

Teoría Econométrica I - EAE- 250-A

Problemas de Especificación y Datos

Tatiana Rosá

Instituto de Economía - Pontificia Universidad Católica de Chile

Septiembre 2021

Introducción

- En esta sección veremos problemas de especificación del modelo de regresión lineal
- En particular, qué sucede cuándo la matriz de varianzas es no escalar
- Veremos el estimador de Mínimos Cuadrados Generalizados (MCG) y MCG Factibles (MCGF)
- Veremos también Mínimos Cuadrados No lineales
- Y algunos problemas comunes de datos

Mínimos Cuadrados Generalizados (Aitken)

- Vamos a relajar el segundo supuesto del modelo de regresión lineal ($V(\hat{\beta}) = \sigma^2 I$).
- Con lo que los supuestos de este modelo son:
 - (i) Modelo lineal en los parámetros
 - (ii) La matriz de varianza y covarianza es *no escalar* ó *no esférica*
$$V(Y|X) = \mathbb{E}((Y - \mathbb{E}(Y|X))(Y - \mathbb{E}(Y|X))'|X) = \Sigma_{N \times N} = \sigma^2 \Omega$$
 - (iii) Regresores estocásticos, es decir $X_{N \times K}$ son aleatorios.
 - (iv) Rango completo de la matriz $X \rightarrow X'X$ también es de rango completo.
 - (v) Independencia en media $E(u|X) = 0$ (condición de exogeneidad)

Propiedades con matriz de varianza no esférica

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\begin{aligned}\mathbb{E}[\hat{\beta}|X] &= (X'X)^{-1}X'\mathbb{E}[Y|X] \\ &= (X'X)^{-1}X'X\beta \\ &= \beta\end{aligned}$$

- El estimador es condicionalmente insesgado
- Dado que los X son aleatorios, dada la Ley de Esperanzas Iteradas, tenemos que $\mathbb{E}[\mathbb{E}[\hat{\beta}|X]] = \mathbb{E}[\hat{\beta}] = \beta$
- Es incondicionalmente insesgado también

Varianza

- Con respecto a la varianza del estimador:

$$V(\hat{\beta}|X) = (X'X)^{-1}X'V(Y|X)X(X'X)^{-1}$$

$$V(\hat{\beta}|X) = (X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}$$

$$V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

- Por lo general $V(\hat{\beta}|X) \neq \sigma^2(X'X)^{-1}$ en este modelo.
- Recordar también que:

$$\text{plim}N^{-1}X'X = \text{plim}N^{-1}\sum x_i'x_i = D$$

$$\text{plim}N^{-1}X'\Omega X = C$$

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 D^{-1} C D^{-1})$$

Aitken propone...

- **Aitken propone lo siguiente: si Ω es conocida:**

1. Calcule la siguiente descomposición:

$$\Omega^{-1} = H' H \text{ Cholesky}$$

— Si la matriz A es simétrica y definida positiva, puede ser descompuesta como:
 $A = LL' = U'U$ donde L es una matriz triangular inferior y U es una matriz triangular superior

2. Después premultiplique por la izquierda el modelo por H :

$$Y^* = HY \text{ y } X^* = HX$$

$$\mathbb{E}[Y^*|X] = HX\beta$$

$$\begin{aligned} V[Y^*|X] &= HV[Y|X]H' &= H[\sigma^2\Omega]H' \\ & &= \sigma^2 H\Omega H' \\ & &= \sigma^2 I \end{aligned}$$

MCG/GLS

- Aplicando OLS al modelo transformado tenemos el siguiente estimador de GLS:

$$\begin{aligned}\hat{\beta}_{GLS} &= (X^{*'}X^*)^{-1}X^{*'}Y^* \\ \hat{\beta}_{GLS} &= (X'H'HX)^{-1}X'H'HY\end{aligned}$$

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

- Este estimador es MELI
- Además:

$$\begin{aligned}S_{GLS}^2 &= (N - k)^{-1}(Y^* - X^*\hat{\beta}_{GLS}^*)(Y^* - X^*\hat{\beta}_{GLS}^*) \\ \mathbb{E}[S_{GLS}^2|X] &= \sigma^2 \text{Tarea: probarlo} \\ S_{GLS}^2 &\xrightarrow{p} \sigma^2\end{aligned}$$

