UNIVERSIDAD NACIONAL DE SALTA
Facultad de Ciencias Económicas,
Jurídicas y Sociales
Instituto de Investigaciones Económicas

Reunión de Discusión Nº 144

Fecha: 26/ 7/2000

Hs.: 16

La Transformación de Box - Cox: Una Nota

Eusebio Cleto del Rey

1. **Introducción**

Los propósitos de este trabajo son los siguientes: 1) Que el autor ordene sus propias ideas sobre el tema, tratando de escribir sobre él; 2) Que quede lo más claro posible el uso que se puede dar a esta transformación en el estudio de la contribución de mejoras; 3) Que los participantes tomen contacto con esta metodología, que les puede ser útil en sus propias investigaciones; 4) Que el autor reciba algunas ideas interesantes, lo cual es el propósito general de estas Reuniones de Discusión.

Por lo tanto, no pretendemos presentar nada original respecto a la transformación de Box - Cox ni a sus aplicaciones, salvo, quizás, respecto a su empleo en el cálculo de las contribuciones de mejoras. A ello se debe que su subtítulo sea: "Una Nota".

La transformación a la que nos referimos se origina en BOX y COX (1964). Hemos encontrado referencias a esa transformación en POWELL (1994) y en GREENE (1999), y aplicaciones a los precios hedónicos en mercados inmobiliarios en AGUIRRE y DE FARIA (1996), AGUIRRE (1998) y AGUIRRE y MACEDO (1999). Los tres últimos artículos son de especial interés en relación al tema de contribución de mejoras, que planteamos más adelante.

La transformación de Box - Cox brinda la posibilidad de correr una regresión entre variables, sin especificar de antemano la forma de la respectiva función, dejando que ella sea estimada, junto a los parámetros, en base a los datos disponibles.

En la Sec. 2 estudiamos la transformación que nos ocupa. La Sec. 3 está dedicada a su aplicación a nuestro problema de contribución de mejoras. La Sec. 4 trata de la estimación y la Sec. 5 contiene las conclusiones.

2. Qué es la Transformación de Box - Cox

Si tenemos una variable Y, cuyos valores son siempre positivos[1], su transformación de Box - Cox está dada por las siguientes ecuaciones:

$$Y^{(\lambda)} = \frac{Y^{\lambda} - 1}{\lambda} \qquad \text{para } \lambda \neq 0 \qquad (1)$$

$$Y^{(\lambda)} = \ln Y \qquad \text{para } \lambda = 0 \qquad (2)$$

Donde: λ es un parámetro, que generalmente toma valores entre - 2 y 2 (GREENE, 1999, PÁG. 417); $Y^{(\lambda)}$ es la variable Y, transformada empleando λ, según la ecuación (1).

[1] AGUIRRE (1998), pág. 6 dice, al presentar estas ecuaciones: "Para valores não negativos da variável...". GREENE (1999, pág. 416, nota al pie 17) dice asl: "Para que esté definida para todos los valores de λ, x tiene que ser estrictamente positiva" (x es nuestra Y). Esto se debe a "que $0^{(\lambda)}$ sólo está definido si λ es estrictamente positiva." (GREENE, 1999, pág. 416). Nótese que λ puede ser cero, o tener cualquier signo.

La ecuación (2) no es sino un caso particular de la ecuación (1), ya que a ella se llega tomando límite a esta última, para $\lambda \to 0$, y aplicando la regla de L'Hôpital. Por otra parte, cuando $\lambda = 1$, la ecuación (1) toma la forma:

$$Y^{(\lambda)} = Y - 1$$

o sea que la variable no se transforma (salvo porque se le resta la unidad). Si $\lambda = \frac{1}{2}$ resulta:

$$Y^{(\lambda)} = 2\sqrt{Y} - 2$$

que equivale a la transformación raíz cuadrada. Si $\lambda = -1$ tenemos:

$$Y^{(\lambda)} = 1 - \frac{1}{Y}$$

que es, aproximadamente, la transformación inversa.

