Tema 2

Mínimos Cuadrados Generalizados

2.1. Modelo de regresión con perturbaciones no esféricas

En el tema de Mínimos Cuadrados Generalizados vamos a relajar dos de las hipótesis básicas sobre la perturbación. Hasta ahora hemos supuesto que la perturbación era homocedástica, $Var(u_t) = \sigma^2 \quad \forall t$, y no autocorrelada, $Cov(u_t, u_s) = 0 \quad \forall t, s \;\; t \neq s$, lo que en la literatura econométrica se conoce como perturbaciones esféricas. En este tema vamos a relajar estas dos hipótesis básicas, permitiremos que exista heterocedasticidad y/o autocorrelación. Escribimos estas hipótesis como:

- heterocedasticidad: $Var(u_t) = \sigma_t^2$
- autocorrelación: $Cov(u_t, u_s) \neq 0, \quad \forall t, s \;\; t \neq s$

Para el propósito de estimación distinguir entre heterocedasticidad y/o autocorrelación no es necesario ya que en ambos casos el modelo se estima de la misma manera, por ello en este tema presentaremos el estimador Mínimo Cuadrático Generalizado y sus propiedades, comunes a ambos casos. En los dos temas siguientes veremos los problemas de heterocedasticidad y autocorrelación por separado particularizando en cada uno de ellos.

Nuestro marco de trabajo queda definido como sigue:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t \quad t = 1, 2, \ldots, T \qquad \Longleftrightarrow \qquad Y_t = X_t' \beta + u_t$$

mantenemos que los regresores $\mathbf{X}_t' = [1 \quad \mathbf{X}_{2t} \; \ldots \; \mathbf{X}_{Kt}]$ son no estocásticos y sobre las perturbaciones suponemos:

$E(u_t) = 0 \quad \forall t$, media cero
$E(u_t^2) = \sigma_t^2 \quad t = 1, 2, \ldots, T$, varianza no constante y/o
$E(u_t u_s) \neq 0 \;\; \forall t, s \quad t \neq s$, covarianzas no nulas,

En términos matriciales el modelo se escribe:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K)}{X} \; \underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{u}$$

y la estructura de **la matriz de varianzas y covarianzas del vector de perturbaciones** que recoge los supuestos anteriores sobre u sería una matriz no escalar como la siguiente:

$$E(uu') = \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

$$= \quad \sigma^2\Omega = \quad \sigma^2 \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1T} \\ w_{21} & w_{22} & \cdots & w_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ w_{T1} & w_{T2} & \cdots & w_{TT} \end{bmatrix}$$

donde

$$\begin{aligned} Var(u_t) &= \sigma_t^2 = \sigma^2 w_{tt}, \quad t = 1, ..., T \\ Cov(u_t, u_s) &= \sigma_{ts} = \sigma_{st} \; = \; \sigma^2 w_{ts}, \quad t \neq s \end{aligned}$$

Es decir, la varianza de la perturbación puede cambiar en cada momento y puede existir autocorrelación entre perturbaciones de distintos momentos del tiempo. Además, podemos suponer que existe un factor de escala común a todos los elementos, el parámetro σ^2, que perfectamente puede tomar el valor de la unidad.

Vamos a empezar viendo ejemplos en los que la hipótesis de perturbaciones esféricas no se cumple:

Ejemplo 2.1 Supongamos una muestra de observaciones de sección cruzada relativas a gastos de consumo familiares, C, y renta disponible, R, de un colectivo de N familias. La perturbación mide la diferencia entre el consumo de una familia y el consumo medio de todas las familias que poseen la misma renta, $u_i = C_i - E(C_i/R_i)$, y σ^2 mide la dispersión de estas observaciones. En familias con rentas bajas, las posibilidades de consumo están restringidas por la renta. Sin embargo, a medida que aumenta la renta se amplían las posibilidades y podrán decidir cuanto consumir y cuanto ahorrar. Así, podemos encontrarnos con familias de rentas altas ahorrativas, con bajo consumo, y otras con alto consumo y poco ahorro. En este caso hay una gran dispersión del consumo y σ^2 será grande mientras que para las rentas bajas σ^2 será pequeña ya que tienen menos posibilidades de variar su consumo dad su renta. En este supuesto la varianza de la perturbación cambia según la renta de las familias, existe **heterocedasticidad** y podemos escribirla por ejemplo como $E(u_i^2) = \sigma^2 R_i$.

En el caso de datos de sección cruzada, y si suponemos que no existe correlación entre perturbaciones de distintos individuos, la matriz de varianzas y covarianzas de la perturbación sería:

$$E(uu') = \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}$$

donde suponemos $E(u_i) = 0 \; \forall i \quad i = 1, \ldots, N, \quad E(u_i^2) = \sigma_i^2$, y $E(u_i u_j) = 0 \quad \forall i, j \;\; i \neq j$.