Observaciones

- Si asumimos normalidad multivariada de y y podemos hacer inferencia exacta” (las distribuciones utilizadas no son aproximaciones sino que las distribuciones reales) y podemos usar los test que conocemos del **modelo normal** de regresión.
- Si no asumimos normalidad tendremos que usar aproximaciones asintóticas. En ese caso podemos usar WLLN, TCL, TC, TMC, DM, etc.

$$\sqrt{N}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, V)$$

donde $V = \sigma^2 \text{plim}(\frac{1}{N} X' \Omega^{-1} X)^{-1}$, el cual desarrollando a través de álgebra de límites del plim tenemos:

$$V = \text{plim} S^2(\frac{1}{N} X' \Omega^{-1} X)^{-1}$$

- Pero conocer la forma de Ω es algo idílico
⇒ Usaremos Mínimos Cuadrados Generalizados Factibles.
- Primero vamos a ver como llegamos a esa distribución asintótica

Distribución asintótica de MCG/GLS

- Recuerde que:

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

$$\hat{\beta}_{MCG} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon \quad (2)$$

$$\text{ya que } Y = X\beta + \varepsilon \quad (3)$$

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) = \sqrt{n}(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon \quad (4)$$

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) = \underbrace{\left(\frac{1}{n}X'\Omega^{-1}X\right)^{-1}}_A \underbrace{\frac{1}{\sqrt{n}}X'\Omega^{-1}\varepsilon}_B \quad (5)$$

Distribución asintótica de MCG/GLS

- Usaremos Ley Débil de Grandes Números (WLLN) y el Teorema Central del Límite (CLT) para encontrar la convergencia en distribución del estimador MCG
- Partiremos trabajando A ,

$$A = \frac{1}{n} X' \Omega^{-1} X \xrightarrow{p} \text{plim} \left(\frac{1}{n} X' \Omega^{-1} X \right)$$

- Esto último por Ley de Grandes Números (tenemos iid y momentos acotados)
- Por teorema de continuidad:

$$\left(\frac{1}{n} X' \Omega^{-1} X \right)^{-1} \xrightarrow{p} \left(\text{plim} \frac{1}{n} X' \Omega^{-1} X \right)^{-1}$$

Distribución asintótica de MCG/GLS

- Por otro lado, buscamos aplicar CLT en B:

$$B = \frac{1}{\sqrt{n}} X' \Omega^{-1} \varepsilon$$
$$B = \sqrt{n} \left(\frac{1}{n} X' \Omega^{-1} \varepsilon \right)$$

- Recuerde que CLT de Lindeberg-Levy (versión univariada) dice que:
 $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \rightarrow N(0, 1),$
- Esto bajo momentos acotados y un proceso generador de datos (p.g.d.) que arroja observaciones i.i.d
- Podemos tomar $\frac{1}{n} X' \Omega^{-1} \varepsilon$ como un promedio (que está multiplicado por \sqrt{n} en la expresión anterior), por lo que necesitamos encontrar su esperanza y varianza (en estricto rigor del argumento de la sumatoria implícita).

Distribución asintótica de MCG/GLS

- (i) $\mathbb{E}(X'\Omega^{-1}\varepsilon) = 0$ Independiente de la especificación del modelo.

- $\mathbb{E}(X'\Omega^{-1}\varepsilon|X) = X'\Omega\mathbb{E}(\varepsilon|X) = 0$ dado que $\mathbb{E}(\varepsilon|X) = 0$

Por Ley de Esperanzas Iteradas (LEI): $\mathbb{E}(X'\Omega^{-1}\varepsilon) = \mathbb{E}(X'\Omega^{-1} \underbrace{\mathbb{E}(\varepsilon|X)}_0)$.

