# Econometría II

## CAPITULO 2: Modelos de regresión no lineales

Profesores: Graciela Sanroman - Mijail Yapor

Guía esquemática para las clases teóricas

Año: 2016

# Índice

1. Aproximándonos a los Modelos No Lineales (MNL)
2. Modelos lineales o linealizables
3. Modelos de Regresión No Lineal
4. Desarrollos de Taylor
5. Mínimos cuadrados no lineales (MCNL)
6. Estimador Método de los momentos
7. Estimador Máximo Verosímil (EMV)
8. Efectos parciales y transformación Box-Cox
9. Contrastes de hipótesis
10. Bibliografía

# Aproximándonos a los Modelos No Lineales (MNL)

Consideremos las dos funciones que figuran en el ejercicio 1 del práctico de este tema:

$$Y_i = B\left[\alpha N_i^r + (1-\alpha)K_i^r\right]^{\frac{1}{r}} \exp^{\mu_i} \qquad CES$$
$$Y_i = A N_i^{\alpha} K_i^{(1-\alpha)} \exp^{\mu_i} \qquad Cobb-Douglas$$

- Los parámetros B y A representan la Productividad Total de Factores (PTF) en cada uno de los modelos.
- En la función de producción Cobb-Douglas, $\alpha$ representa la elasticidad del producto ante cambios en el factor trabajo.
- En la función CES, el parámetro r es parte de la elasticidad de sustitución entre los factores productivos ($\sigma = \frac{1}{1-r}$), y cuando tiende a 0 la función CES es equivalente a la Cobb-Douglas.
- $\mu_i$ representa la heterogeneidad entre empresas que es inobservable. Por ejemplo, las empresas podrían trabajar con distintos niveles de capital humano, capacidades de gestión, etc., y ello las hace heterogéneas entre sí. La heterogeneidad que sí podemos observar es la causada por los regresores incluídos en los modelos (N y K aquí).

# Modelos lineales o linealizables

- Aplicando logaritmos al modelo Cobb-Douglas se tiene:

$$ln(Y_i) = ln(A) + \alpha ln(N_i)(1 - \alpha)ln(K_i) + \mu_i$$

  Que resulta ser una ecuación lineal en los parámetros $ln(A)$ y $\alpha$. Se dice que la función Cobb-Douglas es linealizable, es decir, existe una transformación de la función que es lineal en los parámetros. Notar que esto ocurre sólo si el error entra en la ecuación en forma multiplicativa.
- Comentario general: Aún en el caso en que los regresores estén constituidos por transformaciones no lineales de otras variables, si el **modelo es lineal en los parámetros** se continúa en el marco de los modelos de regresión lineal.
- La función CES no es linealizable, pues aún aplicando la transformación logarítmica (u otras transformaciones), no es posible obtener un modelo lineal en los parámetros.

# Estimación de los modelos de regresión no lineal

Utilizando la base de datos CalonMa.dta y el archivo Taller 3 2016 MNL.do disponibles en el EVA, se estiman ambos modelos.

Cuadro: Modelo Cobb-Douglas

| **Variable** | **Coefficient** | **(Std. Err.)** |
|---|---|---|
| n | 0.642** | (0.010) |
| k | 0.308** | (0.007) |
| Intercept | 1.890** | (0.029) |

Cuadro: Modelo CES

| **Variable** | **Coefficient** | **(Std. Err.)** |
|---|---|---|
| beta0 | 1.734** | (0.014) |
| r | 0.498** | (0.049) |
| a | 0.842** | (0.013) |

Forma general del modelo de regresión no lineal con errores aditivos

$$y_i = f(x_i, \beta) + u_i$$

Notar: el término de error entra de forma aditiva en la ecuación.

El modelo incluye muchas posibilidades

$$y = \beta_0 + \beta_1 \exp(\beta_2 x) + u$$

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2) + u \quad \text{Regresión de Poisson}$$

$$y = A\left[\alpha K^{\gamma} + (1-\alpha) L^{\gamma}\right]^{\frac{1}{\gamma}} + u \quad \text{Funcion producción CES}$$

**Supuestos:**

Sea

$$y_i = f(x_i, \beta) + u_i \quad i = 1, 2...N$$

**Supuesto 1 (independencia en media)**: $E\left(u_i \mid f(x_i, \beta)\right) = 0$. Notar que si se cumple este supuesto tenemos un modelo para la esperanza condicional $E(y_i \mid x_i) = f(x_i, \beta)$

**Supuesto 2 (identificación)**: No existe un vector no nulo $\beta^0 \neq \beta$ tal que $f(x_i, \beta^0) = f(x_i, \beta)$ para todo $x_i$