Ejemplo 2.2 Supongamos que analizamos la función de formación de precios de una acción de un banco con fuerte participación en el negocio de la construcción en los Emiratos Árabes. Creemos que los precios de las acciones dependen de las reservas del banco, de su volumen de negocio y de los beneficios obtenidos, entre otros factores que afectan en la media del nivel de precios. Una quiebra no esperada del negocio en dichos Emiratos afectará negativamente a sus resultados y a la confianza que los agentes del mercado tienen depositada en el banco. Este shock no predecible será recogido por la perturbación del modelo, u_t. Resulta razonable pensar que el shock no va a afectar únicamente a un periodo t, si no que los agentes recuperaran la confianza lentamente por lo que parte de ese shock se mantendrá en el futuro. Si esto es así, lo que ocurre en un período actual depende de lo ocurrido en el periodo pasado y será difícil mantener $E(u_t u_{t-1}) = 0$, es decir, que las perturbaciones de los momentos t y $t-1$ estén incorrelacionadas. Ocurrirá lo contrario, que las perturbaciones están correlacionadas. Para el caso general podemos recoger la existencia de autocorrelación denotándola como :

$$E(u_t u_s) \neq 0 \quad \forall t, s \quad t \neq s$$

En este caso, para datos de serie temporal, y suponiendo que la varianza es constante, escribimos la matriz de varianzas y covarianzas de la perturbación bajo el supuesto de **autocorrelación** como:

$$E(uu') = \Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma^2 \end{bmatrix}$$

Ejemplo 2.3 Supongamos que queremos estimar la demanda de automóviles Y como una función de la renta X, utilizando datos microeconómicos sobre gastos de las familias en dos núcleos geográficos distintos, núcleo urbano y núcleo rural. La función de demanda para las familias del núcleo urbano es:

$$Y_{1i} = \alpha_1 + \beta_1 X_{1i} + u_{1i} \qquad i = 1, \ldots, N_1 \qquad u_{1i} \sim N(0, \sigma_1^2 I_{N_1})$$

La función de demanda para las familias del núcleo rural es:

$$Y_{2j} = \alpha_2 + \beta_2 X_{2j} + u_{2j} \qquad j = 1, \ldots, N_2 \qquad u_{2j} \sim N(0, \sigma_2^2 I_{N_2})$$

Supongamos que la propensión marginal a consumir de ambos núcleos es la misma, $\beta_1 = \beta_2$. En este caso deberíamos estimar la función de demanda en el siguiente modelo conjunto:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} i_{N_1} & 0 & X_1 \\ 0 & i_{N_2} & X_2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \Longleftrightarrow \underset{((N_1+N_2)\times 1)}{Y} = \underset{((N_1+N_2)\times 3)}{X} \underset{(3\times 1)}{\beta} + \underset{((N_1+N_2)\times 1)}{u}$$

Suponiendo ambas muestras independientes, el vector de medias y la matriz de varianzas y covarianzas del vector de perturbaciones del sistema de ecuaciones a estimar es:

$$E(u) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}_{(N_1+N_2)\times 1} \qquad E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

Puede ocurrir que la variación en el gasto en automóviles dada una renta, sea mayor en el núcleo urbano que en el rural, esto es $\sigma_1^2 > \sigma_2^2$, en este caso, en el modelo a estimar la perturbación es **heterocedástica** ya que $\sigma_1^2 \neq \sigma_2^2$.

Además, las muestras no tienen porque ser independientes podría ocurrir que aún en el caso de que la variación en el gasto de automóviles fuese la misma en la zona urbana que en la rural, hubiese factores que siendo no significativos en media en las respectivas demandas implicasen dependencia entre u_1 y u_2, tal que $\sigma_{12} \neq 0$. En este caso las ecuaciones estarían relacionadas a través del término de perturbación y $E(uu') = \Sigma \neq I$ tal que[1].

$$E(uu') = \begin{bmatrix} \sigma^2 I_{N_1} & \sigma_{12} I_N \\ \sigma_{12} I_N & \sigma^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

Como se puede observar hay muchas y diversas situaciones en que pueden aparecer los problemas de heterocedasticidad y/o autocorrelación. En general la heterocedasticidad aparece con mayor probabilidad en datos de sección cruzada mientras que la autocorrelación es más fácil encontrarla con datos de serie temporal. Dado que la existencia de heterocedasticidad y/o autocorrelación implica que algunas de las hipótesis básicas utilizadas para derivar las propiedades del estimador MCO no se cumplen, debemos empezar viendo que implicaciones tiene en el estimador de MCO y sus propiedades el hecho de que la matriz de varianzas y covarianzas de la perturbación no sea ya una matriz escalar.