- (ii) Encontrar la varianza tiene su cuidado.

$$V(X'\Omega^{-1}\varepsilon) = \mathbb{E}(X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X) = V$$

Si no hacemos supuestos adicionales sólo podemos decir que la varianza es V y que por Lindeverg-Levy

$$\frac{1}{\sqrt{n}}X'\Omega^{-1}\varepsilon \xrightarrow{d} N(0, V)$$

- Lo que implica que obtenemos una forma de sandwich para la matriz de varianzas y covarianzas

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) \xrightarrow{d} N(0, A^{-1}VA^{-1})$$

Distribución asintótica de MCG/GLS

- Si hacemos el supuesto que $\mathbb{E}(\varepsilon\varepsilon'|X) = \sigma^2\Omega$, nos queda la expresión típica y que postulamos la clase pasada:

$$\begin{aligned}V(X'\Omega^{-1}\varepsilon) &= \mathbb{E}(X'\Omega^{-1}\mathbb{E}(\varepsilon\varepsilon'|X)\Omega^{-1}X) \text{ por LEI} \\V(X'\Omega^{-1}\varepsilon) &= \sigma^2\mathbb{E}(X'\Omega^{-1}\Omega\Omega^{-1}X) \\&= \sigma^2\mathbb{E}(X'\Omega^{-1}X) \\&= \sigma^2\text{plim}\left(\frac{1}{n}(X'\Omega^{-1}X)\right)\end{aligned}$$

- Esto último por Ley de Grandes Números dado que trabajamos con momentos finitos y un p.g.d. i.i.d.
- En este caso usamos el TCL de Lindeberg-Levy Multivariado:

$$\frac{1}{n}X'\Omega^{-1}\varepsilon \xrightarrow{d} N(0, \text{plim}\sigma^2X'\Omega^{-1}X)$$

Distribución asintótica de MCG/GLS

- Entonces aplicando Slutsky, tenemos:

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) \xrightarrow{d} \text{plim} \left(\frac{1}{n} X' \Omega^{-1} X \right)^{-1} N \left(0, \text{plim} \sigma^2 \left(\frac{1}{n} X' \Omega^{-1} X \right) \right)$$

- Y lo podemos escribir:

$$\sqrt{n}(\hat{\beta}_{MCG} - \beta) \xrightarrow{d} N \left(0, \text{plim} \sigma^2 \left(\frac{1}{n} X' \Omega^{-1} X \right)^{-1} \right)$$

- Que es bajo el caso usual con $\mathbb{E}(\varepsilon\varepsilon'|X) = \sigma^2\Omega$ y asumiendo que Ω es conocido. Ahora, si σ^2 es desconocido, usamos $S_{MCG}^2 = (Y - X\hat{\beta}_{MCG})'(Y - X\hat{\beta}_{MCG})$
- Notar que si estimamos por GLS el modelo transformado ya no tiene una constante: qué pasa entonces?**

Minimoc Cuadrados Generalizados Factibles (MCGF/FGLS)

- Muy lindo el GLS....pero no tenemos Ω ! ¿Qué hacemos?
- Tenemos algun $\hat{\Omega}$ tal que $\hat{\Omega} \xrightarrow{P} \Omega$ no tenemos problema, porque las propiedades asintóticas se mantienen (con muestras grandes claro!!)

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

- Pero... ¿Se puede estimar sin imponer forma funcional de Ω ?
- Recordemos que $\mathbb{E}(\varepsilon \varepsilon') = \sigma^2 \Omega$ y supongamos, por simplicidad, que $\sigma^2 = 1$. Realizaremos el cálculo de $\hat{\Omega}$.

Estimador de Omega y Distribución Asintótica

- Primero, se obtienen residuos de una regresión auxiliar MCO y se usan de tal manera que se tiene:

$$\hat{\Omega} = \frac{1}{N} \hat{\varepsilon}_{MCO} \hat{\varepsilon}_{MCO}' \xrightarrow{p} \Omega$$

- Usando los mismos argumentos anteriores, la varianza asintótica de MCGF es: $\frac{1}{N} A^{-1} B A^{-1}$

Donde $A = \text{plim} \left(\frac{1}{N} X' \hat{\Omega}^{-1} X \right)$ y $B = \text{plim} \left(\frac{1}{N} X' \hat{\Omega}^{-1} \hat{\varepsilon}_{MCG} \hat{\varepsilon}_{MCG}' \hat{\Omega}^{-1} X \right)$ siendo estas estimaciones robustas.

