CURSO 2011/2012

Tema 5: Modelo clásico de regresión: extensiones

Aránzazu de Juan Fernández

ECONOMETRÍA I

ESQUEMA DEL TEMA 5

- ❑ **Motivación**
- ❑ **Funciones de regresión no lineales**
- ❑ **Interacciones entre variables**
- ❑ **Errores de especificación**
- ❑ **Multicolinealidad**
- ❑ **Heterocedasticidad**

BIBLIOGRAFÍA BÁSICA

[1] **Wooldridge, J.M. (2006).** *Introducción a la Econometría. Un enfoque moderno.* **Thomson. Cap. 6 - 9**

[2] **Stock, J.H. y Watson, M.W. (2008).** *Introduction to Econometrics* **(2 Ed.) Pearson International Edition, Cap. 6, 16**

[5] **Novales, A. (1993).** *Econometría*. **Cap. 5, 6 y 10**

1.- MOTIVACIÓN

Hasta ahora se han hecho supuestos ó hipótesis que nos han permitido encontrar estimaciones de los parámetros en situaciones ideales

Sin embargo, en numerosas ocasiones no se cumplen estas hipótesis de manera que es conveniente conocer en qué situaciones no se cumplen y qué propiedades tienen los estimadores en estos casos.

En este tema veremos el incumplimiento de ciertas hipótesis del Modelo Clásico de Regresión y analizaremos cuáles son las propiedades de los estimadores

2.- FUNCIONES DE REGRESIÓN NO LINEALES

- Todo lo anterior ha sido lineal en las *X*'s
- La aproximación de que la función de regresión sea lineal puede ser satisfactoria para algunas variables pero no para otras.
- El enfoque de regresión múltiple se puede extender para tener en cuenta funciones de regresión que no sean lineales en una o más de una *X*.

Si una relación entre *Y* y *X* es no lineal:

- El efecto sobre *Y* de un cambio en *X* depende del valor de *X* – es decir, el efecto marginal de *X* no es constante
- Una regresión lineal está mal especificada – la forma funcional no es la correcta
- El estimador del efecto en *Y* de *X* es sesgado – ni siquiera es correcto en media.
- La solución a este problema es estimar una función de regresión que no sea lineal en *X*

La función de regresión poblacional no lineal general

$$Y_i = f(X_{1i}, X_{2i}, \ldots, X_{ki}) + u_i,\ i = 1, \ldots, n$$

Hipótesis:

- **$E(u_i| X_{1i}, X_{2i}, \ldots, X_{ki}) = 0$ (igual que antes); implica que f es la esperanza condicionada de Y dadas las X's.**

- **$(X_{1i}, \ldots, X_{ki}, Y_i)$ son i.i.d. (igual que antes).**

- **existencia de "un número suficiente" de momentos (misma idea; dependiendo de cada f específica).**

- **No hay multicolinealidad perfecta (misma idea; dependiendo de cada f específica).**

Efecto esperado en Y_i de un cambio en X_1 en el modelo de Regresión No Lineal

El cambio esperado en Y, ΔY, asociado con el cambio en $X_1, \Delta X_1$, manteniendo constantes $X_2, X_3, \ldots, X_k$, es la diferencia entre el valor de la función de regresión poblacional antes y después de cambiar X_1, manteniendo $X_2, X_3, \ldots, X_k$ constantes:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \ldots, X_k) - f(X_1, X_2, \ldots, X_k)$$

El estimador de esta diferencia es la diferencia entre ambas predicciones. Sea $\hat{f}\left(X_1, X_2, \ldots, X_k\right)$ la predicción de Y basada en el estimador de la función de regresión poblacional. Entonces, la predicción del cambio en Y es:

$$\Delta \hat{Y} = \hat{f}\left(X_1 + \Delta X_1, X_2, \ldots, X_k\right) - \hat{f}\left(X_1, X_2, \ldots, X_k\right)$$

Funciones no lineales de una sola variable independiente

Dos enfoques complementarios:

1. **Polinomios en *X***

   La función de regresión poblacional se aproxima por un polinomio cuadrático, cúbico, o de mayor orden.

2. **Transformaciones Logarítmicas**

   - ***Y* y/ó *X* se transforman tomando sus logaritmos**
   - **Esto proporciona una interpretación de "porcentajes" que tiene más sentido en muchas aplicaciones**

1. Polinomios en *X*

Se aproxima la función de regresión poblacional por un polinomio:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \ldots + \beta_r X_i^r + u_i$$

Es el modelo de regresión clásico pero con regresores que son potencias de X

Se estima igual, aunque los coeficientes son más difíciles de interpretar

- **Se estima por MCO con los nuevos regresores**
- **Para interpretar la función de regresión estimada:**
  - **Dibujar la predicción como función de x**
  - **Calcular la predicción $\frac{\Delta Y}{\Delta X}$ para varios valores de x**
- **Las hipótesis sobre el grado r pueden ser contrastadas mediante estadísticos t- y F- sobre la(s) variable(s) o bloques de variables apropiadas.**

2. Funciones logarítmicas de *Y* y/ó *X*

- **Las transformaciones logarítmicas permiten estudiar relaciones en términos de "porcentajes" (como elasticidades), en vez de linealmente.**

$$\ln(x + \Delta x) - \ln(x) = \ln\left(1 + \frac{\Delta x}{x}\right) \cong \frac{\Delta x}{x}$$

- **Tres casos**

Caso	Función de regresión poblacional
I. Lineal – log	$Y_i = \beta_1 + \beta_2 ln(X_i) + u_i$
II. Log - lineal	$ln(Y_i) = \beta_1 + \beta_2 X_i + u_i$
III. Log - log	$ln(Y_i) = \beta_1 + \beta_2 ln(X_i) + u_i$

La interpretación del coeficiente de pendiente difiere en cada caso

La interpretación se encuentra aplicando la regla «antes y después»:

calcular el cambio en Y para un cambio en X

Caso I. Lineal - Log

$$Y_i = \beta_1 + \beta_2 \, ln(X_i) + u_i \qquad \text{(b)}$$

Ahora cambia X:

$$Y_i + \Delta Y = \beta_1 + \beta_2 \, ln(X_i + \Delta X) \qquad \text{(a)}$$

Restar (a) – (b)

$$\Delta Y = \beta_2 \left[ln(X_i + \Delta X) - ln(X_i) \right]$$

Ahora

$$ln(X_i + \Delta X) - ln(X_i) \cong \frac{\Delta X}{X_i}$$

Entonces:

$$\Delta Y \cong \beta_2 \frac{\Delta X}{X} \qquad \text{ó} \qquad \beta_2 \cong \frac{\Delta Y}{\Delta X / X} \quad \left(\Delta X \;\; pequeño \right)$$

Así, un incremento del 1% en X (multiplicando X por 1.01) se asocia con un cambio de $0.01\beta_2$ en Y.

