

• **Resumen.** Al tratar de simplificar los problemas de estimación en relaciones no-lineales, el economista recurre con frecuencia a utilizar transformaciones sobre las variables originales de forma tal que el modelo con las nuevas variables sea lineal. Sin embargo, estas transformaciones producen sesgos importantes al obtener, mediante proceso de retransformación, el modelo original ajustado. Este artículo muestra cómo corregir en forma fácil gran parte de estos sesgos en algunos modelos frecuentemente utilizados. En la práctica estas correcciones actúan sobre los sesgos proporcionando modelos más adecuadamente ajustados.

Fitting No-Linear Models and the Transformation Bias

• **Abstract.** *When the economist tries to simplify non-linear relations frecuently uses transformations of original variables in order to linearize the model. Nevertheless, the filled original model obtained by retransformation contains significant biases. This paper shows how to correct those biases in some models frecuently used. The corrective actuates over the biases in order to improve the fittness of the model.*

INTRODUCCION

Es de uso frecuente entre los economistas utilizar transformaciones de las variables originales para tratar de simplificar los problemas de estimación en el ajuste de una relación no-lineal. Estas transformaciones son generalmente elegidas de forma tal que el modelo expresado en términos de las variables transformadas sea lineal. Ahora, el interés del economista no radica, en general, en este nuevo modelo, sino en la relación no-lineal original y por tanto debe aplicar alguna clase de transformación contraria que le devuelva la relación original, después que el modelo linearizado ha sido ajustado. El propósito de este artículo es mostrar que el proceso de retransformación hacia la relación no-lineal produce sesgos en el modelo ajustado original que pueden llegar a ser importantes y que son en gran parte corregibles. Se presentará un ejemplo para mostrar la necesidad de corregir dichos sesgos.

1. El proceso de linearización

Conocemos que existe una relación no-lineal entre X y Y, digamos $Y = f(X)$, donde consideramos que X es una variable no estocástica (este conocimiento puede estar basado en la teoría económica o en el uso de diagramas de dispersión entre los datos de las variables X y Y). Elegimos una transformación lineal adecuada ya sea sobre la variable dependiente o la dependiente o sobre ambas. Se ajusta entonces el modelo lineal y se prueba su validez siempre y cuando se cumplan las hipótesis necesarias acerca del término de perturbación del modelo linearizado. Este proceso está descrito en muchos textos de econometría (véase H. Theil 1978) o de estadística (véase Draper-Smith 1981). Generalmente la atención se concentra en la elección de la transformación apropiada para lograr linearidad y un comportamiento adecuado de los residuales (véase Box-Cox. 1964 y Mosteller-Tukey 1977).

Una vez que el modelo ha sido ajustado, ¿cuáles son las razones por las cuales se generan sesgos al tratar de obtener el modelo original ajustado?

Ilustremos esto con un ejemplo. Supongamos que el modelo no-lineal adecuado es

$$Y = Ae^{\beta_1 X}e^{\mu} \qquad (1)$$

Este modelo es fácilmente linearizado tomando el logaritmo natural,

$$\ln Y = \beta_o + \beta_1 X + \mu \qquad (2)$$

Utilizando mínimos cuadrados ajustamos (2) y obtenemos

$$\hat{\ln} Y = \hat{\beta}_o + \hat{\beta}_1 X \qquad (3)$$

y el modelo inicial ajustado es

$$\hat{Y} = \hat{A}e^{\hat{\beta}_1 X} \qquad (4)$$

donde $\hat{A} = e^{\hat{\beta}_o}$

Si nuestro interés es obtener el estimador de la respuesta media de Y dado X, el estimador generalmente utilizado es (4).

En realidad (4) es un estimador de la respuesta mediana de Y dado X, como lo veremos pronto. Por tanto el sesgo al estimar la respuesta media de Y utilizando directamente el modelo retransformado tiene la misma base teórica que el sesgo introducido al utilizar la mediana muestral como estimador de la media poblacional.