La transformación de Box - Cox puede ser aplicada a la variable dependiente o a las variables independientes de una regresión. La especificación más general es la siguiente:

$$Y^{(\lambda_0)} = \beta_0 + \beta_1\ X_1^{(\lambda_1)} + \beta_2\ X_2^{(\lambda_2)} + \ldots + \beta_k\ X_k^{(\lambda_k)} + \varepsilon \qquad (3)$$

Donde: Y es la variable dependiente; X_j son variables independientes; β_0 es la coordenada al origen, a estimar; β_j son los parámetros a estimar; λ_0 y los λ_j son las constantes definidas en la ecuación (1); ε es el término de error; y el subíndice toma los valores: $j = 1, 2, \ldots, k$.

Según GREENE (1999, pág. 416), una especificación tan general como la ecuación (3) sería muy incómoda, debido a lo complicado de los cálculos necesarios para la estimación. Ello lleva a proponer la siguiente forma:

$$Y^{(\lambda_0)} = \beta_0 + \beta_1\ X_1^{(\lambda_1)} + \beta_2\ X_2^{(\lambda_1)} + \ldots + \beta_k\ X_k^{(\lambda_1)} + \varepsilon \qquad (4)$$

En la que se impone el mismo valor de λ para todas las variables independientes, pero se permite que éste difiera del de la variable dependiente.

Pero que λ_0 no sea igual a λ_1 (en ecuación (4)) es, todavía "una complicación mayor que la necesaria" (GREENE, 1999, pág. 419), por lo cual se puede especificar:

$$Y^{(\lambda)} = \beta_0 + \beta_1 \; X_1^{(\lambda)} + \beta_2 \; X_2^{(\lambda)} + \ldots + \beta_k \; X_k^{(\lambda)} + \varepsilon \tag{5}$$

Por último, en muchos casos puede ser suficiente la siguiente forma:

$$Y^{(\lambda)} = \beta_0 + \beta_1 \; X_1 + \beta_2 \; X_2 + \ldots + \beta_k \; X_k + \varepsilon \tag{6}$$

Las ecuaciones (3), (4), (5) y (6) fueron tomadas de AGUIRRE (1998, pág. 7), quien las atribuye a Spitzer.

Es necesario advertir que las variables dummies no pueden ser sometidas a la transformación de Box - Cox, debido a que pueden tomar el valor cero. Esto se ve confirmado por la siguiente afirmación de GREENE (1999, pág. 416, nota al pie 18): "En la mayoría de las aplicaciones, algunos regresores (por ejemplo, variables ficticias) no serán transformados."

3. Aplicación al Cálculo de Contribuciones de Mejoras

En DEL REY (1999a) y DEL REY (1999b) propusimos (mediante la ecuación (13) de esos trabajos), siguiendo a MELONI y RUIZ NÚÑEZ (1998), estimar una función del tipo:

$$P = p(A_1, A_2, \ldots A_k) + u \tag{7}$$

Donde: P es el precio del terreno, por metro cuadrado; A_j, para $j = 1, 2, \ldots k$, son los atributos del terreno, en alguna forma medidos; u es un componente al azar, sujeto a los supuestos acostumbrados de normalidad, media nula y varianza constante.

A continuación dijimos: "La función (13) puede tomar cualquiera de las formas acostumbradas para su estimación. Para fijar ideas, proponemos a continuación dos de ellas." Y trabajamos con una función lineal y otra semilogarítmica.

La transformación de Box - Cox nos permite plantear de un modo general la ecuación (7), y dejar que su forma sea determinada por los datos. Debido a que la mayoría de las A_j son variables dicotómicas, parece adecuado especificar (7) como la ecuación (6), mutatis mutandis, o sea del siguiente modo:

$$P^{(\lambda)} = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + ... + \beta_k A_k + u \qquad (8)$$

Si la estimación de λ resulta consistente con la unidad debemos emplear una función lineal, en tanto que si es consistente con cero la especificación debe ser semilogarítmica. Puede ser que no se acepte ninguna de esas hipótesis, y debamos usar una función de cualquier otra forma.