Caso II. Log - lineal

$$ln(Y_i) = \beta_1 + \beta_2 X_i + u_i \qquad \text{(b)}$$

Ahora cambia X: $$ln(Y_i + \Delta Y) = \beta_1 + \beta_2 (X_i + \Delta X) \qquad \text{(a)}$$

Restar (a) – (b): $$ln(Y_i + \Delta Y) - ln(Y_i) = \beta_2 \Delta X$$

Así $$\frac{\Delta Y}{Y} \cong \beta_2 \Delta X \qquad \text{ó} \qquad \beta_2 \cong \frac{\Delta Y / Y}{\Delta X} \quad (\Delta X \;\; pequeño)$$

Así, un cambio en X en una unidad (ΔX = 1) se asocia con un cambio en E igual a 100 β_2 % (Y se incrementa en un factor de 1+ β_2)

Caso III. Log - log

$$ln(Y_i) = \beta_1 + \beta_2 \, ln(X_i) + u_i \qquad \text{(b)}$$

Ahora cambia X: $$ln(Y_i + \Delta Y) = \beta_1 + \beta_2 \, ln(X_i + \Delta X) \qquad \text{(a)}$$

Restamos: $$ln(Y_i + \Delta Y) - ln(Y_i) = \beta_2 \left[ln(X_i + \Delta X) - ln(X_i) \right]$$

Así $$\frac{\Delta Y}{Y} \cong \beta_2 \frac{\Delta X}{X} \quad o \quad \beta_2 \cong \frac{\Delta Y / Y}{\Delta X / X} \quad (\Delta X \ \ pequeño)$$

Así, un cambio del 1% en X se asocia con un cambio de β_2 % en Y. Es una elasticidad.

3.- INTERACCIONES ENTRE VARIABLES INDEPENDIENTES

Para los casos en los que $\frac{\Delta Y}{\Delta X_1}$ puede depender de X_2, buscamos cómo modelizar estas interacciones.

(a) Interacciones entre dos variables binarias

$$Y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + u_i$$

Las variables D_{1i} y D_{2i} son binarias.
β_2 es el efecto de cambiar de $D_1 = 0$ a $D_1 = 1$. En esta especificación, este efecto no depende del valor de D_2.
Para permitir que el efecto de cambiar D_1 dependa de D_2, se incluye el «término interacción» $D_1 \times D_2$ como un regresor adicional:

$$Y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 \left(D_{1i} \times D_{2i} \right) u_i$$

Interpretación de los coeficientes:

Regla General: comparar los distintos casos

$$E\left(Y_i \middle| D_{1i}=0, D_{2i}=d_2\right)=\beta_1+\beta_3 d_2 \qquad (b)$$

$$E\left(Y_i \middle| D_{1i}=1, D_{2i}=d_2\right)=\beta_1+\beta_2+\beta_3 d_2+\beta_4 d_2 \qquad (a)$$

Restar (a)-(b)

$$E\left(Y_i \middle| D_{1i}=1, D_{2i}=d_2\right)-E\left(Y_i \middle| D_{1i}=0, D_{2i}=d_2\right)=\beta_2+\beta_4 d_2$$

- El efecto de D_1 depende de d_2
- β_4 = incremento del efecto de D_1, cuando $D_2 = 1$

(b) Interacciones entre variables continuas y binarias

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + u_i$$

D_i es binaria y X_i es continua. Para permitir que el efecto de X dependa de D, se incluye el «término interacción», $D_i \times X_i$ como regresor:

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 \left(D_i \times X_i \right) + u_i$$

Interpretación de los coeficientes:
Regla general: comparar varios casos

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 \left(D_i \times X_i \right) \qquad (b)$$

Ahora cambia X

$$Y_i + \Delta Y = \beta_1 + \beta_2 D_i + \beta_3 (X_i + \Delta X) + \beta_4 \left(D_i \times (X_i + \Delta X) \right) \qquad (a)$$

Restar (a)-(b) $\Delta Y = \beta_3 \Delta X + \beta_4 D \Delta X$ o $\dfrac{\Delta Y}{\Delta X} = \beta_3 + \beta_4 D$

El efecto de X depende de D y $\beta_4 =$ incremento del efecto de X, cuando $D = 1$