2. Algunos sesgos debido a la transformación y su remedio

Logaritmo

Si elegimos la transformación logarítmica como la apropiada, el modelo lineal que ajustamos es de la forma

$$l\hat{n}Y = \hat{\beta}_O + \hat{\beta}_1 X$$

donde $\hat{\beta}_O$ y $\hat{\beta}_1$ son los estimadores mínimos cuadráticos de los coeficientes del modelo lineal. Si hacemos la transformación contraria obtenemos

$$\hat{Y} = \hat{A}e^{\hat{\beta}_1 X} \qquad (5)$$

donde $\hat{A} = e^{\hat{\beta}_O}$

Veamos el sesgo de este modelo transformado.

$$lnY = \beta_O + \beta_1 X + \mu \qquad (6)$$

el cual después de retransformarlo se convierte en el modelo

$$Y = Ae^{\beta_1 X}\mu^*$$

donde $A = e^{\beta_O}$ y $\mu^* = e^{\mu}$

Bajo el supuesto de normalidad en el término de perturbación $\mu^* = e^{\mu}$ tiene una distribución lognormal (véase Mood, Graybill y Boes 1974). Entonces la respuesta media de Y es simplemente

$$E(Y) = Ae^{\beta_1 X} E(\mu^*),$$ bajo el supuesto que X es una variable aleatoria no-estocástica.

$$E(Y) = Ae^{\beta_1 X} e^{1/2\sigma^2} \qquad (7)$$

Ahora la respuesta mediana de Y, *Med(Y)*, está dada por

$$Med(Y) = Ae^{\beta_1 X} Med(\mu^*)$$

$$= Ae^{\beta_1 X} \qquad (8)$$

dado que la mediana de μ^* es 1.

Generalmente se acostumbra a utilizar $\hat{A}e^{\hat{\beta}_1 X}$ como el estimador de $E(Y)$, es decir utilizamos $M\hat{e}d(Y)$ como estimador de $E(Y)$.

Comparando (7) y (8) podemos ver el sesgo que se introduce al estimador $E(Y)$ utilizando $M\hat{e}d(Y)$. Se puede probar fácilmente que $M\hat{e}d(Y)$

es un estimador consistente de $E(Y)$ pero que subestima a $E(Y)$. El factor de sesgo en este caso es $e^{1/2\sigma 2}$ el cual es multiplicado y crece exponencialmente con σ^2.

Un remedio simple es aplicar un estimador de este factor al estimador retransformado (5)

$$E(\hat{Y}) = \hat{A}e^{\hat{\beta}_1 X}\ e^{1/2\,\hat{\sigma}^2}$$

$$E(\hat{Y}) = \hat{A}e^{\hat{\beta}_1 X\ +\ 1/2\,\hat{\sigma}^2}$$

con $\hat{A} = e^{\hat{\beta}o}$ y donde $\hat{\beta}\sigma, \hat{\beta}_1$ y $\hat{\sigma}^2$ son los estimadores mínimo-cuadráticos del modelo linearizado (6). Este modelo permanece aún con un sesgo, puesto que $\hat{A}$ y $e^{\hat{\beta}}$ no son estimadores insesgados de A y e^{β_1}, respectivamente, pero al aplicar el factor de ajuste se reduce la mayor parte del sesgo.

Como un caso particular de este modelo se encuentra el modelo logarítmico doble o el modelo de elasticidad constante; para obtener esto basta hacer X igual a lnX en el modelo (1) el cual es de la forma

$$Y = AX^{\beta_1}\ e\mu \qquad (9)$$

Para este modelo $\quad E(Y) = AX^{\beta_1}\ e^{1/2\,\hat{\sigma}^2}$, con $A = e^{\beta o}$

3. Un ejemplo

Para ilustrar la necesidad de corregir sesgos debido a la transformación, veamos un ejemplo basado en datos reales de los ingresos fiscales de los municipios del departamento de Antioquia (Colombia) en el cual se pretende investigar la incidencia del ingreso total de los municipios sobre el número de alumnos matriculados en establecimientos oficiales del respectivo municipio. Para el estudio se excluyeron los municipios del Area Metropolitana (Medellín, Barbosa, Bello, Caldas, Copacabana, Envigado, Girardota, Itagüí, La Estrella y Sabaneta) con el objeto de controlar el problema estadístico derivado del flujo de personas que viven en un municipio pero estudian en otro.

En este caso tenemos:

Y = Número de alumnos matriculados en los establecimientos oficiales de las cabeceras municipales del departamento de Antioquia en 1980 (sin incluir los municipios del Area Metropolitana).

X = Ingresos totales de los fiscos municipales del departamento de Antioquia (sin incluir los municipios del Area Metropolitana), en millones de pesos en 1980.