**Supuesto 3 (homocedasticidad y no autocorrelación):**
3.1 $E\left[u_i^2 \mid f(x_j, \beta) \quad j = 1, 2...N)\right] = \sigma^2 \quad \forall i$
3.2 $E\left[u_i u_j \mid f(x_i, \beta), f(x_j, \beta) \quad i, j = 1, 2...N\right] = 0 \quad \forall i \neq j$

# Modelos de regresión no lineal: aspectos teóricos

- Lo que caracterizará a un modelo de regresión no lineal es el método utilizado para estimar sus parámetros.
- Comparado con el modelo lineal surgen dos complicaciones:
  - no existe una solución análitica cerrada para los estimadores, por lo cual es necesario recurrir a métodos de cálculo númerico.
  - el efecto parcial de una variable deja de ser el coeficiente asociado a dicha variable.
- Veremos varias formas para estimar modelos de regresión no lineales con errores aditivos:
  - Regresión linealizada a través de desarrollos de Taylor.
  - Mínimos cuadrados no lineales.
  - Método de los momentos.
  - Máxima verosimilitud.

# La regresión linealizada: Desarrollos de Taylor

Desarrollo de Taylor en un entorno de $a$:

$$
\begin{aligned}
g(x) &= g(a) + g'(a)(x-a) + \frac{g''(a)}{2}(x-a)^2 + \ldots\ldots \\
&= \sum_{s=0}^{\infty} \frac{g_s(a)}{s!}(x-a)^s
\end{aligned}
$$

Para una función de dos variables, $x_1, x_2$, en un entorno $(a, b)$ el desarrollo de orden 1 es:

$$
\begin{aligned}
g(x_1, x_2) = {} & g(a, b) + g_1(a, b)\,(x_1 - a) + g_2(a, b)\,(x_2 - b) \\
& + O(.)
\end{aligned}
$$

# La regresión linealizada: Desarrollos de Taylor

Aquí estamos interesados en escribir la forma general no lineal como una función lineal en el vector $\beta$, para lo cual utilizaremos una aproximación de Taylor de orden 1 en torno a un valor $\beta^0$

Para el modelo

$$y_i = f(x_i, \beta) + u_i$$

Se tiene que:

$$f(x_i, \beta) = f(x_i, \beta^0) + \sum_{j=0}^{k} \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_j}(\beta_j - \beta_j^0) + O(p)$$

Agrupando términos:

$$f(x_i, \beta) = \left[f(x_i, \beta^0) - \sum_{j=0}^{k} \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_j}\beta_j^0\right] + \sum_{j=0}^{k} \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_j}\beta_j + O(p)$$

Notar: el primer término es una función de los datos y de los valores $\beta^0$, no de los parámetros $\beta$.

Entonces podemos escribir:

$$f(x_i, \beta) = f^0 - x^{0\prime}\beta^0 + x^{0\prime}\beta + O(p)$$

donde:

$$\beta^0 \text{ es un valor dado} f^0 = f(x_i, \beta^0) \; x^0 = \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta}$$

Podemos ahora obtener:

$$y^0 = y - f^0 + x^{0\prime}\beta^0$$

y definir el siguiente modelo de regresión lineal

$$y^0 = x^{0\prime}\beta + u^0$$

Así, para un valor dado $\beta^0$ es posible obtener $y^0$ y $x^0$ y obtener una estimación $\beta$ a través de una regresión lineal por MCO de $y^0$ sobre $x^0$.

**Ejemplo:**

$$f\left(x_i,\beta\right)=\beta_0+\beta_1\exp(\beta_2 x_i) \tag{1}$$

$$\frac{\partial f\left(x_i,\beta\right)}{\partial\beta_0}=1$$

$$\frac{\partial f\left(x_i,\beta\right)}{\partial\beta_1}=\exp(\beta_2 x_i)$$

$$\frac{\partial f\left(x_i,\beta\right)}{\partial\beta_2}=\beta_1 x_i\exp(\beta_2 x_i)$$

**Ejemplo (cont.)**

Entonces para un valor dado del vector $\beta^0$ es posible calcular

$$x_{0i}^0 = \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_0} = 1$$

$$x_{1i}^0 = \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_1} = \exp(\beta_2^0 x_i)$$

$$x_{2i}^0 = \frac{\partial f\left(x_i, \beta^0\right)}{\partial \beta_2} = \beta_1^0 x_i \exp(\beta_2^0 x_i)$$

Notación:

- $\beta^0$ refiere al vector de valores $\beta_0^0, \beta_1^0$ y $\beta_2^0$,
- $\beta_j$ con $j = 0, 1, 2$ son los parámetros:

**Ejemplo (cont.)**

$$y_i^0 = y_i - f(x_i, \beta^0) + \beta_0^0 x_{0i}^0 + \beta_1^0 x_{1i}^0 + \beta_2^0 x_{2i}^0$$

donde

$$\begin{aligned}
f(x_i, \beta^0) &= \beta_0^0 + \beta_1^0 \exp(\beta_2^0 x_i) \\
\beta_0^0 x_{0i}^0 &= \beta_0^0 \\
\beta_1^0 x_{1i}^0 &= \beta_1^0 \exp(\beta_2^0 x_i) \\
\beta_2^0 x_{1i}^0 &= \beta_2^0 \beta_1^0 x_i \exp(\beta_2^0 x_i)
\end{aligned}$$

Luego se regresa $y_i^0$ sobre $x_{0i}^0, x_{1i}^0$ y $x_{2i}^0$ para estimar $\beta_0, \beta_1$ y $\beta_2$. La precisión de la estimación dependerá de cuan buena fue la aproximación inicial $\beta^0$. Es posible comparar estos valores con los estimados e iterar hasta lograr un acercamiento de acuerdo a algún criterio prefijado. Este procedimiento se conoce con el nombre de Mínimos Cuadrados No Lineales.

# Mínimos cuadrados no lineales (MCNL)

Si se cumple el supuesto de independencia en media de los errores respecto a las X, podemos operar a través del método de los mínimos cuadrados. Los valores de los parámetros que minimizan la suma al cuadrado de los residuos será

$$SCR(\beta) = \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (y_i - f(x_i, \beta))^2$$

Las CPO son:

$$\frac{\partial SCR(\beta)}{\partial \beta} = -2 \sum_{i=1}^{N} (y_i - f(x_i, \beta)) \frac{\partial f(x_i, \beta)}{\partial \beta} = 0$$

Este sistema de ecuaciones no lineales no posee una solución explícita.

**Definición:** Un modelo de regresión no lineal es aquel cuyas condiciones de primer orden para la estimación por mínimos cuadrados de sus parámetros son funciones no lineales de esos parámetros.

En el caso del modelo para la CES (en logaritmos), se tiene que:

$$y_i = ln(Y_i) = ln(B) + \frac{1}{r}\left[\alpha N_i^r + (1-\alpha)K_i^r\right] + \mu_i$$

Por lo que:

$$f\left(x_i, \beta\right) = ln(B) + \frac{1}{r}\left[\alpha N_i^r + (1-\alpha)K_i^r\right]$$

y las CPO están dadas por:

$$-2\sum_{i=1}^{N}\left(y_i - ln(B) + \frac{1}{r}\left[\alpha N_i^r + (1-\alpha)K_i^r\right]\right)\frac{\partial f\left(x_i, \beta\right)}{\partial \beta} = 0$$

El gradiente o pseudo regresores $\frac{\partial f(x_i,\beta)}{\partial \beta}$ es el vector resultante de derivar la función CES respecto de sus tres parámetros: $ln(B)$, $\alpha$ y $r$.

# MCNL

Para estimar el modelo a través de MCNL podemos:

- iterar en $\beta^0$ el modelo de regresión linealizado, es decir, a partir de un $\beta^0$ inicial tomar en cada iteración al $\beta$ como nuevo $\beta^0$,
- estimar utilizando algún algoritmo de minimización de funciones no lineales (Gauss-Newton, Newton-Raphson, BHHH).

**Varianza del estimador MCNL**

$$Var\left(\widehat{\beta}_{MCNL}\right) = \sigma^2 \left[\sum_{i=1}^{N} \frac{\partial f\left(x_i, \beta\right)}{\partial \beta_j} \frac{\partial f\left(x_i, \beta\right)}{\partial \beta_j}'\right]^{-1}$$

Donde una estimación consistente del parámetro $\sigma^2$ se obtiene a través de

$$\widehat{\sigma}^2 = \frac{e'e}{N} \text{ con } e = y - f\left(x, \widehat{\beta}_{MCNL}\right)$$

Con las estimaciones puntuales de los parámetros del modelo no lineal y de su matriz de varianzas y covarianzas es posible realizar los contrastes de significación individual y conjunta de los parámetros.utilizando los estadísticos z y chi-cuadrado, ya que la solución es asintótica.

Las condiciones bajo las que los resultados anteriores son válidos incluyen la existencia de un único mínimo global de la función SCR($\beta$) y la no singularidad de la matriz límite.

**Propiedades asintóticas del estimador MCNL**

Al igual que en el estimador MCO del modelo de regresión lineal, la consistencia requiere $plim(\widehat{\beta} - \beta) = 0$.

$$
\begin{aligned}
\widehat{\beta}_{MCNL} - \beta &= \left(X^{0\prime}X^{0}\right)^{-1} X^{0\prime}u \\
&= \left(\frac{X^{0\prime}X^{0}}{n}\right)^{-1} \frac{X^{0\prime}u}{n} \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} x_i^0 x_i^{0\prime}\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} x_i^0 u_i\right]
\end{aligned}
$$

Por lo tanto, $\widehat{\beta}_{MCNL}$ será consistente si límite en probabilidad de este resultado es nulo.