Interacciones binaria-continua: las dos líneas de regresión

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i \times X_i) + u_i$$

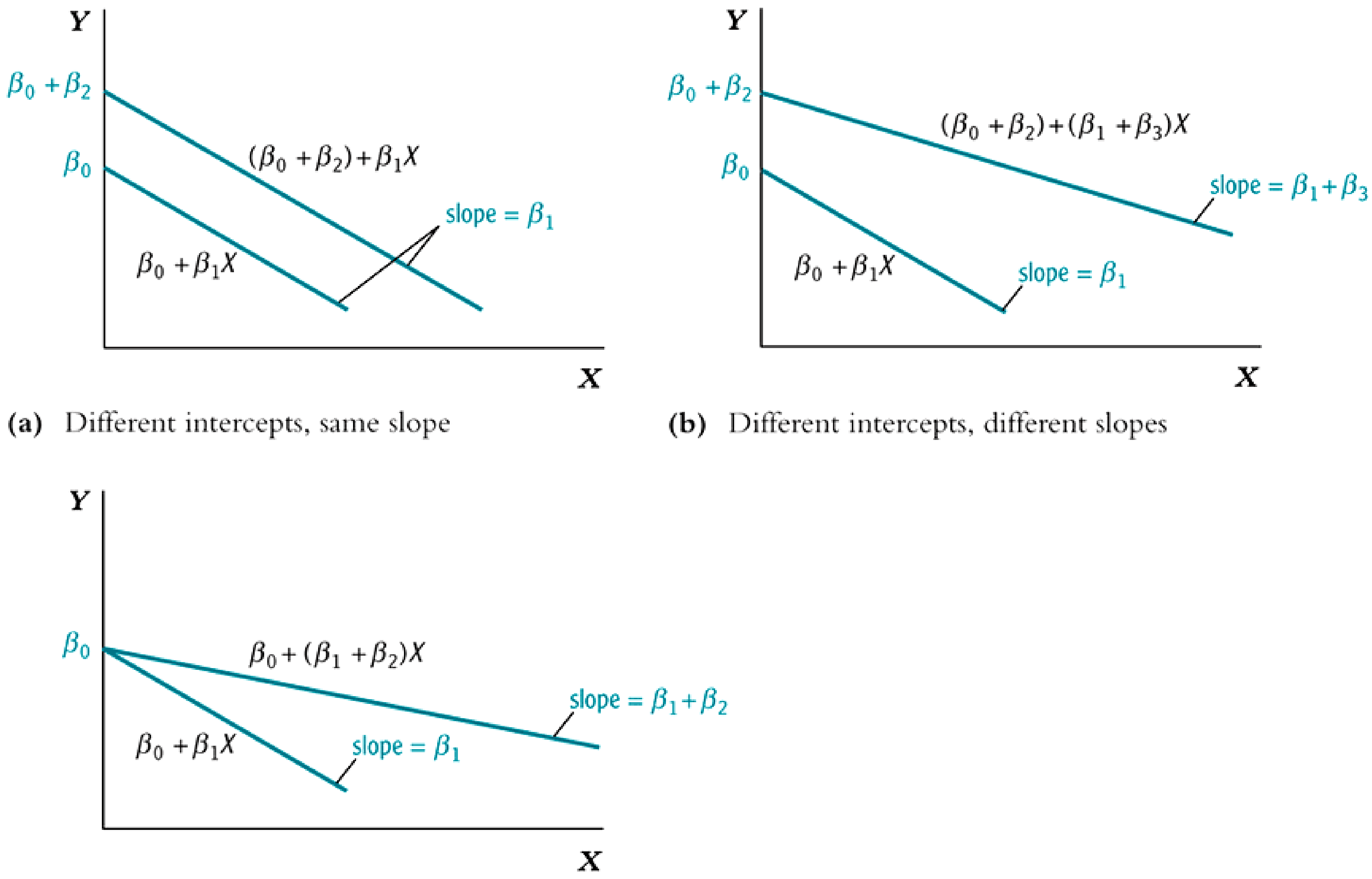
Observaciones con $D_i = 0$:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Observaciones con $D_i = 1$:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 + \beta_3 X_i + \beta_4 X_i + u_i \\ &= (\beta_1 + \beta_2) + (\beta_3 + \beta_4) X_i + u_i \end{aligned}$$

FIGURE 6.8 **Regression Functions Using Binary and Continuous Variables**



(a) Different intercepts, same slope

(b) Different intercepts, different slopes

(c) Same intercept, different slopes

Interactions of binary variables and continuous variables can produce three different population regression functions: (a) $\beta_0 + \beta_1 X + \beta_2 D$ allows for different intercepts but has the same slope; (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D)$ allows for different intercepts and different slopes; and (c) $\beta_0 + \beta_1 X + \beta_2 (X \times D)$ has the same intercept but allows for different slopes.

(c) Interacciones entre dos variables continuas

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + u_i$$

X_{1i} y X_{2i} son continuas. El efecto de X_1 no depende de X_2 y el efecto de X_2 no depende X_1.
Para permitir que el efecto de X_1 dependa de X_2 se incluye el «término de interacción» $X_1 \times X_2$ como regresor

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \beta_4 \left(X_{1i} \times X_{2i} \right) + u_i$$

Regla general: comparar los diferentes casos

$$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 \left(X_1 \times X_2 \right) \qquad (b)$$

Ahora cambia X_1

$$\Delta Y = \beta_1 + \beta_2 \left(X_1 + \Delta X_1 \right) + \beta_3 X_2 + \beta_4 \left[\left(X_1 + \Delta X_1 \right) \times X_2 \right] \qquad (a)$$

Restar (a) – (b):

$$\Delta Y = \beta_2 \Delta X_1 + \beta_4 X_2 \Delta X_1 \quad o \quad \frac{\Delta Y}{\Delta X_1} = \beta_2 + \beta_4 X_2$$

- **El efecto de X_1 depende de X_2**
- $\beta_4 =$ **incremento del efecto de X_1 a partir de un cambio unitario en X_2**

Aplicación Stock y Watson

Interpretación de los coeficientes de las variables ficticias - Ejemplo

Relación entre el salario hora trabajada percibido por el trabajador n-ésimo (W_i) y su nivel de estudios (variable cualitativa representada por 3 variables ficticias):

$$E_{i1} = \begin{cases} 1 & \text{Sin estudios ó sólo con estudios primarios (EP)} \\ 0 & \text{En caso contrario} \end{cases}$$

$$E_{i2} = \begin{cases} 1 & \text{Con estudios medios (no superiores) (EM)} \\ 0 & \text{En caso contrario} \end{cases}$$

$$E_{i3} = \begin{cases} 1 & \text{Con estudios superiores (ES)} \\ 0 & \text{En caso contrario} \end{cases}$$

$$W_i = \beta_1 E_{i1} + \beta_2 E_{i2} + \beta_3 E_{i3} + u_i$$

La matriz de regresores es:

$$X = \begin{bmatrix} \underset{[n_1\times 1]}{1} & 0 & 0 \\ 0 & \underset{[n_2\times 1]}{1} & 0 \\ 0 & 0 & \underset{[n_3\times 1]}{1} \end{bmatrix}$$

donde $\underset{[n_j\times 1]}{1}$ es un vector columna de unos de dimensión igual al número de trabajadores con educación de nivel $j.\left(n_j\right)$

Las ecuaciones normales serán: $X'X\hat{\beta} = X'W$, es decir:

$$\begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum_{i\in EP} w_i \\ \sum_{i\in EM} w_i \\ \sum_{i\in ES} w_i \end{bmatrix}$$

Por lo tanto,

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n_j} w_j}{n_j} = \bar{w}_j$$

es decir es el salario medio en cada nivel de educación.