Se eligió una muestra aleatoria de 39 municipios y los datos fueron tomados del *Anuario Estadístico de Antioquia. 1980.* El modelo empleado fue

$Y = AX^{\beta_1} e^{\mu}$, o sea el modelo o doble logarítmico.

El modelo linearizado es

$$l\hat{n} Y = \beta_O + \beta_1 \ lnX + \mu, \text{ con } A = e^{\beta_O}$$

El modelo linearizado ajustado es

$$l\hat{n}y = 4.128471 + 0.926417 \ lnX$$

para éste se obtuvo un buen comportamiento de los residuales y no aparecieron datos extremos, es decir datos que influencian la regresión.

El modelo retransformado ajustado es

$$\hat{Y} = e^{4.128471} X^{0,926417}$$

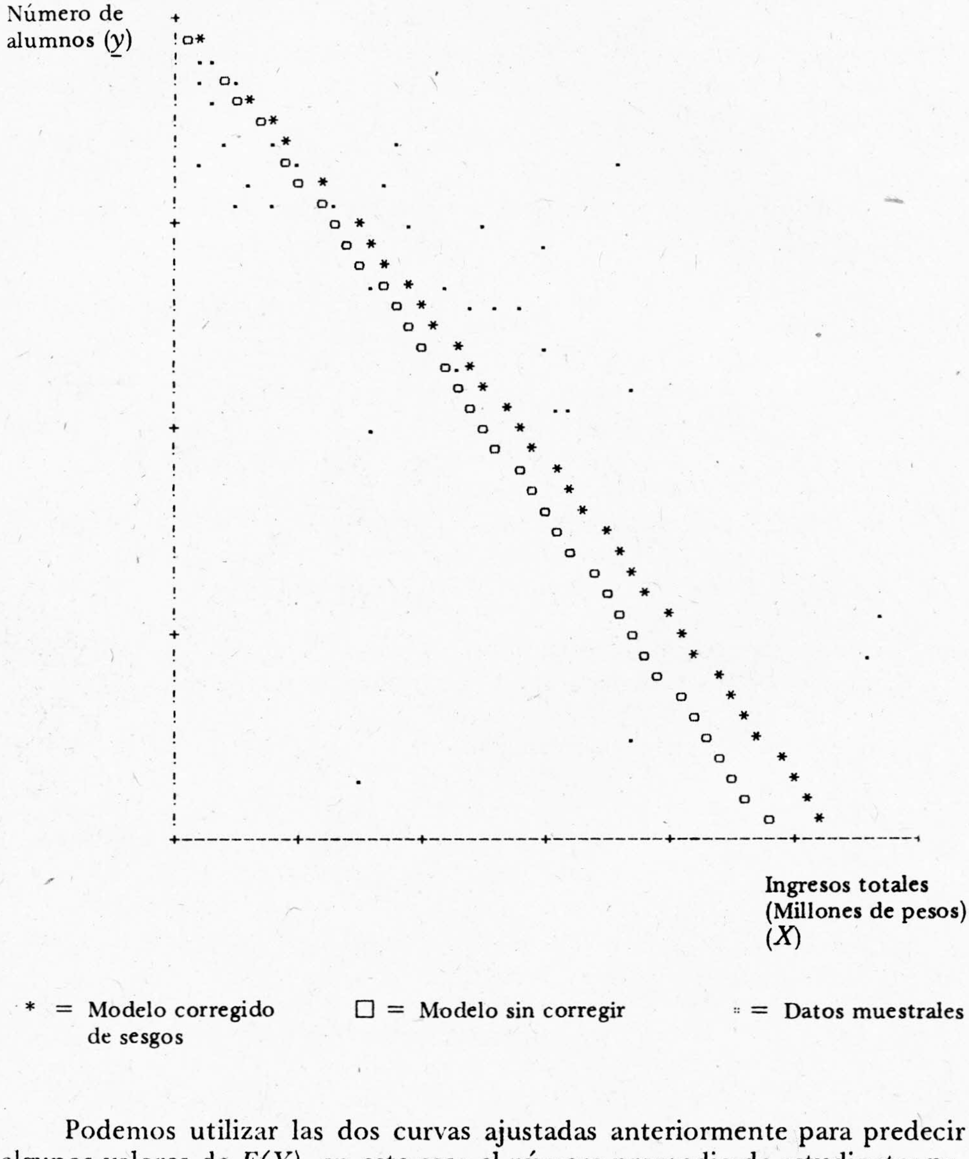
$$\hat{Y} = 62.082926 \ X^{0,926417}$$

El modelo retransformado corregido, utilizando el factor de corrección que para este caso es $e^{1/2\hat{\sigma}^2}$, es de la forma

$$\hat{Y} = 67.80\ 7377 \ X^{0,926417}$$

La curva más baja, como estimador de la respuesta media, parece estar sesgada hacia abajo (véase Gráfico No. 1). Cuando utilizados el factor para la reducción de sesgos se obtiene la curva más alta, la cual es un estimador con menos sesgo de la curva $E(Y)$.