- Esto implica que la distribución asintótica de $\hat{\beta}_{MCGF}$ es:

$$\sqrt{N}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1})$$

Supuestos detrás

- Esto último bajo las condiciones que $\mathbb{E}(X'\epsilon) = 0$, que Ω es definida positiva y que $\mathbb{E}(X'\Omega^{-1}X)$ es no singular (demostración *Wooldridge "Econometric Analysis of Cross Section and Panel Data"* Capítulo 7, Teorema 7.3)
- La primera condición es fundamental para establecer la consistencia del estimador de MCGF
- Entonces si $\mathbb{E}(X'\Omega^{-1}\epsilon\epsilon'\Omega^{-1}X) = \mathbb{E}(X'\Omega^{-1}X)$ el estimador de la varianza asintótica de $\hat{\beta}_{MCGF}$ será:

$$\text{As. } V(\hat{\beta}_{MCGF}) = \frac{1}{N} \hat{A}^{-1} = (X'\hat{\Omega}^{-1}X)^{-1}$$

- Todo esto con el supuesto que los residuos del modelo de MCO en promedio no son distintos de los de MCGF.

Heterocedasticidad

- Ocorre cuando Ω es diagonal con distintos valores para cada entrada de la misma
- Los elementos de la diagonal distintos se dan en el caso de efectos pares (especialmente en educación), datos agrupados
- Ejemplo, en datos agrupados el modelo es de la forma:

$$y_{ij} = x'_{ij}\beta + \varepsilon_{ij} \text{ con } i = 1, \dots, n \wedge j = 1, \dots, m_i$$

donde i indexa al grupo i -ésimo y m_i al número de integrantes del grupo i

- Además $\mathbb{E}(\varepsilon_{ij}) = 0 \wedge V(\varepsilon_{ij}) = \sigma^2$
- Si observamos sólo los promedios para cada grupo i , es decir sumamos sobre j , tenemos que:

$$\bar{y}_i = \frac{1}{m_i} \sum_j y_{ij}; \quad \bar{x}_i = \frac{1}{m_i} \sum_j x_{ij}; \quad \bar{\varepsilon}_i = \frac{1}{m_i} \sum_j \varepsilon_{ij}$$

- Lo que implica que: $\text{var}(\bar{\varepsilon}_i) = \frac{m_i \sigma^2}{m_i^2} = \frac{\sigma^2}{m_i}$. En este caso particular tendríamos que $\Omega = \text{diag}\left\{\frac{\sigma^2}{m_i}\right\}$.

Heterocedasticidad Multiplicativa: Modelo de coeficientes aleatorios

- **Heterocedasticidad Multiplicativa:** En los años 80 se modeló de la siguiente forma:

$$\begin{aligned}y_i &= x_i' \beta + u_i \text{ donde } u_i = c_i \varepsilon_i \\ \mathbb{E}(\varepsilon_i) &= 0 \wedge V(\varepsilon_i) = \sigma^2 \\ c_i^2 &= h(z_i \theta)\end{aligned}$$

- Tomaremos como ejemplo de referencia el Modelo de Coeficientes Aleatorios

$$\begin{aligned}y_i &= \alpha_i + x_i \beta_i \\ \mathbb{E}(\alpha_i) &= \alpha \wedge V(\alpha_i) = \sigma^2 \\ \mathbb{E}(\beta_i) &= \beta \wedge V(\beta_i) = \Gamma_{k-1 \times k-1}\end{aligned}$$

- En este modelo no hay errores puesto que los coeficientes son variables en cada observación

Modelo de coeficientes aleatorios

- Estos coeficientes atrapan toda la heterogeneidad de las observaciones
- Todas las observaciones se asumen en principio que no tienen elementos comunes (en el caso de MCO asumíamos que había una constante común a todas las observaciones)
- Esto hace, además, que los coeficientes aleatorios tengan correlación con los x_i (que sería por ejemplo en un modelo de educación, donde a mayores habilidades, mayor α_i , implica que estudio mas, mayor x_i).
- Este modelo no lo podemos estimar puesto que tenemos $k \times n$ estimadores.
- Con una pequeña operación matemática obtenemos:

$$y_i = \alpha + x_i\beta + u_i$$

Donde $u_i = (\alpha_i - \alpha) + x_i(\beta_i - \beta)$

Modelo de coeficientes aleatorios

- Tenemos además que $cov(\alpha_i, \beta_i) = \gamma$.
- Supongamos que los x_i son no estocásticos, entonces:

$$\begin{aligned} \mathbb{E}(u_i) &= 0 \\ V(u_i) = \mathbb{E}(u_i^2) &= \underbrace{\mathbb{E}((\alpha_i - \alpha)^2)}_{\sigma^2} + x_i' \underbrace{\mathbb{E}(\beta_i - \beta)(\beta_i - \beta)'}_{\Gamma} x_i + 2cov(x_i(\beta_i - \beta), \alpha_i) \\ V(u_i) &= \sigma^2 + x_i' \Gamma x_i + 2x_i \gamma \end{aligned}$$

- Reparametrizando obtenemos:

$$V(u_i) = \sigma^2(1 + z_i' \theta_i)$$

- Que viene de la definición de h , donde para este caso es una función lineal.
- Heterocedasticidad Exponencial:**

$$\sigma_i^2 = \sigma^2 \exp\{x_i' \theta\}$$

Test de heteroscedasticidad

- Prueba de hipótesis:

$$H_0 : \theta = 0$$

$$H_a : \theta \neq 0$$

- Bajo la nula se cumplen los supuestos de Gauss-Markov ($\mathbb{E}(\varepsilon_i^2) = \sigma^2$). La mayoría de los tests se basan en estimar una regresión auxiliar.

Pasos

1. $\varepsilon_i^2 = \sigma^2 + z_i' \delta + r_i$ donde r_i es un ruido blanco. Este modelo está basado en el modelo de Coeficientes Aleatorios. Se asume que $\mathbb{E}(r_i|z_i) = 0$ y $V(r_i|z_i) = \tau$, es decir, no hay endogeneidad. Dado que ε_i es desconocido, obtendremos $\hat{\varepsilon}_i$ de una estimación MCO.
2. Estime $\hat{\varepsilon}_i^2 = \sigma^2 + z_i' \delta + r_i$ por MCO. Donde z_i contiene a x_i y distintas combinaciones
3. Nos estamos preguntando si x_i y sus combinaciones ayudan a explicar $\hat{\varepsilon}_i$.
4. Podemos hacer esto porque sabemos que dado que $\hat{\beta}_{MCGF} \xrightarrow{p} \beta \Rightarrow \hat{\varepsilon}_i \xrightarrow{p} \varepsilon$ (en el fondo decimos que la heterocedasticidad no daña la convergencia en probabilidad). De la estimación mencionada obtenemos el R^2 .

Breusch-Pagan / White

- Breusch-Pagan proponen el siguiente test:

$$T = nR^2 \xrightarrow{d} \chi^2_{p=\dim\{\theta\}}$$

- Si R^2 es grande, indicaría presencia de heterocedasticidad (donde los z_i estarían explicando en una alta proporción los $\hat{\varepsilon}_i$).
- En este caso se rechazaría la hipótesis nula
White propone muy parecido al test de Breusch y Pagan pero en lugar de z 's los regresores de la regresión auxiliar son x 's además de x^2 y $x_i * x_j$

F test

- Una alternativa es un test F donde mide conjuntamente si los δ de $\varepsilon_i^2 = \sigma^2 + z_i' \delta + r_i$ son distintos de cero:

$$F = (n - k) \frac{R^2/p}{1 - R^2} \xrightarrow{d} F_{p, n-k}$$

- Este test se realiza bajo la nula:

$$\begin{aligned} H_0 : \delta &= 0 \Rightarrow 1 - R^2 \simeq 1 \\ \Rightarrow F &= \frac{n - k}{n} \frac{nR^2/p}{1 - R^2} \simeq nR^2/p \end{aligned}$$

- Bajo la nula, el Test F es similar al Test de Breusch-Pagan/White.
- Esto explota el hecho que cuando $n - k$ crece, tenemos que $\chi_p^2 \simeq p \times F_{p, n-k}$.
- Ejemplo: si $F(2, 71) = 2.05$, el valor correspondiente χ_2^2 es $2 \times 2.05 = 4.1$ y note que las tail-probabilities son muy parecidas:
 - $F(2, 71) = 2.05 \Rightarrow p = .1363$
 - $\chi_2^2 = 4.1 \Rightarrow p = .1287$
- Con n grandes no deberían haber diferencias entre los resultados de los tests

LM test

- Este ultimo lo obtenemos si asumimos normalidad
- Definimos el estadístico S

$$S = \frac{SSE}{2\hat{\sigma}^4}$$

Donde:

$$- ESS = \sum_{i=1}^n [(z_i - \bar{z})' \hat{\delta}]^2$$

– $2\hat{\sigma}^4$ es el cuarto momento de ε_i bajo una distribución normal estándar.