Se debe tener cuidado con problemas de multicolinealidad exacta que pueden aparecer - TRAMPA DE LAS VARIABLES FICTICIAS

Si el modelo lineal completo es:

$$W_i = \beta_1 + \beta_2 z_i + \beta_3 x_i + \beta_4 E_{i1} + \beta_5 E_{i2} + \beta_6 E_{i3} + u_i$$

Como $1 = E_{i1} + E_{i2} + E_{i3}, \forall n$ **habrá multicolinealidad exacta y no es posible la estimación de todos los parámetros del modelo**

Soluciones:

- **Reemplazar la constante por** $E_{i1}+E_{i2}+E_{i3}$**:**

$$\begin{aligned} W_i &= \beta_1\left(E_{i1}+E_{i2}+E_{i3}\right)+\beta_2 z_i+\beta_3 x_i+\beta_4 E_{i1}+\beta_5 E_{i2}+\beta_6 E_{i3}+u_i \\ &= \beta_2 z_i+\beta_3 x_i+\left(\beta_1+\beta_4\right)E_{i1}+\left(\beta_1+\beta_5\right)E_{i2}+\left(\beta_1+\beta_6\right)E_{i3}+u_i \\ &= \beta_2 z_i+\beta_3 x_i+\delta_1 E_{i1}+\delta_2 E_{i2}+\delta_3 E_{i3}+u_i \end{aligned}$$

Que es un modelo sin término constante. Aquí $\delta_1=\left(\beta_1+\beta_4\right)$ **es una combinación del salario «autónomo»** $\left(\beta_1\right)$ **y del nivel de estudios** $\left(\beta_4\right)$

- **Reemplazar** E_{i2} **por** $\left(1-E_{i1}-E_{i3}\right)$ **, por ejemplo:**

$$\begin{aligned} W_i &= \beta_1+\beta_2 z_i+\beta_3 x_i+\beta_4 E_{i1}+\beta_5(1-E_{i1}-E_{i3})+\beta_6 E_{i3}+u_i \\ &= \left(\beta_1+\beta_5\right)+\beta_2 z_i+\beta_3 x_i+\left(\beta_1-\beta_5\right)E_{i1}+\left(\beta_6-\beta_5\right)E_{i3}+u_i \\ &= \theta_0+\beta_2 z_i+\beta_3 x_i+\theta_1 E_{i1}+\theta_3 E_{i3}+u_i \end{aligned}$$

4.- ERRORES DE ESPECIFICACIÓN

2 casos: Omisión de variable relevante e inclusión de variable irrelavante

1.- Omisión de variable relevante

Modelo correcto: $$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i$$

Modelo incorrecto: $$Y_i = \beta_1^* + \beta_2^* X_{i2} + u_i^*$$

Estimando en el modelo incorrecto:

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^{n} Y_i \left(X_{i2} - \bar{X}_2 \right)}{\sum_{i=1}^{n} \left(X_{i2} - \bar{X}_2 \right)^2}$$

Sustituyendo Y_i por el modelo correcto:

$$\hat{\beta}_2^* = \frac{\beta_2 \sum_{i=1}^{n} \left(X_{i2} - \bar{X}_2 \right)^2 + \beta_3 \sum_{i=1}^{n} X_{i3} \left(X_{i2} - \bar{X}_2 \right) + \sum_{i=1}^{n} u_i \left(X_{i2} - \bar{X}_2 \right)}{\sum_{i=1}^{n} \left(X_{i2} - \bar{X}_2 \right)^2}$$

¿Qué propiedades tiene?

$$E\left[\hat{\beta}_2^* \middle| X \right] = \beta_2 + \beta_3 \frac{\sum_{i=1}^{n} X_{i3} \left(X_{i2} - \bar{X}_2 \right)}{\sum_{i=1}^{n} \left(X_{i2} - \bar{X}_2 \right)^2}$$

$$= \beta_2 + \beta_3 \frac{S_{X_2 X_3}}{S_{X_2}^2}$$

Será un estimador sesgado y el sesgo depende de la correlación entre las variables X_2 y X_3

$$Var\left(\hat{\beta}_2^* \middle| X\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_{i2} - \bar{X}_2\right)^2}$$

Mientras que la varianza del estimador en el modelo correcto será:

$$Var\left(\hat{\beta}_2 \middle| X\right) = \frac{\sigma^2}{\left(1 - R_{23}^2\right)\sum_{i=1}^{n}\left(X_{i2} - \bar{X}_2\right)^2}$$

donde R_{23}^2 es el R^2 de la regresión de X_2 sobre X_3

Cuando se omite una variable relevante, ésta entra a formar parte de las perturbaciones. Si la correlación entre X_2 y X_3 es nula, se cumplirán las hipótesis, por lo que la estimación en el modelo incorrecto no presentará problemas.

Por el contrario, si $R_{23}^2 \cong 1$, la varianza del estimador en el modelo correcto se hace enorme, implicando una bajísima precisión en las estimaciones. Además la información de X_3 ya está incluida en X_2 por lo que excluirla del modelo no supone un coste elevado. Este problema es la MULTICOLINEALIDAD

2.- Inclusión de variable irrelevante

Modelo correcto: $$Y_i = \beta_1^* + \beta_2^* X_{i2} + u_i^*$$

Modelo incorrecto: $$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i$$

Estimando en el modelo incorrecto y hallando las propiedades del estimador:

Si se cumplen las hipótesis:

$$E\left[\hat{\beta}_2 \middle| X\right] = \beta_2^*$$ **Insesgado**

$$E\left[\hat{\beta}_3 \middle| X\right] = 0$$ **Insesgado ya que no debería estar en el modelo**