# MCNL

Las condiciones que asegurarán las consistencia son:
1) $\text{plim}\left[\frac{1}{n}\sum_{i=1}^{n} x_i^0 u_i\right] = 0.$
2) plim $\frac{1}{n}\sum_{i=1}^{n} x_i^0 x_i^{0'}$ converge a una matriz finita e invertible.

Estimación de la varianza de los errores:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{N}\left(y_i - f\left(x_i, \widehat{\beta}\right)\right)^2$$

La corrección por grados de libertad (n-k) no está justificada en este caso, porque los resultados sólo tienen validez asintótica.

**Distribución asintótica del estimador MCNL**

Si los pseudo-regresores definidos anteriormente se comportan bien, entonces:

$$\widehat{\beta}_{MCNL} \xrightarrow{D} N\left[\beta, \frac{\sigma^2}{n}\left(Q^0\right)^{-1}\right]$$

donde,

$$Q^0 = p\lim\left(\frac{X^{0\prime}X^0}{n}\right)$$

La estimación de la matriz de varianzas y covarianzas del estimador MCNL estará dada entonces por:

$$\widehat{Avar}\left(\widehat{\beta}_{MCNL}\right) = \widehat{\sigma}^2\left(X^{0\prime}X^0\right)^{-1}$$

# Estimador Método de los momentos

En términos más generales, si se cumple el supuesto

$$E(u_i \,|f(x_i, \beta)\,) = 0 \Rightarrow$$

$$\begin{aligned}
&E(u_i) = E(y_i - f(x_i, \beta)) = 0 \\
&E(\tfrac{\partial f(x_i,\beta)}{\partial \beta} u_i) = 0 \Rightarrow \\
&\quad E(\tfrac{\partial f(x_i,\beta)}{\partial \beta_1} u_i) = E\left[\tfrac{\partial f(x_i,\beta)}{\partial \beta_1}(y_i - f(x_i, \beta)\right] = 0 \\
&\quad E(\tfrac{\partial f(x_i,\beta)}{\partial \beta_2} u_i) = E\left[\tfrac{\partial f(x_i,\beta)}{\partial \beta_2}(y_i - f(x_i, \beta)\right] = 0 \\
&\vdots \\
&\quad E(\tfrac{\partial f(x_i,\beta)}{\partial \beta_k} u_i) = E\left[\tfrac{\partial f(x_i,\beta)}{\partial \beta_k}(y_i - f(x_i, \beta)\right] = 0
\end{aligned}$$

# Estimador Método de los momentos

Si se cuenta con una muestra iid $\{y_i, x_i\}_{i=1,\dots N}$ es posible utilizar el método de los momentos (el principio de analogía) para estimar los coeficientes del modelo:

$$E(u_i) = 0 \overset{\text{ppio analogía}}{\longrightarrow} \frac{1}{N}\sum_{i=1}^{N}\left[y_i - f(x_i,\widehat{\beta})\right] = 0$$

$$E\left(\frac{\partial f(x_i,\beta)}{\partial \beta_1}u_i\right) = 0 \overset{\text{ppio analogía}}{\longrightarrow} \frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial f(x_i,\beta)}{\partial \beta_1}\right|_{\beta=\widehat{\beta}}\left[y_i - f(x_i,\widehat{\beta})\right] = 0$$

$$E\left(\frac{\partial f(x_i,\beta)}{\partial \beta_2}u_i\right) = 0 \overset{\text{ppio analogía}}{\longrightarrow} \frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial f(x_i,\beta)}{\partial \beta_2}\right|_{\beta=\widehat{\beta}}\left[y_i - f(x_i,\widehat{\beta})\right] = 0$$

$$\vdots$$

$$E\left(\frac{\partial f(x_i,\beta)}{\partial \beta_k}u_i\right) = 0 \overset{\text{ppio analogía}}{\longrightarrow} \frac{1}{N}\sum_{i=1}^{N}\left.\frac{\partial f(x_i,\beta)}{\partial \beta_k}\right|_{\beta=\widehat{\beta}}\left[y_i - f(x_i,\widehat{\beta})\right] = 0$$

# Estimador Método de los momentos

Notar:

- las condiciones anteriores son idénticas a las que dan lugar a los estimadores MCNL, por lo cual las propiedades de este estimador son iguales a las del MCNL
- es necesario resolver estas ecuaciones por métodos de cálculo númerico