Lo dicho en el párrafo anterior condiciona el modo de computar la contribución de mejoras correspondiente a un lote de terreno. Veamos en los párrafos siguentes como se realiza ese cómputo, en los tres casos planteados. Para fijar ideas, suponemos que el atributo para el que $j = 1$ es pavimento, esto es, si la calle que pasa frente al terreno está o no pavimentada. En tal caso, la correspondiente variable estará definida como:

$A_1 = 1$ si la calle está pavimentada

$A_1 = 0$ si la calle no está pavimentada.

A) En el caso en que se acepte una de las hipótesis presentadas en DEL REY (1999a) y DEL REY (1999b), tenemos:

1) Aceptamos la hipótesis de que (8) es simplemente lineal ($\lambda = 1$), tal que:

$$P = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + ... + \beta_k A_k + u \qquad (9)$$

Entonces $\hat{\beta}_1$ (la estimación del coeficiente de A_1) es el monto a pagar por contribución de mejoras por cada metro cuadrado de terreno. Para saber cuanto debe pagar (como máximo) determinado terreno multiplicamos $\hat{\beta}_1$ por la superficie del mismo (S).

2) Aceptamos la hipótesis de que la función (8), tiene la forma que emplean MELONI y RUIZ NÚÑEZ (1998), o sea la semilogarítmica ($\lambda = 0$), entonces:

$$\ln P = \beta_0 + \beta_1 \, A_1 + \beta_2 \, A_2 + \ldots + \beta_k \, A_k + u \qquad (10)$$

Donde: $\ln$ simboliza logaritmos neperianos.

Entonces calculamos la contribución de mejoras como vemos a continuación: Para determinado lote de terreno, cuya calle se planea pavimentar, será cierto que:

$P_1 = P' \, e^{\hat{\beta}_1}$ cuando la calle esté pavimentada, y

$P_2 = P'$ en tanto no lo esté.

Donde:

$$P' = e^{\hat{\beta}_0} \, e^{\hat{\beta}_2 A_2'} \ldots e^{\hat{\beta}_k A_k'}$$

Siendo: A_2', ..., A_k' los valores de las otras variables independientes, correspondientes al lote para el que se quiere calcular la contribución de mejoras; $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_k$ las estimaciones de los respectivos parámetros.

Entonces:

$$\Delta P = P_1 - P_2 = P' \, (e^{\hat{\beta}_1} - 1) \qquad (11)$$

Es necesario puntualizar aquí lo siguiente: a) ΔP es el incremento del precio del lote considerado, por metro cuadrado, debido a la existencia de pavimento en su calle, y es, una vez multiplicado por S, lo que se le puede cobrar (como máximo) al dueño de este lote, por una sola vez, como contribución de mejoras. b) $(e^{\hat{\beta}_1} - 1)$ es el tanto por uno de incremento en el precio por metro cuadrado del terreno considerado, debido al pavimento. Así, por ejemplo, en MELONI y RUIZ NÚÑEZ (1998) encontramos que el coeficiente estimado es (con nuestra simbología) $\hat{\beta}_1 = 0{,}48$, de donde resulta que $(e^{\hat{\beta}_1} - 1) \approx 0{,}61$, o sea un 61 % de aumento en el precio debido a la existencia de pavimento. c) Por lo tanto, habiendo estimado la función (10) y conocidas las características A_2', ..., A_k' del lote para el que se debe calcular la contribución, computamos P' y le aplicamos el porcentaje de b). Multiplicamos el resultado por S, y obtenemos el total a pagar (como máximo) en concepto de contribución de mejoras por pavimento. d) Puesto que, al estimar la ecuación (10) con los supuestos de que u se distribuye normalmente, con media nula y varianza constante, trabajamos con un modelo lognormal, debemos corregir nuestro cálculo de la contribución por el sesgo pertinente (BRADU y MUNDLAK, 1970; DEL REY, 1983).