Si se no cumplen las hipótesis y $E\left[u_i \middle| X_{3i}\right] \neq 0$ **:**

$$E\left[\begin{matrix}\hat{\beta}_2 \\ \hat{\beta}_3\end{matrix} \middle| X\right] = \begin{pmatrix}\beta_2^* \\ 0\end{pmatrix} + \left(X'X\right)^{-1} X' E\left[\underbrace{u \middle| X}_{\neq 0}\right]$$

por lo que las estimaciones serán sesgadas. Lo crucial de este resultado no es si las variables son relevantes, es la correlación con las perturbaciones

La varianza en el modelo correcto:

$$Var\left(\hat{\beta}_2^* \middle| X\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(X_{i2} - \bar{X}_2\right)^2}$$

La varianza en el modelo incorrecto:

$$Var\left(\hat{\beta}_2 \middle| X\right) = \frac{\sigma^2}{\left(1 - R_{23}^2\right)\sum_{i=1}^{n}(X_{2i} - \bar{X}_2)^2}$$

donde R_{23}^2 es el R^2 de la regresión de X_2 sobre X_3

Si la correlación entre los regresores es nula, la inclusión de la variable irrelevante no tiene ningún efecto. Pero a medida que la correlación crece en valor absoluto, la varianza del estimador se incrementa; disminuyendo consiguientemente la precisión de los estimaciones con las consecuencias que esto tiene sobre la inferencia en el modelo.

5.- MULTICOLINEALIDAD

2 tipos de multicolinealidad:

Estricta ó perfecta: $\left|X'X\right| = 0 \quad rango\left(\underset{k\times k}{X'X}\right) < k$

- **Incumplimiento del supuesto [H4]**
- **El sistema de ecuaciones normales tiene infinitas soluciones**

No estricta ó imperfecta: $\left|X'X\right| \cong 0 \quad rango\left(\underset{k\times k}{X'X}\right) = k$

- **Cumplimiento del supuesto [H4]**
- **Existe alta correlación entre los regresores**
- **A mayor correlación, menor determinante $\left|X'X\right|$ y mayor gravedad del problema**

Ejemplo de multicolinealidad estricta ó exacta: Sea

$$Y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

donde $x_{2i} = \lambda x_{3i}$ **. Entonces:**

$$Y_i = \beta_1 + \beta_2 \left(\lambda x_{3i}\right) + \beta_3 x_{3i} + u_i$$
$$= \beta_1 + \left(\lambda \beta_2 + \beta_3\right) x_{3i} + u_i$$

Tiene una solución para β_1 **y otra para** $\lambda \beta_2 + \beta_3$ **pero ésta última es compatible con infinitas combinaciones de valores para** β_2 **y** β_3**.**

Multicolinealidad no estricta ó aproximada - EFECTOS

- **Las varianzas y covarianzas de los estimadores se hacen grandes:**

$$Var\left(\hat{\beta}\middle|X\right)=\sigma^2\left(X'X\right)^{-1}=\sigma^2\frac{\left[Adj\left(X'X\right)\right]}{\left|X'X\right|}$$

Poca precisión que implica que pequeñas variaciones en la muestra provocan grandes variaciones en la estimación

- **Propensión a aceptar casi cualquier hipótesis sobre los parámetros, en concreto la hipótesis de significatividad individual:** $H_0 : \beta_i = 0$

$$t=\frac{\hat{\beta}_i}{ES\left(\hat{\beta}_i\right)}\sim t_{(n-k)}$$

- **Pero no afecta al contraste de significación conjunta:**

$$F=\frac{n-K}{n-1}\frac{R^2}{\left(1-R^2\right)}\sim F_{(k-1,n-k)}$$

Multicolinealidad no estricta ó aproximada - CAUSAS

- **Variables explicativas no estacionarias que evolucionen de forma parecida – Transformar los datos**

- **Inclusión de variables explicativas cuya información ya está incorporada en otras variables explicativas – Excluir dichas variables**

- **Inclusión de retardos de variables explicativas – frecuentemente correlación elevada (Series Temporales)**

- **Problemas de escala de los datos -** $\left|X'X\right|$ **depende de las unidades de medida - cambiar la escala**

Multicolinealidad no estricta ó aproximada - DETECCIÓN

- **Examen de la correlación entre las variables explicativas del modelo**
  - **Cálculo de las correlaciones simples entre pares de variables**
  - **Regresiones entre variables explicativas**

- **Análisis del tamaño de *X'X***
  - **Examinar el determinante (producto de auto-valores en matrices simétricas)**

  ¿Cuándo es pequeño? Examinar los tamaños relativos de los auto-valores: *Número de Condición*

$$Si\ \sqrt{\frac{mayor\ autovalor}{menor\ autovalor}} > 20 \Rightarrow$$ **problemas**

Multicolinealidad no estricta ó aproximada - SOLUCIONES

- **Difícil solución ya que supone que se dispone de insuficiente información para estimar TODOS los parámetros del modelo**
- **Algunas soluciones:**
    - **Obtener más datos**
    - **Imponer restricciones**
    - **Suprimir variables**
    - **Transformar variables**

6.- HETEROCEDASTICIDAD

Se produce cuando en el modelo

$$Y = X\beta + u$$

La hipótesis [H4.1] pasa a ser:

$$\left[H\,4.1b\right] \qquad E\left[u_i^2 \,\middle|\, X\right] = \sigma_i^2$$

Así la hipótesis [H4] será:

$$E\left[uu' \,\middle|\, X\right] = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ & & & & \sigma_n^2 \end{bmatrix} = \sigma^2 \Omega$$

¿Qué propiedades tendrá ahora el estimador MCO?

$$\hat{\beta}_{MCO} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'u$$

$$E\left[\hat{\beta}_{MCO} \middle| X\right] = \beta$$

$$Var\left(\hat{\beta}_{MCO} \middle| X\right) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Si reformulamos la [H5] de la siguiente forma: $u \big| X \sim N\left(0, \sigma^2 \Omega\right)$

$$\hat{\beta}_{MCO} \big| X \sim N\left(\beta, \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}\right)$$

Puede que este estimador, bajo las nuevas hipótesis, ya no sea el de mínima varianza, como lo era el MCO bajo las hipótesis anteriores. Por ello, se puede transformar el modelo de manera que se cumpla que la matriz de varianzas-covarianzas de *u* sea escalar.