- Y se puede demostrar:

$$S \xrightarrow{d} \chi_{dim\{\theta\}}^2$$

- Al asumir normalidad tenemos también que el cuarto momento se relaciona con el segundo momento (al cuadrado)

$$\tau = var(\varepsilon_i^2) = 2 [var(\varepsilon_i)]^2 = 2\sigma^4$$

- Se puede demostrar entonces que “S” es equivalente al estadístico de Breuch-Pagan.
- Si no asumimos normalidad simplemente cambiamos σ^4 por $\hat{\tau}$

¿Qué hacemos cuando rechazamos la Nula?

1. Estimamos por MCO y obtenemos la matriz de varianzas y covarianzas robusta (Eicker-Huber-White)
2. Estimamos por métodos robustos a la heteroscedasticidad: MCG/MCGF
 - Cuando solo tenemos heteroscedasticidad (la matriz de varianzas sigue siendo diagonal) se le llama Mínimos Cuadrados Ponderados Factibles ya que:

$$\Omega = \begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{pmatrix}; H^{-1} = \begin{pmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sqrt{\omega_n} \end{pmatrix}$$

Es como si ponderáramos nuestras observaciones por la inversa de su varianza

- 2.1 Si sabemos la forma de la heteroscedasticidad: estimamos los parámetros de la estructura de la heteroscedasticidad. Usamos eso para estimar por MCGF/MCGP

Como Aplicamos FGLS/MCGF con Hetersocedasticidad conocida?

1. Heterocedasticidad Multiplicativa

- $\mathbb{E}(\varepsilon_i^2 | Z_i) = \sigma_i^2 = \sigma^2 h(Z_i' \theta)$
- Pasos:
 - Encontrar una manera de estimar " θ ":
 - ⇒ Estime $\hat{\varepsilon}_{MCO}^2 = \hat{\sigma}^2 h(Z_i' \hat{\theta})$ (puede ser por MCNL)
 - H se asume conocida
 - Reemplazamos $\hat{\Omega} = \text{diag} \{ h(Z_i' \hat{\theta}) \}$ en $\hat{\beta}_{MCGF} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} Y$
- Equivalentemente, si queremos expresarlo como mínimos cuadrados ponderados factibles (MCPF), podemos transformar el modelo:

$$y^* = \frac{y_i}{\sqrt{h(Z_i' \hat{\theta})}} \text{ y } x^* = \frac{x_i}{\sqrt{h(Z_i' \hat{\theta})}}$$

- y aplicar MCO a este modelo transformado.

Heteroscedasticidad multiplicativa: Modelo de Coeficientes Aleatorios

1.1 Modelo de Coeficientes Aleatorios

- Aca la estimación por MCO implica que necesariamente el término de error será heterocedástico con una forma funcional conocida:

$$\text{var}(\varepsilon_i|X_i) = \sigma^2 + 2X_i'\gamma + X_i'\Gamma X_i$$

- La expresión anterior se puede reparametrizar, escribiéndola como:

$$\begin{aligned}\text{var}(\varepsilon_i|X_i) &= \sigma^2 + \sigma^2 \left(\frac{2}{\sigma^2} X_i'\gamma + \frac{1}{\sigma^2} X_i'\Gamma X_i \right) \\ &= \sigma^2 \left(1 + \frac{2}{\sigma^2} X_i'\gamma + \frac{1}{\sigma^2} X_i'\Gamma X_i \right) \\ &= \sigma^2(1 + Z_i'\theta)\end{aligned}$$

Modelo Coeficientes Aleatorios: Pasos

1. Estimamos $\hat{\varepsilon}_i^2 = \sigma^2 + Z_i' \delta + r_i$, con $\mathbb{E}(r_i | Z_i) = 0$, donde $\delta = \sigma^2 \theta$
2. El intercepto nos dará un estimador para σ^2 lo que implica que *theta* está identificado y así, tenemos $\hat{\theta}$ y $\hat{\sigma}^2$.
3. Estimamos por MCGF o MCPF
 - Notar que podría ocurrir algo muy grave: que el estimador de σ^2 tenga valor negativo.
 - Sin embargo, dado que la variable dependiente es definida en los reales no negativos (al ser una función cuadrática), el intercepto no puede ser negativo, por construcción.