# Estimador Máximo Verosímil (EMV)

Análogamente que en el MRL es necesario agregar un supuesto respecto a la distribución de probabilidad de las perturbaciones, siendo lo más usual suponer normalidad

$$u \mid x \sim N(0, \sigma^2 I)$$

Bajo los supuestos de muestras *iid*, normalidad, independencia en media, homocedasticidad y no autocorrelación, la función de verosimilitud está dada por:

$$\mathcal{L}(u, \sigma^2) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} u' u\right)$$

# EMV

La función de verosimilitud es:

$$\mathcal{L}(y, X \mid \beta, \sigma^2) \\ = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(y - f(X,\beta))'(y - f(X,\beta))\right)$$

Tomando logaritmos obtenemos la log-verosimilitud:

$$L(y, X \mid \beta, \sigma^2) = ln\mathcal{L}(y, X \mid \beta, \sigma^2) \\ = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - f(x_i,\beta))'(y_i - f(x_i,\beta))$$

Notar: Maximizar la verosimilitud, bajo los supuestos de errores aditivos y normalmente distribuidos, equivale a minimizar la suma al cuadrado de los residuos.

# EMV

Las condiciones de primer orden son:

$$\frac{\partial L}{\partial \beta} = -\frac{1}{2\sigma^2}\frac{\partial SCR(\beta)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i - f\left(x_i, \beta\right)\right)\frac{\partial f\left(x_i, \beta\right)}{\partial \beta} = 0$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}u'u = 0$$

- No hay una solución análitica cerrada para $\beta$, pero la solución coincide con la obtenida mediante MCNL
- $\widehat{\sigma}^2_{MV} = \frac{e'e}{n}$ con $e = \left(y_i - f\left(x_i, \widehat{\beta}\right)\right)$

**Varianza del estimador Máximo Verosímil**

Para estimar la matriz de covarianzas es necesario calcular las derivadas segundas:

$$
\begin{aligned}
\frac{\partial^2 L}{\partial\beta\partial\beta'} &= -\frac{1}{2\sigma^2}\frac{\partial^2 SCR(\beta)}{\partial\beta\partial\beta'} \\
\frac{\partial^2 L}{\partial\beta\partial\sigma^2} &= \frac{1}{2\sigma^4}\frac{\partial SCR(\beta)}{\partial\beta} \\
\frac{\partial^2 L}{\partial\sigma^2\partial\sigma^2} &= \frac{n}{2\sigma^4} - \frac{SCR(\beta)}{\sigma^6}
\end{aligned}
$$

$$E\left(\frac{\partial SCR(\beta)}{\partial\beta}\right) = -2\sum_{i=1}^{N} E\left[\left(y_i - f(x_i, \widehat{\beta})\right)\frac{\partial f\left(x_i, \beta\right)}{\partial\beta}\right]$$

$$= -2\sum_{i=1}^{N} E\left[e_i \frac{\partial f\left(x_i, \beta\right)}{\partial\beta}\right] = 0$$

por lo cual,

$$E\left(\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\right) = \frac{1}{2\sigma^4} E\left(\frac{\partial SCR(\beta)}{\partial\beta}\right) = 0$$

# EMV

La matriz de información es diagonal en bloques:

$$\frac{\partial^2 SCR(\beta)}{\partial\beta\partial\beta'} = -2\sum_{i=1}^{N}\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}' + 2\sum_{i=1}^{N}\left(y - f(x_i,\widehat{\beta})\right)\frac{\partial f\left(x_i,\beta\right)}{\partial\beta\partial\beta'}$$
$$= -2\sum_{i=1}^{N}\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}' + 2\sum_{i=1}^{N}u_i\frac{\partial f\left(x_i,\beta\right)}{\partial\beta\partial\beta'}$$

como, $E\left(\frac{\partial SCR(\beta)}{\partial\beta}\right) = 0$

$$E\left(-\frac{\partial^2 L}{\partial\beta\partial\beta'}\right) = \frac{1}{\sigma^2}E\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}\frac{\partial f\left(x_i,\beta\right)}{\partial\beta}'$$

La matriz de información es:

$$I(\theta) = \begin{pmatrix} \frac{1}{\sigma^2} \frac{\partial f(x_i,\beta)'}{\partial \beta} \frac{\partial f(x_i,\beta)}{\partial \beta} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}$$

Y por tanto la matriz de varianzas y covarianzas es:

$$I^{-1}(\theta) = \begin{pmatrix} \sigma^2 \left[ \frac{\partial f(x_i,\beta)'}{\partial \beta} \frac{\partial f(x_i,\beta)}{\partial \beta} \right]^{-1} & 0 \\ 0 & \frac{2\widehat{\sigma}^4}{N} \end{pmatrix}$$