¿Qué transformación utilizar?

Ω es una matriz simétrica, definida positiva por lo que puede descomponerse de la siguiente forma:

$$\Omega = VV' \rightarrow V^{-1}\Omega\left(V^{-1}\right)' = I_n$$

Siendo V una matriz cuadrada, no singular. Además: $\Omega^{-1} = \left(V^{-1}\right)'V^{-1}$

Si transformamos el modelo premultiplicando por V^{-1} tendremos:

$$V^{-1}Y = V^{-1}X\beta + V^{-1}u \Rightarrow Y^* = X^*\beta + u^*$$

donde $Y^* = V^{-1}Y;\; X^* = V^{-1}X;\; u^* = V^{-1}u$ y se cumplirá:

$$Var\left(u^* \middle| X\right) = \sigma^2 V^{-1}\Omega\left(V^{-1}\right)' = \sigma^2 I_n$$

Con esta transformación, se cumplirán las hipótesis iniciales, de manera que el estimador MCO en el modelo transformado será el de mínima varianza.

El estimador será: $\hat{\beta} = \left(X^{*\prime} X^{*}\right)^{-1} X^{*\prime} Y^{*}$

Deshaciendo las transformaciones:

$$\hat{\beta}_{MCG} = \left(X' \Omega^{-1} X\right)^{-1} X' \Omega^{-1} Y$$

ESTIMADOR DE MÍNIMOS CUADRADOS GENERALIZADOS

Propiedades del estimador MCG

- **Es insesgado** $E\left[\hat{\beta}_{MCG} \mid X\right] = \beta$
- **Varianza es mínima** $Var\left(\hat{\beta}_{MCG}\right) = \sigma^2 \left(X' \Omega^{-1} X\right)^{-1}$
- **Más eficiente que el MCO**

$$\left(X' \Omega^{-1} X\right)^{-1} - \left(X' X\right)^{-1} X' \Omega X \left(X' X\right)^{-1}$$ **es semidefinida negativa**

Estimación de σ^2

$$\hat{\sigma}^2_{MCG} = \frac{\hat{u}^{*\prime} \hat{u}^{*}}{n-k} = \frac{\hat{u}'_{MCG} \Omega^{-1} \hat{u}_{MCG}}{n-k}$$ **con** $\hat{u}_{MCG} = Y - X\hat{\beta}_{MCG}$

La suma de cuadrados de residuos será:

$$SCR = \hat{u}'_{MCG}\Omega^{-1}\hat{u}_{MCG} = Y'\Omega^{-1}Y - \hat{\beta}'_{MCG}X'\Omega^{-1}Y$$

Así

$$\hat{\sigma}^2_{MCG} = \frac{\hat{u}'_{MCG}\Omega^{-1}\hat{u}_{MCG}}{n-k} = \frac{Y'\Omega^{-1}Y - \hat{\beta}'_{MCG}X'\Omega^{-1}Y}{n-k}$$

Pero, no se puede utilizar el coeficiente de determinación como medida de ajuste ya que el modelo transformado puede no tener constante, no estando acotado entre 0 y 1. Además, con él mediríamos la capacidad de explicar y*, que no es la variable de interés.

La heterocedasticidad es la regla, no la excepción de manera que se debe siempre realizar un análisis exhaustivo en cualquier aplicación.

(EJEMPLOS)

HETEROCEDASTICIDAD - CAUSAS

- **La presencia de atípicos: observaciones que son muy diferentes en relación a las demás observaciones de la muestra**

- **Errores de especificación en el modelo – omisión de variable relevante**

- **El hecho de que cuanto mayores sean los valores de alguna de las variables del modelo provoca que habitualmente sea mayor la dispersión absoluta del modelo**

- **En los modelos con datos de sección cruzada, es habitual que las unidades muestrales presenten diferentes valores según el tamaño de dicha unidad provocando una mayor dispersión en el modelo.**

HETEROCEDASTICIDAD - CONSECUENCIAS

- **Estimaciones incorrectas de la matriz de varianzas-covarianzas de los estimadores**

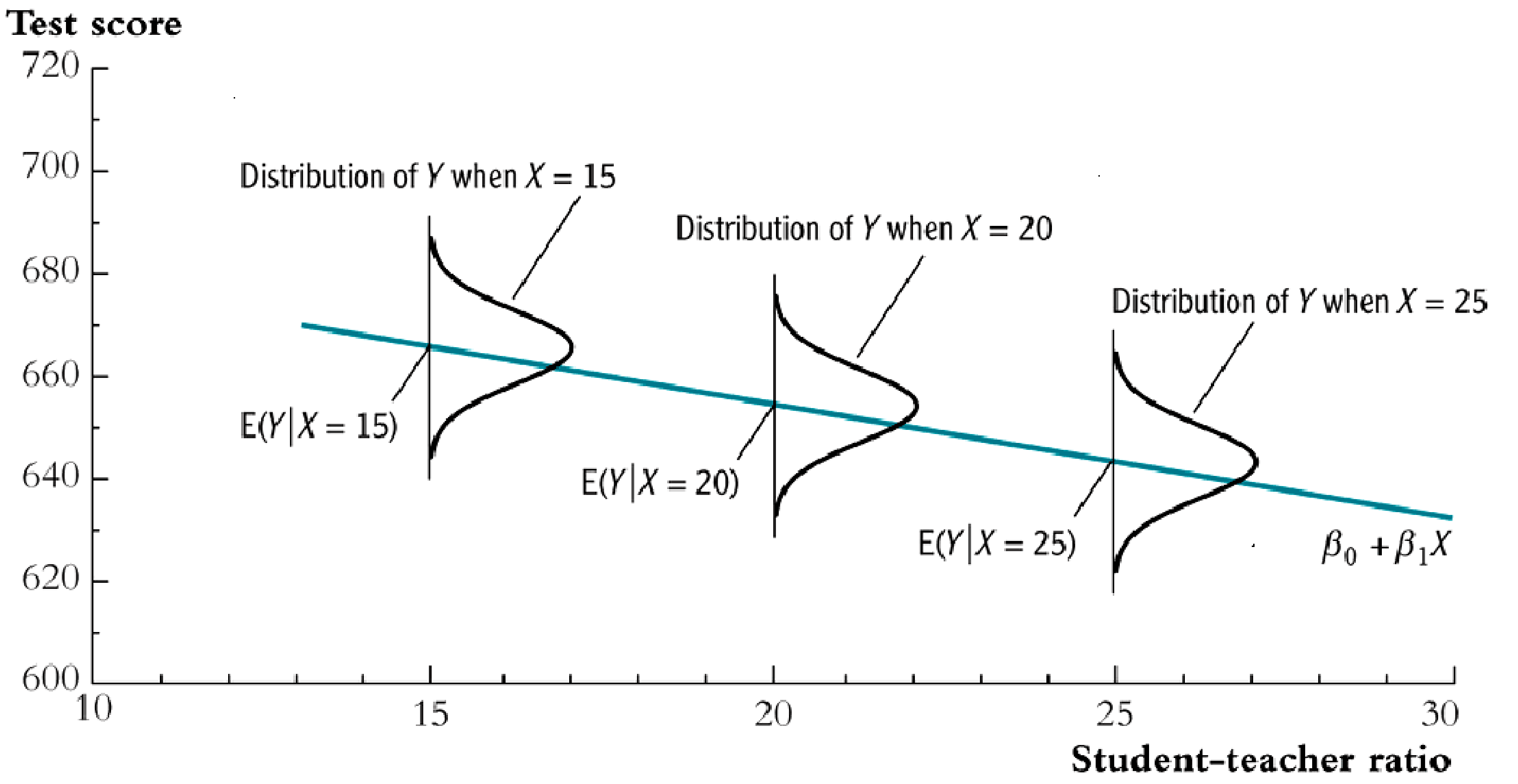
- **El estimador MCO de la varianza del modelo σ^2 será sesgado, sesgando así las estimaciones de las varianzas y covarianzas de los estimadores de los parámetros del modelo**