Gráfico No. 1
COMPARACION GRAFICA DE MODELOS AJUSTADOS



* = Modelo corregido de sesgos   □ = Modelo sin corregir   ∶ = Datos muestrales

Podemos utilizar las dos curvas ajustadas anteriormente para predecir algunos valores de $E(Y)$, en este caso el número promedio de estudiantes para un nivel X de ingreso total.

Veamos algunos ejemplos y cuál es la predicción en ambos casos.

Cuadro No. 1
PREDICCIONES UTILIZANDO LOS MODELOS AJUSTADOS

Número	Valor del Ingreso (Millones de pesos)	Predicciones del número medio de estimaciones Sin corregir	Corregido	Valor Real
1	5.0	275.8	301.2	411
2	6.0	326.5	356.6	371
3	6.9	366.6	400.4	355
4	8.3	441.0	481.6	448
5	9.7	509.5	556.5	686
6	9.8	514.4	561.8	795
7	10.1	528.9	577.7	592
8	11.0	577.8	630.5	614
9	11.7	602.1	622.0	1045
10	12.2	630.1	688.2	1198
11	15.9	805.3	879.6	1054
12	16.2	819.4	894.9	1079
13	20.0	996.0	1088.0	1513
14	37.0	1764.1	1923.5	2065
15	39.1	1853.5	2024.4	1850

FUENTE: Departamento de Antioquia, Departamento Administrativo de Planeación. *Anuario estadístico de Antioquia 1980.* Cálculos nuestros.

Observando el Cuadro No. 1 vemos que el modelo corregido predice mejor en la mayoría de los presentados (exceptuando dos) el valor promedio Y, y que el modelo sin corregir subestima constantemente el valor medio de Y.

4. Otros modelos

Potencias positivas fraccionarias

Si la transformación $Y^{1/n}$ es la apropiada, el modelo retransformado observado es de la forma

$$Y = (\beta_O + \beta_1 X + \mu)^n,$$ para cualquier n entero positivo y valores positivos de Y.

Si $\mu \sim n(o, \sigma^2)$, entonces $E(Y) = E(\beta o + \beta_1 X + \mu)^n$ el cual es el n-ésimo momento de una variable normal con media $\beta o + \beta_1 X$ y varianza σ^2. Los momentos de la distribución normal son bien conocidos (véase Kendall y Stuart 1969. Vol. 1) y proporcionan entonces el valor esperado de Y, dado X, para cualquier valor de n.

Por ejemplo, si $n = 2$ (es decir, utilizamos la transformación raíz cuadrada sobre Y), obtenemos que

$E(Y)$ = Segundo momento de una distribución normal con media $\beta o + \beta_1 X$ y varianza σ^2.

$$E(Y) = (media)^2 + varianza = (\beta o + \beta_1 X)^2 + \sigma^2 \qquad (10)$$

Por tanto el estimador con menos sesgos se obtiene sustituyendo $\hat{\beta} o, \hat{\beta}_1$ y $\hat{\sigma}^2$ en (10)

$$E(\hat{Y}) = (\hat{\beta} o + \hat{\beta}_1 X)^2 + \hat{\sigma}^2$$

La inversa

Cuando empleamos la transformación inversa $1/Y$, el modelo es

$$1/Y = \beta o + \beta_1 X + \mu$$

para valores positivos de Y.

Se puede probar que $E(Y)$ puede aproximarse por la cantidad

$$E(Y) = \frac{1}{\beta o + \beta_1 X} \left(1 + \frac{\sigma^2}{(\beta o + \beta_1 X)^2}\right) \text{ (véase Miller 1984)}$$

Un estimador de sesgos bajos se obtiene reemplazando los estimadores mínimo-cuadráticos $\hat{\beta} o, \hat{\beta}_1$ y $\hat{\sigma}^2$ en la expresión anterior.