# Efectos parciales en los Modelos de regresión no lineal

Partiendo del modelo $E(y \mid x) = f(x, \beta)$ obtenemos el efecto parcial asociado a la variable $x_j$

$$\frac{\partial E(y \mid x)}{\partial x_j} = \frac{\partial f(x, \beta)}{\partial x_j}$$

Notar:

- En general la estimación puntual del efecto parcial dependerá de $x$.
- Para calcular el efecto parcial asociado a una variable será en general necesario aplicar el teorema del Slutsky, que plantea que si $w_n \xrightarrow{p} c$ y $h(w_n)$ es continuo en $c$, entonces $h(w_n) \xrightarrow{p} h(c)$.
- Para calcular el error estándar del efecto parcial se utiliza el método delta:

$$ee\left[\frac{\partial E(y \mid x)}{\partial x_j}\right] = h(x, \beta)' Var\,(\beta)\, h(x, \beta)$$

$$\text{siendo } h(x, \beta) = \partial \frac{\partial E(y \mid x)}{\partial x_j} \Big/ \partial \beta$$

**Ejemplo:**
Considere nuevamente el modelo $E(y \mid x) = f(x, \beta) = \beta_0 + \beta_1 \exp(\beta_2 x)$

El efecto parcial asociado a la variable x es:

$$\frac{\partial E(y \mid x)}{\partial x} = \beta_1 \times \beta_2 \times \exp(\beta_2 x)$$

La estimación puntual del efecto está dada por:

$$\frac{\partial \widehat{E(y \mid x)}}{\partial x} = \widehat{\beta}_{1MCNL} \times \widehat{\beta}_{2MCNL} \times \exp(\widehat{\beta}_{2MCNL} x)$$

Es necesario dar valores a la $x$ para estimarlo, siendo las dos alternativas más utilizadas:

- estimar el efecto en $\overline{x}$ (el valor medio de $x$ en la muestra):
  $\frac{\partial \widehat{E(y|x)}}{\partial x} = \widehat{\beta}_{1MCNL} \times \widehat{\beta}_{2MCNL} \times \exp(\widehat{\beta}_{2MCNL} \overline{x})$
- estimar la media de los efectos parciales
  $\frac{\partial \widehat{E(y|x)}}{\partial x} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\beta}_{1MCNL} \times \widehat{\beta}_{2MCNL} \times \exp(\widehat{\beta}_{2MCNL} x_i)$

**Método Delta (en el ejemplo bajo análisis)** Tenemos que el efecto parcial asociado a la variable x es:

$$\frac{\partial E(y \mid x)}{\partial x} = \beta_1 \times \beta_2 \times \exp(\beta_2 x)$$

Entonces:

$$h(x,\beta) = \frac{\partial \frac{\partial E(y|x)}{\partial x}}{\partial \beta} = \begin{pmatrix} 0 \\ \beta_2 \exp(\beta_2 x) \\ \beta_1 \exp(\beta_2 x)\left[1 + \beta_2 x\right] \end{pmatrix}$$

$$Var(\beta) = \begin{pmatrix} Var\left(\beta_0\right) & cov\left(\beta_0, \beta_1\right) & cov\left(\beta_0, \beta_2\right) \\ & Var\left(\beta_1\right) & cov\left(\beta_1, \beta_2\right) \\ & & Var\left(\beta_2\right) \end{pmatrix}$$

Los errores estándar de los efectos parciales también deben evaluarse en valores especificos de las $x$

# Transformación de Box-Cox

Consideremos el siguiente modelo lineal:

$$y = \alpha + \beta g(x) + u$$

donde,

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

La transformación de Box-Cox, $g(x)$, se utiliza para generalizar los modelos lineales dándole diferentes valores al parámetro $\lambda$. Los casos especiales de interés son:

- Modelo lineal: $\lambda = 1$
- Modelos log-lineal o semi-logarítmico $\lambda = 0$
- Otros potenciales valores como $\lambda = -1$

# Transformación de Box-Cox

- Para un valor dado de $\lambda$, se utiliza dicho valor en la transformación y el modelo se convierte en lineal, estimándose mediante MCO.
- Sin embargo, si $\lambda$ es tomado como un parámetro desconocido el modelo se transforma en uno no lineal en los parámetros.
- Para evitar esta situación, una alternativa es estimar el modelo para distintos valores de $\lambda$ (que en general se espera esté entre -1 y 1) y así obtener estimaciones para $\alpha$ y $\beta$, eligiendo el modelo que obtenga la menor suma del cuadrado de los residuos.
  Sin embargo, como se efectuó un proceso de búsqueda sobre un rango de valores de $\lambda$, la matriz de covarianzas de los coeficientes $\alpha$ y $\beta$ que se obtiene en la regresión condicionada que se ha seleccionado como óptima no es válida y tenderá a subestimar las verdaderas varianzas.