Heteroscedasticidad Multiplicativa

1.2 Heteroscedasticidad Exponencial

- Aplicando logaritmo y suponiendo independencia entre μ_i y X :

$$\varepsilon_i = \mu_i^2 \exp\{X_i' \theta\}$$

$$\ln \varepsilon_i = 2 \ln \mu_i + X_i' \theta$$

- ¿Qué ocurre si μ_i es independiente de X_i ? Apliquemos esperanza:

$$\begin{aligned}\mathbb{E}[\ln(\varepsilon_i)^2 | X] &= \mathbb{E}[2 \ln \mu_i | X] + X_i' \theta \\ &= \mathbb{E}[2 \ln \mu_i] + X_i' \theta \\ &= \alpha + X_i' \theta\end{aligned}$$

- Recordamos que $\mathbb{E}[g(x)] = \int g(x)f(x)dx$. Para nuestro caso :

$$\alpha = 2 \int \ln \mu f(\mu) d\mu$$

- Estimamos la regresión:

$$\ln(\hat{\varepsilon}_i^2) = \alpha + X_i' \theta + \epsilon_i$$

- Después estimamos por MCGF/MCPF

Matriz de Varianzas y Covarianzas Consistentes

- $\hat{\beta}_{MCGF} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$
- $var(\hat{\beta}_{MCGF}) = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}var(Y)\hat{\Omega}^{-1}X(X'\hat{\Omega}^{-1}X)^{-1}$
- Supongamos que no sabemos la forma de la heteroscedasticidad :
 $\sigma_i^2 \neq \sigma^2 h(z_i'\theta)$
- Definimos: $\Sigma = \hat{\Omega}^{-1}var(Y)\hat{\Omega}^{-1}$
- Entonces: $var(\hat{\beta}_{MCGF}) = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Sigma}X(X'\hat{\Omega}^{-1}X)^{-1}$
- Sabemos que tanto Ω como $Var(Y)$ son diagonales. Entonces,

$$\Sigma = \hat{\Omega}^{-1}var(Y)\hat{\Omega}^{-1} = diag \left\{ \frac{1}{h(x_i'\hat{\theta})} \right\} diag\{\sigma_i^2\} diag \left\{ \frac{1}{h(z_i'\hat{\theta})} \right\}$$

- Pero aun desconocemos σ_i

Matriz de Varianzas y Covarianzas Consistentes

- Reemplazamos σ_i^2 por el error estimado.
- ¿Qué error estimado deberíamos usar? El de MCO lo usamos para estimar θ
- Ahora debemos usar el de **MCGF**

$$\hat{\varepsilon}_i = y_i - x_i' \hat{\beta}_{MCGF}$$

- Obtenemos una estimación de Σ ,

$$\hat{\Sigma} = \text{diag} \left\{ \frac{\hat{\varepsilon}_i^2}{h(z_i' \theta)^2} \right\}$$

- Asintóticamente:

$$\sqrt{n} \left(\hat{\beta}_{MCGF} - \beta \right) \xrightarrow{d} N(0, A^{-1} B A)$$

- Donde $A = p \lim \frac{1}{n} (X' \hat{\Omega}^{-1} X)^{-1}$ y $B = p \lim \frac{1}{n} (X' \Sigma X)$

Matriz de Varianzas y Covarianzas Consistentes

- Sólo cuando $\sigma_i^2 = \sigma^2 h(z_i' \theta)$ (no hay error de especificación en la estructura de la heterocedasticidad), tenemos que

$$\Sigma = \sigma^2 \Omega^{-1}$$

- Y de esta manera:

$$\sqrt{n} \left(\hat{\beta}_{MCGF} - \beta \right) \xrightarrow{d} N(0, A^{-1})$$