2.2. Propiedades del estimador de MCO

Sea el Modelo de Regresión Lineal General, $Y = X\beta + u$, donde se mantienen las hipótesis básicas salvo que[2]:

$$E(uu') = \Sigma = \sigma^2\Omega, \text{ donde } \Omega \neq I \tag{2.1}$$

El estimador MCO de β, $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$, en muestras finitas sigue siendo lineal e insesgado, pero no es varianza mínima. En muestras grandes o asintóticas es consistente. Prueba:

- **Lineal:** dado que X es no estocástica el estimador MCO es lineal en u, ya que se puede escribir como una combinación lineal del vector de perturbaciones u y la matriz de constantes $D = (X'X)^{-1}X'$, siendo u lo único aleatorio de su expresión:

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

- **Insesgado:** dado que X es no estocástica y $E(u) = 0$ el estimador MCO es insesgado.

$$E(\hat{\beta}_{MCO}) = \beta + E[\,(X'X)^{-1}X'u\,] = \beta + (X'X)^{-1}X'\,E(u) = \beta$$

[1]Estas no son las únicas posibilidades. Por ejemplo podemos pensar en $\sigma_1^2 > \sigma_2^2$ y $\sigma_{12} \neq 0$ o cualquier otra combinación de posibilidades.

[2]Para el propósito de la demostración es indiferente que en la matriz de varianzas y covarianzas exista o no factor de escala distinto de la unidad. En lo que sigue supondremos que $E(uu') = \sigma^2\Omega$.

- **Matriz de varianzas y covarianzas:**

$$\begin{aligned} V(\hat{\beta}_{MCO}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\ &= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'\,E(uu')\,X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \end{aligned}$$

  Bajo heterocedasticidad y/o autocorrelación el estimador MCO **no** es el estimador con varianza mínima entre los estimadores lineales e insesgados. Si $\Omega \neq I$ veremos que existe un estimador que alcanza la cota de varianza mínima, y que es el eficiente entre los estimadores lineales e insesgados en esta situación.

- **Distribución en muestras finitas[3]:** Si las perturbaciones tienen una distribución normal $u \sim N(0,\, \sigma^2\Omega)$

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

- **Consistente:** El estimador MCO bajo perturbaciones no esféricas es consistente. Vamos a demostrar la consistencia del estimador utilizando las condiciones suficientes de consistencia. Sean:

$$\begin{aligned} \lim_{T\to\infty} \frac{X'X}{T} &= Q \quad \text{finita, definida positiva} \\ \lim_{T\to\infty} \frac{X'\Omega X}{T} &= Z \quad \text{finita, definida positiva} \end{aligned}$$

  Entonces:

$$\begin{aligned} \lim_{T\to\infty} E(\hat{\beta}_{MCO}) &= \beta \\ \lim_{T\to\infty} V(\hat{\beta}_{MCO}) &= \lim_{T\to\infty} \frac{\sigma^2}{T}\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'\Omega X}{T}\right)\left(\frac{X'X}{T}\right)^{-1} = \\ &= 0 \cdot Q^{-1} \cdot Z \cdot Q^{-1} = 0, \end{aligned}$$

  por lo que el estimador es consistente: $\operatorname{plim} \hat{\beta}_{MCO} = \beta$.

2.2.1. Estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$

El estimador MCO de β es lineal e insesgado en muestras finitas. Su matriz de varianzas y covarianzas se define $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Un estimador de la matriz de varianzas y covarianzas del estimador MCO es $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ siendo $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K}$. Utilizar este estimador para hacer inferencia cuando la matriz de varianzas y covarianzas de la perturbación del modelo no es escalar, $E(uu') = \sigma^2\Omega,\ \Omega \neq I$, no es adecuado. Los estadísticos t y F habituales para

[3] Es importante recordar que $\hat{\beta}_{MCO}$ es lineal en u. Esto implica que su distribución en muestras finitas coincide con la distribución de u, independientemente de cuál sea ésta. En nuestro caso como suponemos a u normal el estimador MCO tendrá distribución normal.

hacer inferencia sobre β definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ son inapropiados ya que $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$ es un estimador sesgado e inconsistente. Dichos estadísticos ahora no tienen las distribuciones t-Student y F-Snedecor habituales por lo tanto, la inferencia en base a ellos no es válida.

Por todo ello parece más adecuado buscar un nuevo estimador del vector de coeficientes β que tuviera en cuenta que $E(uu') = \sigma^2\Omega$, con Ω definida bien para el caso de heterocedasticidad, bien para el caso de autocorrelación, bien para ambos, y que fuese lineal, insesgado, de varianza mínima y consistente. Este estimador es el de Mínimos Cuadrados Generalizados y a su vez permite proponer un estimador insesgado para σ^2 y realizar inferencia válida. El único requisito para poder obtenerlo es que Ω sea conocida. En el caso de que no exista factor de escala $\sigma^2\Omega = \Sigma$ y requerimos que Σ sea conocida.