# Transformación de Box-Cox

Como alternativa puede utilizarse la estimación por máxima verosimilitud:

$$lnL(y, X \mid \beta, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}u\prime u$$

Donde el jacobiano de la transformación es:

$$\left|\frac{du}{dy}\right| = 1$$

Sustituyendo el error del modelo considerando la transformación de Box-Cox:

$$lnL(y, X \mid \beta, \sigma^2) =$$
$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum\left(y_i - \alpha - \beta g(x_i)\right)^2$$

Las C.P.O son:

$$\frac{\partial lnL}{\partial \alpha} = \frac{u_i}{\sigma^2} = 0$$

$$\frac{\partial lnL}{\partial \beta} = \frac{u_i}{\sigma^2} g(x_i) = \frac{u_i}{\sigma^2} \frac{x_i^\lambda - 1}{\lambda} = 0$$

$$\frac{\partial lnL}{\partial \lambda} = \frac{u_i}{\sigma^2} \beta \frac{\partial g(x_i)}{\partial \lambda} = \frac{u_i}{\sigma^2} \beta \frac{\lambda x_i^\lambda \ln x_i - x_i^\lambda + 1}{\lambda^2} = 0$$

$$\frac{\partial lnL}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} u_i^2 = 0$$

Resolución:

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i-\widehat{\alpha}-\widehat{\beta}\frac{x_i^{\widehat{\lambda}}-1}{\widehat{\lambda}}\right)=0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i-\widehat{\alpha}-\widehat{\beta}\frac{x_i^{\widehat{\lambda}}-1}{\widehat{\lambda}}\right)\frac{x_i^{\lambda}-1}{\lambda}=0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i-\widehat{\alpha}-\widehat{\beta}\frac{x_i^{\widehat{\lambda}}-1}{\widehat{\lambda}}\right)\frac{\lambda x_i^{\lambda}\ln x_i-x_i^{\lambda}+1}{\lambda^2}=0$$

$$-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{N}\left(y_i-\widehat{\alpha}-\widehat{\beta}\frac{x_i^{\widehat{\lambda}}-1}{\widehat{\lambda}}\right)^2=0$$

Que no tiene soluciòn analítica, por lo puede tratarse como un modelo de regresión no lineal.

Veremos cómo evaluar la adecuación de un conjunto de hipótesis del tipo

$$H_0 : R(\theta) - r = 0$$

donde $R(\theta)$ es un vector columna de $q$ funciones continuas en el dominio de $\theta$.

El ejemplo más simple es $R(\theta)$ lineal: $R(\theta) = R\theta$ donde $R$ es una matriz $q \times k$, con $q \leq k$ y $r$ un vector $q \times 1$.

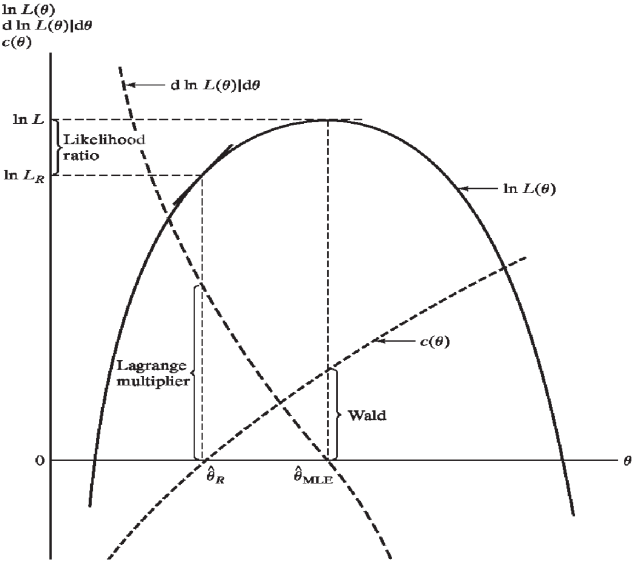
Pero aquí estamos interesados en funciones más generales. Una condición que debe cumplirse si $Ho$ contiene más de una restricción es que las restricciones deben ser funcionalmente independientes. Esto se cumple si la matriz $q \times k$ y $r(\theta) = \frac{\partial R(\theta)}{\partial \theta}$ tiene rango completo.

Veremos cuatro alternativas:

- Prueba F.
- Contraste de Wald(W).
- Contraste de multiplicadores de Lagrange (ML).
- Contraste de razón de verosimilitudes (RV).

Estas alternativas son asintóticamente válidas y asintóticamente equivalentes.