En este caso el factor de ajuste es $\left(1 + \frac{\hat{\sigma}^2}{(\hat{\beta} o + \hat{\beta}_1 X)^2}\right)$

Potencias fraccionarias inversas

Si la transformación $Y^{-1/n}$. es la adecuada, la media de Y puede ser obtenida combinando los métodos usados para la transformación inversa las potencias fraccionarias. Por ejemplo para la raíz cuadrada inversa obtenemos.

$$E(\hat{Y}) = \frac{1}{(\hat{\beta}_o + \hat{\beta}_1 X)^2 + \hat{\sigma}^2} \left(1 + \frac{2\hat{\sigma}^2 + 4(\hat{\beta}_o + \hat{\beta}_1 X)^2 \hat{\sigma}^2}{\{(\hat{\beta}_o + \hat{\beta}_1 X)^2 + \hat{\sigma}^2\}^2}\right)$$

5. Extensión a algunos modelos no lineales con más de dos variables

Si la relación entre Y y las variables X_1, $X_2 . . .$, X_k es de alguna de las siguientes formas no-lineales y las variables explicativas X no son estocásticas los procedimientos anteriores para corregir sesgos son fácilmente extendidos así,

Si $Y = Ae^{\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta k\, Xk}\ e\mu$

entonces $E(\hat{Y}) = \hat{A}e^{\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta} k X k + 1/2\, \hat{\sigma}^2}$

donde $\hat{A} = e^{\hat{\beta}_o}$ y $\hat{\beta}_1, \hat{\beta}_2 \ldots, \hat{\beta}_k, \hat{\sigma}^2$, son los estimadores por mínimos cuadrados de $\beta_o, \beta_1, \ldots, \beta_k$ y σ^2 en el modelo

$$ln\ Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \mu$$

Para el modelo de elasticidad constante $E(Y) = \hat{A} X_1^{\hat{\beta}_1} X_2^{\hat{\beta}_2} \ldots X_k^{\beta k} e^{1/2\, \hat{\sigma}^2}$

Si $Y = (\beta_o + \beta_1 X_1 + \ldots \beta_k X_k + \mu)^2$

entonces

$$E(Y) = (\hat{\beta}_o + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k)^2 + \hat{\sigma}^2$$

Si $Y = (\beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)^{-1}$

entonces

$$E(\hat{Y}) = \frac{1}{\hat{\beta}_O + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k} \left(1 + \frac{\hat{\sigma}^2}{(\hat{\beta}_O + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k)^2}\right)$$

Conclusiones

Cuando utilizamos variables transformadas pueden ocurrir sesgos importantes en la estimación del valor medio de Y. Para los casos expuestos estos sesgos pueden ser reducidos, en proporción, utilizando los sencillos procedimientos dados.

Como observación final, para el buen funcionamiento de estos factores de corrección de sesgos es necesario que la hipótesis de normalidad en el término de perturbación del modelo sea satisfecha con buena aproximación; de otro modo estas correcciones no actuarán adecuadamente. Por tanto el predictor para $E(Y)$ no es resistente a alejamientos severos de la normalidad de μ.

BIBLIOGRAFIA

Box, G.E.P. y Cox D.R. (1964). "An Analysis of transformations". *Journal of the Royal Statistical Society.* Serie B, 26.

Departamento de Antioquia, Departamento Administrativo de Planeación. *Anuario Estadístico de Antioquia 1980.*

Draper, N.R. y Smith H. (1981). *Applied Regression Analysis.* New York, John Wiley and Sons.

Miller Don M. (1983)."Reducing Transformations Bias in Curve Fitting". Instituts of Statistics Working Paper, Virginia Commonwealth University.

____________ (1984). "Reducing Transformation Bias in Curve Fitting". *The American Statistician.* Vol. 38.

Mood, A. Graybill, F. y Boes, D. (1974). *Introduction to the Theory of Statistics.* Tokyo, Mac-Graw Hill - Kogakusha.

Mosteller, F. y Tukey, J. (1977). *Data Analysis and Regression.* Reading, Addison-Weseey.

Kendall, M. y Stuart, A. (1969). *The Advanced Theory of Statistics.* Londres, Charles Griffin.

Theil, H. (1978). *Introduction to Econometrics.* New Jersey, Prentice - Hall.