B) En el caso en el que se rechacen ambas hipótesis planteadas en a), la función (8) tendrá la forma general:

$$P^{(\lambda)} = \frac{P^{\lambda} - 1}{\lambda} = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \ldots + \beta_k A_k + u \qquad (12)$$

Entonces calculamos la contribución de mejoras para determinado lote de terreno, cuya calle se planea pavimentar, del sigiente modo:

$$P_1 = \left[\hat{\lambda}\left(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 X_2' + \ldots + \hat{\beta}_k X_k'\right) + 1\right]^{\frac{1}{\hat{\lambda}}}$$ cuando la calle esté pavimentada, y

$$P_2 = \left[\hat{\lambda}\left(\hat{\beta}_0 + \hat{\beta}_2 X_2' + \ldots + \hat{\beta}_k X_k'\right) + 1\right]^{\frac{1}{\hat{\lambda}}}$$ en tanto no lo esté.

Donde: $\hat{\lambda}$ es el valor estimado de λ, y los otros símbolos tienen el mismo significado que en el caso anterior.

Entonces:

$$\Delta P = P_1 - P_2 \qquad (13)$$

ΔP es el incremento del precio del lote considerado, por metro cuadrado, debido a la existencia de pavimento en su calle, y es, una vez multiplicado por S, lo que se le puede cobrar (como máximo) al dueño de este lote, por una sola vez, como contribución de mejoras.

Nótese que la transformación que nos permite pasar de $P^{(\lambda)}$ estimado a P_1 y P_2 no es lineal. El valor esperado de la transformación lineal de una variable (o estimador) es el valor esperado de esa variable, sometido a la misma tranformación lineal. Pero en este caso no podemos decir cuál es el valor esperado de las transformaciones, debido a que ellas son no lineales. En conclusión, podemos esperar que P_1 y P_2 sean sesgados, pero no conocemos su sesgo, para corregirlo (lo que es posible hacer en el caso semilogarítmico). Es curioso que GREENE (1999, pág. 422, ecuación (10-47)) presenta una transformación similar, pero no habla del posible sesgo.

4. Estimación

Si consideramos a λ de la ecuación (8) como un parámetro desconocido a estimar, esa ecuación no es lineal en los parámetros. Esto nos lleva a realizar nuestra estimación de (8) mediante el método de mínimos cuadrados no lineales. Si la variable al azar u de (8) tiene distribución normal, lo obtenido por mínimos cuadrados no lineales es, además, la estimación por máxima verosimilitud (GREENE, 1999, pág. 417).

Existen paquetes computacionales que realizan los cálculos necesarios para obtener lo enunciado en el párrafo anterior. Entre ellos se encuentra el LIMDEP[2], cuya subrutina BOXCOX realiza la transformación que aquí proponemos. Permite, además, hacer el test para las hipótesis nulas $\lambda = 1$ y $\lambda = 0$, o sea para las hipótesis de linealidad y ecuación semilogarítmica, respectivamente. Por supuesto que es también posible realizar el test para cualquier otra hipótesis referente a λ.

5. Conclusiones

A manera de conclusiones, describiremos el estado actual del Proyecto de Investigación sobre contribución de mejoras del que este trabajo forma parte. En esa descripción se podrá ver que lo presentado aquí es un avance notable del Proyecto.

1) La transformación de Box - Cox permite resolver el problema de la falta de una determinada forma para la función hedónica, pues ella no surge de la teoría. Esa transformación nos permite dejar que le forma surja de los datos.

---

[2] Este paquete econométrico está disponible en Salta.

2) El cálculo de la contribución que corresponde a cada propietario de un lote de terreno es matemáticamente diferente para cada una de las posibles formas de la función hedónica, por lo que lo tratado en este trabajo tiene gran importancia metodológica.

3) Siguiendo con la descripción del estado actual del Proyecto, diremos que solamente nos falta contar con una base de datos para Salta, que contenga los precios y las características de una muestra de terrenos. Ello nos permitiría trabajar sobre una aplicación piloto de nuestro método, lo que nos llevaría, a la vez, a un probable perfeccionamiento del mismo.

4) Fue elevado un pedido de financiación, para relevar los datos a los que se refiere 3), al Instituto "Horacio J. Fereyra", de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata, sin que se tenga aún noticias del resultado de esa gestión. Existen otras fuentes posibles de financiamiento, que serán exploradas en el caso de no tener éxito con la presentación antes mencionada.