# Contrastes de hipótesis



**Prueba F**

Sean:

$$\widehat{\theta}_{NR} \text{ el estimador en el modelo no retringido}$$

$$\widehat{\theta}_{R} \text{ el estimador en el modelo retringido}$$

El estadístico análogo a la prueba F en el contexto de restricciones lineales está dado por:

$$F(q, n-k) = \frac{\left[SCR\left(\widehat{\theta}_R\right) - SCR\left(\widehat{\theta}_{NR}\right)\right]/q}{SCR\left(\widehat{\theta}_{NR}\right)/(n-k)}$$

Esta fórmula es muy similar a la vista en el contexto de modelos y contrastes lineales, sin embargo en el contexto de modelos de regresión no lineales ni el numerador ni el denominador se distribuyen exactamente chi-cuadrado por lo cual la distribución $F$ es sólo aproximada (es aconsejable utilizar estos contrastes sólo si contamos con una muestra grande).

**Contraste de Wald**

Este contraste requiere estimar exclusivamente el modelo no restringido. El vector $R(\widehat{\theta}_{NR}) - r$ indicará hasta qué punto los estimadores no restringidos ajustan a la hipótesis nula. El estadístico (y su distribución asintótica bajo la hipótesis nula) estará dado por :

$$W = \left(R(\widehat{\theta}_{NR}) - r\right)' \left[r(\widehat{\theta}_{NR})\widehat{V}r(\widehat{\theta}_{NR})'\right]^{-1} \left(R(\widehat{\theta}_{NR}) - r\right) \overset{a}{\sim} \chi^2_q$$

donde,

$$r(\theta) = \frac{\partial R(\theta)}{\partial \theta} \qquad y \qquad \widehat{V} = \widehat{aVar}(\widehat{\theta}_{NR})$$

Si las restricciones son válidas el estadístico $W$ será asintóticamente equivalente a $q$ veces el estadístico $F$.

**Contraste de Wald: puntualizaciones**

- El comportamiento del estadístico $W$ en muestras pequeñas puede ser muy errático.

- Debido a que el contraste no involucra la hipótesis alternativa, el estadístico de Wald no es invariante a como la hipótesis es formulada. En los casos en los cuales hay más de un camino para especificar $R(\theta) = r$, *Wald* podría llevar a conclusiones diferentes.

**Contraste de Multiplicadores de Lagrange (ML)**

El estadístico ML (y su distribución asintótica bajo la hipótesis nula) está dado por:

$$ML = \frac{e_*^{'} X_*^0 \left(X_*^{0\prime} X_*^0\right)^{-1} X_*^0 e_*}{e_*^{'} e_* / n} \overset{a}{\sim} \chi_J^2$$

con

$$e_i^* = y_i - f(x_i, \widehat{\theta}_R)$$

$$e_*^{'} = (e_1^*, e_2^* .....e_n^*)' \qquad X_*^0 = \begin{pmatrix} x_{10}^0 & x_{11}^0 & .... & x_{1k}^0 \\ x_{20}^0 & x_{21}^0 & .... & x_{2k}^0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n0}^0 & x_{n1}^0 & .... & x_{nk}^0 \end{pmatrix}$$

$$x_{ij}^0 = \frac{\partial f\left(x_i, \widehat{\theta}_R\right)}{\partial \beta j} \text{ con } i = 1, 2, ....n \quad j = 1, 2...k$$

Notar que este contraste es también n veces el $R^2$ no centrado de la regresión de $e_*$ sobre $X_*^0$.

**Contraste de la razón de verosimilitudes (RV)**
Para proceder por este camino es necesario conocer (o suponer conocida) la distribución de probabilidad de las perturbaciones.

Se procede a estimar el modelo:

- no restringido
- sujeto a restricciones del tipo $R(\theta) - r = 0$.

**Contraste de la razón de verosimilitudes (RV)**

$L(w \mid \widetilde{\theta})$ es la versosimilitud del modelo restringido, $\widetilde{\theta}$ es el vector con las estimaciones del modelo restringido.

$L(w \mid \widehat{\theta})$ es la versosimilitud del modelo no restringido. $\widehat{\theta}$ es el vector con las estimaciones del modelo no restringido.

Definimos el ratio de verosimilitudes como:

$$\lambda = \frac{L(w \mid \widetilde{\theta})}{L(w \mid \widehat{\theta})}$$

Donde esperaremos rechazar la hipótesis nula cuando $\lambda$ sea muy pequeña.

El estadístico del contraste (y su distribución asintótica bajo la hipótesis nula)

$$RV = -2 \ln \lambda = -2 \ln \left[ \frac{L(w \mid \widetilde{\theta})}{L(w \mid \widehat{\theta})} \right] \overset{a}{\sim} \varkappa_q^2$$

# Bibliografía:

Greene, W. - “Análisis Econométrico”, 3era. ed., Prentice Hall Iberia SRL, 1998. Capítulo 10