- $\hat{u}'\hat{u}/\sigma^2$ **ya no seguirá una distribución chi-cuadrado con (n – k) grados de libertad, como consecuencia de que no se cumple la hipótesis [H4] sino que**

$$\left[H4.b\right] \quad E\left[uu'\middle|X\right] = \sigma^2\Omega \quad y \quad u\middle|X \sim N\left(0, \sigma^2\Omega\right)$$

Homoscedasticidad en un gráfico:

FIGURE 4.4 The Conditional Probability Distributions and the Population Regression Line



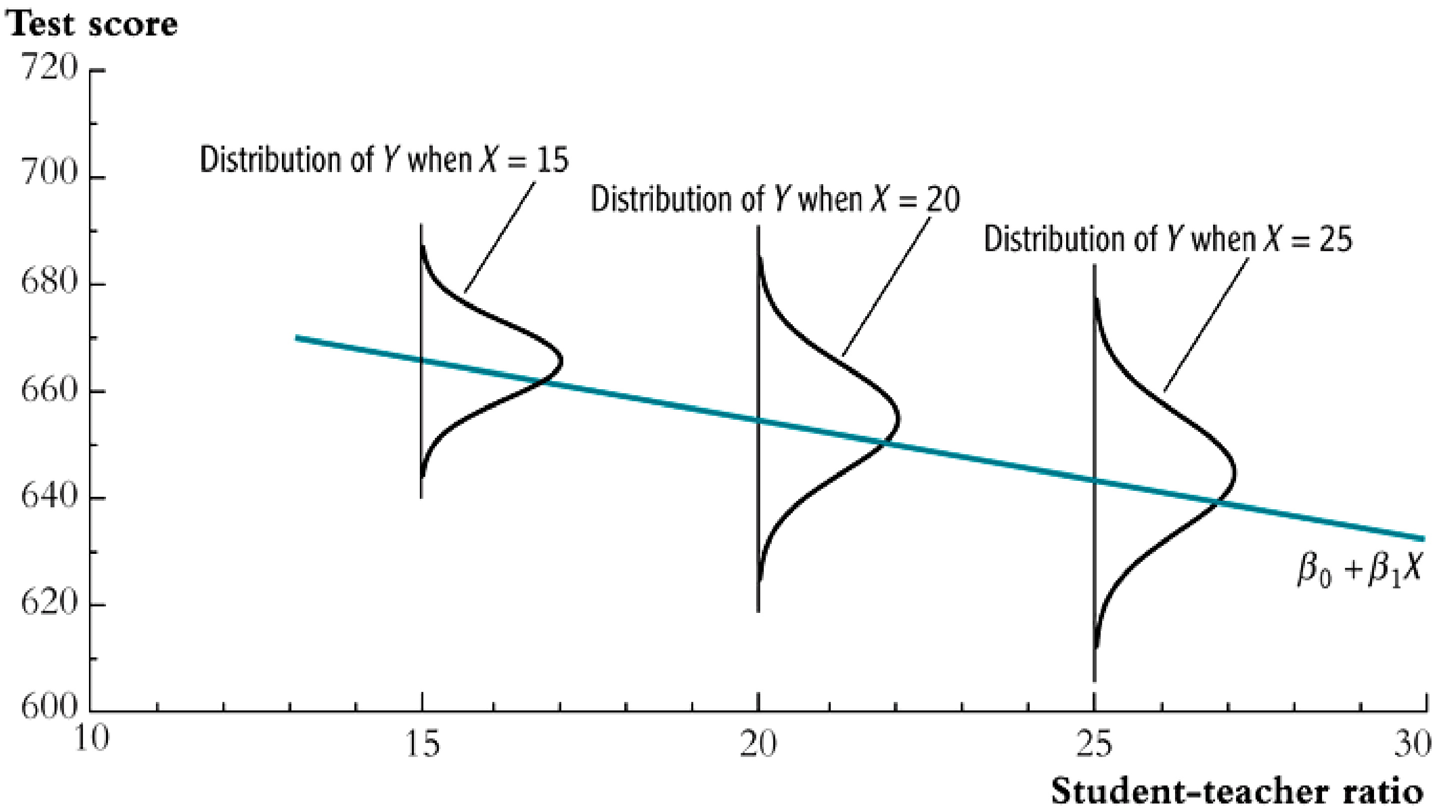
$E(u|X=x) = 0$ (u satisface la Hipótesis [H1] de MCO)

La varianza de u no cambia con (depende de) x

Heteroscedasticidad en un gráfico:

FIGURE 4.7 **An Example of Heteroskedasticity**

Like Figure 4.4, this shows the conditional distribution of test scores for three different class sizes. Unlike Figure 4.4, these distributions become more spread out (have a larger variance) for larger class sizes. Because the variance of the distribution of u given X, $\text{var}(u|X)$, depends on X, u is heteroskedastic.