Referencias:

AGUIRRE, Antonio (1998): "Uma Nota sobre a Transformação Box - Cox", Economia em Revista, Vol. 6, Nº 2, Dezembro, pág. 3 a 18.

AGUIRRE, Antonio y DE FARIA, Diomira M. C. P. (1996): "A Utilição dos "Preços Hedõnicos" na Avaliação Social de Projetos", Texto para Discussão Nº 103, Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR), Facultade de Ciências Econômicas, Universidade Federal de Minas Gerais, Junho.

AGUIRRE, Antonio y MACEDO, Paulo B. M. (1999): "Hedonic Price Estimation in the Housing Market of Belo Horizonte, Brazil", mimeo, Belo Horizonte, Agosto.

BOX, G. E. y COX, D. R. (1964): "An Analysis of Transformations", Journal of the Royal Statistical Society, Vol. 26, Nº 2, pág. 211 a 252. **Nota**: No pudimos disponer de este artículo, por lo que la referencia es tomada de los otros trabajos citados.

BRADU, Dan and MUNDLAK, Yair (1970): "Estimation in Lognormal Linear Models", Journal of the American Statistical Association, Vol. 65, Nº 329.

DEL REY, Eusebio Cleto (1983): "Problemas de Cómputo de la Corrección por Sesgo en el Caso Lognormal", Económica, Vol. 29, Nº 1,.pág. 27 a 43.

DEL REY, Eusebio Cleto (1999a): "La Contribución de Mejoras - Síntesis y Resumen", Asociación Argentina de Economía Política (A. A. E. P.): Anales: XXXIV Reunión Anual, Rosario (Santa Fe), pág. 242 a 249.

DEL REY, Eusebio Cleto (1999b): "La Contribución de Mejoras", A. A. E. P.: Anales: XXXIV Reunión Anual, Rosario (Santa Fe), en CD y en el Web site: http://www.aaep.org.ar.

GREENE, William H. (1999): Análisis Econométrico, Madrid, Prentice Hall, Tercera Edición.

MELONI, Osvaldo y RUIZ NÚÑEZ, Fernanda (1998): "Determinantes de los Precios de Mercado de los Terrenos en San Miguel de Tucumán", A. A. E. P.: Anales: XXXIII Reunión Anual, Mendoza, en CD y en el Web site: http://www.aaep.org.ar.

POWELL, James L. (1994): "Estimation of Semiparametric Models", en: ENGLE, R. J. y McFADDEN, D. L. (Ed.): Handbook of Econometrics, Amsterdam, Elsevier, Vol. IV, Cap. 41, pág. 2443 a 2521.

Universidad Nacional de Salta
Facultad de Ciencias Económicas,
Jurídicas y Sociales
Instituto de Investigaciones Económicas
Buenos Aires 177
4400 Salta
Argentina

REUNIONES DE DISCUSIÓN

Nº	Fecha	Autor	Título
135	18/11/99	Eduardo Antonelli	"Una Modelización de los Paradigmas Neoclásico y Keynesiano II"
136	29/ 3/00	Mauricio Ortín	"Origen del Prejuicio Anticapitalista en Marx"
137	12/ 4/00	Jorge A. Paz	"Cálculo del Ingreso Pleno"
138	3/ 5/00	Eduardo Antonelli y Cristina Egüez	"Un Ejercicio de Simulación de la Economía de Salta"
139	17/ 5/00	Lidia Rosa Elías de Dip	"Concentración y Patrones de Aglomeración en la Provincia de Salta"
140	24/ 5/00	Lidia Rosa Elías de Dip	"Proximidad Regional en la Provincia de Salta"
141	7/ 6/00	Eduardo Antonelli	"Una Nota sobre Aspectos Macroeconómicos del Desempleo"
142	21/ 6/00	Eduardo Antonelli	"Déficit Fiscal: Axiomas y Tabúes"
143	28 / 6/00	Vicente E. Rocha y Hugo H. Andías	"Fortalecimiento de las Finanzas Municipales. El Autoavalúo"
144	26/ 7/ 00	Eusebio Cleto del Rey	"La Transformación de Box - Cox: Una Nota "