Transformaciones

Econometría I

Departamento de Actuaría Física y Matemáticas
Universidad de las Américas Puebla

Dra. Daniela Cortés Toto

Índice

Índice

Transformaciones para linealizar el modelo

- Cuando el supuesto de linealidad entre la variable respuesta y las variables regresoras no se cumple, en algunos casos una función no lineal se puede linealizar con una transformación adecuada.
- A estos modelos no lineales se les llama intrínsecamente o transformablemente lineales.

Transformaciones intrínsecamente lineales

La siguiente tabla muestra algunas opciones que pueden ayudar a verificar el supuesto de linealidad.

Tabla 1. Funciones linealizables y su forma lineal correspondiente.

Función linealizable	Transformación	Forma lineal
$y = \beta_0 x^{\beta_1}$	$y' = log(y)$,$x' = log(x)$	$y' = log(\beta_0) + \beta_1 x'$
$y = \beta_0 e^{\beta_1 x}$	$y' = ln(y)$	$y' = log\left(\beta_0\right) + \beta_1 x$
$y = \beta_0 + \beta_1 log\left(x\right)$	$x' = log\left(x\right)$	$y' = y = \beta_0 + \beta_1 x'$
$y = \frac{x}{\beta_0 x - \beta_1}$	$y' = \frac{1}{y}$,$x' = \frac{1}{x}$	$y' = \beta_0 - \beta_1 x'$

Transformaciones intrínsecamente lineales

- Algunas otras transformaciones útiles son las recíprocas, por ejemplo, para el siguiente modelo

$$y = \beta_0 + \beta_1 \left(\frac{1}{x}\right) + \varepsilon$$

- se puede linealizar usando la transformación recíproca $x' = 1/x$, de esta forma el modelo linealizado quedaría

$$y = \beta_0 + \beta_1 x' + \varepsilon$$

Consideraciones

Cuando se emplean transformaciones como las anteriores, el estimador de mínimos cuadrados tiene propiedades de mínimos cuadrados con respecto a los datos transformados, y no a los datos originales.

Índice

Transformación de potencia y sus inconvenientes

- Una clase transformaciones mayormente utilizada para corregir el supuesto de normalidad y varianza constante es la transformación potencia
$$y^{\lambda}$$
- Donde λ es un parámetro que debe estimarse, por ejemplo, usando máxima verosimilitud de manera simultánea a la estimación de los parámetros del modelo.
- Sin embargo, esta transformación conlleva a una dificultad: $\lambda = 0$
- Si $\lambda \to 0$, entonces $y^{\lambda} \to 1$.

Transformación de potencia corregida

- Para resolver esta situación se propone usar $\left(y^{\lambda}-1\right)/\lambda$ como variable respuesta.
- Si $\lambda \to 0$, entonces $\left(y^{\lambda}-1\right)/\lambda \to ln(y)$
- Sin embargo, si λ cambia, los valores de $\left(y^{\lambda}-1\right)/\lambda$ cambian de forma dramática, dificultando la comparación de los estadísticos de resumen para modelos con distintos valores de λ.
- Por lo tanto, se propuso la siguiente transformación

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y}\, ln(y), & \lambda = 0 \end{cases} \tag{1}$$

Transformación de potencia corregida

- En la transformación descrita en (10) se tiene que $\dot{y} = ln^{-1}\left[(1/n)\sum_{i=1}^{n} ln\,(y_i)\right]$ es el promedio geométrico de las observaciones.
- Posteriormente se puede usar $y^{(\lambda)}$ como respuesta en el modelo de regresión.
- El divisor $\dot{y}^{\lambda-1}$se relaciona con el jacobiano de la transformación que convierte la variable de respuesta y en y^{λ}.
- Es de hecho un factor de escala que asegura que las sumas de cuadrados de residuales sean comparables para modelos con distintos valores de λ.

Índice

Estimación puntual del parámetro λ

- El estimador de λ por máxima verosimilitud es equivalente al valor de λ para el cual es mínima la suma de cuadrados de los residuales del modelo ajustado $SS_{Res}(\lambda)$.
- El valor de λ se estima ajustando un modelo a y^{λ} para distintos valores de λ, posteriormente se grafica la $SS_{Res}(\lambda)$ en función de λ y encontrando el valor que minimiza λla $SS_{Res}(\lambda)$.
- En general, son suficientes entre 10 a 15 valores de λ para encontrar el óptimo. Si se desea se puede hacer una segunda iteración con una malla más fina.

Estimación puntual del parámetro λ

- No se puede seleccionar λ sólo comparando en forma directa las sumas de cuadrados de residuales de las regresiones de y^{λ}respecto a x, porque para cada λ, la suma de cuadrados de residuales se mide en una escala distinta.
- Una vez seleccionado un valor de λ, el analista queda libre para ajustar el modelo usando a y^{λ} como variable de respuesta si $\lambda \neq 0$.
- Si $\lambda = 0$, se usa en y como variable de respuesta.
- Es admisible usar $y^{(\lambda)}$ como respuesta para el modelo final, este modelo tendrá una escala diferente y un origen trasladado en comparación del que usa y^{λ}.
- La mayor parte de los analistas prefieren usar y^{λ} o $ln(y)$ como respuesta.

Índice

Transformación de Box-Cox

A partir del análisis anterior se define la transformación de Box-Cox como a continuación

$$y^{\lambda} = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ ln(y), & \lambda = 0 \end{cases} \tag{2}$$

En donde el parámetro λ se estima como se describió anteriormente.

Índice

Definición del intervalo de confianza para λ

- Al aplicar el método de máxima verosimilitud al modelo de regresión para estimar λ,lo que se está maximizando es

$$L(\lambda) = -\frac{1}{2} n \ln\left[SS_{Res}(\lambda)\right] \tag{3}$$

- O equivalentemente, se está minimizando la función $SS_{Res}(\lambda)$.
- Un intervalo de confianza de $100(1-\alpha)\,\%$ para λ es el de todos aquellos valores que satisfacen la desigualdad

$$L\left(\hat{\lambda}\right) - L(\lambda) \leq \frac{1}{2}\chi^2_{\alpha,1}/n \tag{4}$$

- En donde, $\chi^2_{\alpha,1}$ es el punto porcentual superior de la distribución ji cuadrada con un grado de libertad.

Método gráfico para estimar el intervalo de confianza para λ

- Los valores del intervalo de confianza se pueden ver gráficamente sobre una línea que se traza sobre $L(\lambda)$ en función de λ de la recta

$$L\left(\hat{\lambda}\right) - \frac{1}{2}\chi^2_{\alpha,1} \tag{5}$$

en la escala vertical

- Esta línea corta a $L(\lambda)$en dos puntos, cuyos lugares en el eje de λ definen los dos extremos aproximados del intervalo.
- Dado que se está minimizando $SS_{Res}(\lambda)$ y graficando en función de λ, entonces la línea se debe graficar a la altura de

$$SS^* = SS_{Res}\left(\hat{\lambda}\right) e^{\chi^2_{\alpha,1}/n} \tag{6}$$

donde $\hat{\lambda}$ es el valor de λ que minimiza la suma de cuadrados de residuales.

References I

Montgomery (2006)
Introducción al análisis de regresión lineal. 2001, John Wiley & Sons.