***E*(*u*|*X*=*x*) = 0 (*u* satisface la Hipótesis [H1] de MCO)**

La varianza de *u* depende de *x*.

HETEROCEDASTICIDAD - DETECCIÓN

H_0 *: ausencia de heterocedasticidad*

1.- Contraste de Goldfeld y Quandt (1965)

- Suponen que la magnitud de la varianza de *u* depende de z_t , generalmente una variable explicativa del modelo
- Si la dependencia es positiva (mayor varianza cuando mayores son los valores de z_t)

a) Ordenar de menor a mayor las observaciones por los valores de z_t

b) Omitir p valores de la mitad de la muestra

c) Estimar el modelo original con las $\frac{n-p}{2}$ primeras observaciones y con las $\frac{n-p}{2}$ finales

d) Calcular las sumas residuales para cada regresión. Entonces, bajo el supuesto de homocedasticidad y normalidad del error,

$$\lambda = \frac{SR_2}{SR_1} \sim F_{(m,m)} \quad con \quad m = \frac{n-p}{2} - k$$

- Comentarios:

- **Si la relación es negativa, ordenar las observaciones de mayor a menor**

- **Elección de p (Harvey y Phillips (1974) no eliminar más de tercio de las observaciones)**

- **No rechazar la hipótesis no significa que no haya heterocedasticidad. Hay que realizar el contraste con todas las variables explicativas.**

2.- Contraste de Breusch y Pagan

Pasos del contraste:

a) Estimar el modelo original por MCO y obtener los residuos

b) Obtener los residuos normalizados al cuadrado:

$$\hat{e}_i^2 = \frac{\hat{u}_i^2}{\hat{\sigma}_u^2} \quad i = 1, 2, \ldots, n \quad \text{donde } \hat{\sigma}_u^2 = \frac{SCR}{n}$$

c) Estimar una regresión de $\hat{e}_i^2$ sobre las variables $z_{i1}, z_{i2}, \ldots, z_{ip}$ con constante y obtener la suma de cuadrados explicada en dicha regresión.

d) Bajo la hipótesis nula de homocedasticidad y la normalidad de *u*, el cociente $SCE/2$ calculado para la regresión de c):

$$\frac{SCE}{2} \sim \chi_{(p)}$$

Si los residuos son homocedásticos, *SCE* será pequeña ya que las variables *z* no tendrían poder explicativo. Así si SCE/2 es menor que el valor de la chi-cuadrado, no se rechaza la hipótesis nula de homocedasticidad.

Si se rechaza la homocedasticidad, se transformaría el modelo con $\sqrt{z_i'\hat{\alpha}}$ en el que la aplicación de MCO nos llevaría a obtener el estimador MCG.

La lista de variables a incluir en *z* debería ser:

- **Corta**
- **Incluir pocas variables que no estén incluidas en los regresores**

3.- Contraste de Glejser

Etapas:

a) Estimar el modelo original por MCO y obtener los residuos

b) Estimar: $|\hat{u}_i| = \delta_0 + \delta_1 z_i^h + v_i$ **ó**

$$\hat{u}_i^2 = \delta_0 + \delta_1 z_i^h + v_i$$

para distintos valores de *h*, $h = \left\{-1, 1, \frac{1}{2}, -\frac{1}{2}\right\}$ **y escoger el valor de *h* que proporcione mejor regresión**

c) Encontrado el esquema, transformar el modelo para aplicar MCO

Comentarios:

- **Si no se encuentra ningún esquema válido, repetir el proceso con otra variable.**
- **Cuidado con la transformación de las observaciones: será distinta si se utilizan residuos en valor absoluto ó residuos al cuadrado.**

4.- Contraste de White (1980)

(a) Estimar el modelo por MCO ignorando la posible heterocedasticidad

(b) Estimar una regresión del cuadrado de los residuos MCO sobre una constante, los regresores del modelo original, sus cuadrados y sus productos cruzados de 2º orden.

(c) Al aumentar el tamaño muestral, $nR^2 \sim \chi^2_{p-1}$ siendo p el número de regresores del modelo en b) y R^2 el coeficiente de determinación del modelo en b)

Cuando n crece, el $R^2 \to 0$ bajo la hipótesis nula de homocedasticidad. Sólo cuando la varianza del término de error del modelo dependa de las variables explicativas, el R^2 no tenderá a cero. En este caso, $nR^2 > \chi^2_{(p-1)}$

HETEROCEDASTICIDAD - CORRECCIÓN

Se distinguen dos situaciones:

- Cuando el esquema de heterocedasticidad es conocido: Mínimos Cuadrados Ponderados (MCP) ya son ELIO.

EJEMPLOS ….

- **Errores estándar robustos a heterocedasticidad de White**

En el modelo de regresión múltiple:

$$Y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i \quad con \quad E\left[uu' \middle| X\right] = \sigma_i^2$$

La varianza de cualquier coeficiente estimado puede calcularse como:

$$Var\left(\hat{\beta}_j\right) = \frac{\sum \hat{\omega}_{ji}^2 \hat{u}_i^2}{\left(\sum \hat{\omega}_{ji}^2\right)^2}$$

donde $\hat{u}_i$ son los residuos del modelo original, y $\hat{\omega}_{ji}$ son los residuos de la regresión auxiliar de x_j sobre el resto de las variables explicativas del modelo

White demostró que el estimador de la varianza propuesto es un estimador consistente de la varianza del coeficiente, es decir, que a medida que el tamaño muestral aumenta el estimador propuesto converge a la varianza del coeficiente.