2.3. Método de Mínimos Cuadrados Generalizados (MCG)

Supongamos el MRLG $Y = X\beta + u$ donde $E(u) = 0$ y $E(uu') = \sigma^2\Omega$, siendo Ω conocida y se cumplen el resto de hipótesis básicas. Si lo que queremos es estimar los coeficientes β desconocidos de forma que el estimador sea eficiente, una manera adecuada de proceder es transformar el modelo tal que sus perturbaciones sean esféricas, es decir, de media cero, varianza constante y covarianzas cero. El estimador de MCO en el modelo así transformado será lineal, insesgado, de varianza mínima o eficiente y consistente. Veámoslo.

Dado que $\Sigma = \sigma^2\Omega$ es simétrica y semidefinida positiva, y hemos supuesto que Ω es conocida, existe una matriz no singular P tal que $\Omega = PP'$. Además, P es conocida y no estocástica. La inversa de la matriz P se utiliza como matriz de transformación del modelo original dado que:

$$\begin{aligned}
\Omega &= PP' \\
\Omega^{-1} &= (PP')^{-1} = (P')^{-1}P^{-1} \\
P^{-1}\Omega(P')^{-1} &= P^{-1}PP'(P')^{-1} = I
\end{aligned}$$

Premultiplicando el modelo por P^{-1} obtenemos el siguiente modelo transformado:

$$\underbrace{P^{-1}Y}_{Y^\star} = \underbrace{P^{-1}X}_{X^\star}\beta + \underbrace{P^{-1}u}_{u^\star} \Longleftrightarrow Y^\star = X^\star\beta + u^\star$$

Este modelo transformado tiene perturbaciones, $u^\star$, esféricas, es decir, de media cero, varianza constante y covarianzas cero:

$$\begin{aligned}
E(u^\star) &= E(P^{-1}u) = P^{-1}E(u) = 0 \\
E(u^\star u^{\star\prime}) &= E(P^{-1}uu'(P^{-1})') = P^{-1}E(uu')(P^{-1})' = \\
&= \sigma^2 P^{-1}\,\Omega\,P^{-1\prime} = \sigma^2 P^{-1}PP'(P')^{-1} = \sigma^2 I.
\end{aligned}$$

Por lo tanto, en el modelo transformado se cumplen las hipótesis básicas, y el estimador MCO tendrá las propiedades habituales de linealidad, insesgadez y varianza mínima. Si definimos el estimador de MCO en el modelo transformado, y sustituimos las matrices transformadas por su expresión en términos de las variables originales del modelo, obtenemos la expresión del estimador de Mínimos Cuadrados Generalizados, MCG, en el modelo de interés. Así:

$$\begin{aligned}
\hat{\beta}_{MCO} &= (X^{\star'} X^{\star})^{-1} X^{\star'} Y^{\star} = & (2.2)\\
&= (X^{'}(P')^{-1} P^{-1} X)^{-1} X^{'}(P')^{-1} P^{-1} Y = \\
&= (X^{'}\Omega^{-1} X)^{-1} X^{'}\Omega^{-1} Y = \hat{\beta}_{MCG} & (2.3)
\end{aligned}$$

Si Ω (ó Σ en su caso) es conocida el estimador es inmediatamente calculable.

• **Criterio de estimación y función objetivo:**
El estimador de MCG dado por la expresión (2.3) puede ser obtenido minimizando la siguiente función objetivo:

$$\underset{\hat{\beta}}{Min}\; \hat{u}^{'}\,\Omega^{-1}\,\hat{u} = \underset{\hat{\beta}}{Min}\,(Y - X\hat{\beta})'\,\Omega^{-1}\,(Y - X\hat{\beta}) = \underset{\hat{\beta}}{Min}\,(Y^{\star} - X^{\star}\hat{\beta})'(Y^{\star} - X^{\star}\hat{\beta})$$

Utilizando las condiciones de primer orden podemos derivar el sistema de ecuaciones normales $X^{'}\Omega^{-1}Y = \hat{\beta}(X^{'}\Omega^{-1}X)^{-1}$ de donde obtenemos $\hat{\beta}_{MCG} = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}Y$.

2.3.1. Propiedades del estimador de MCG

Como ya se ha dicho el estimador de MCG es un estimador lineal, insesgado y de varianza mínima en muestras finitas, en muestras grandes es consistente y eficiente asintóticamente. Las propiedades del estimador MCG podemos demostrarlas alternativamente en el modelo transformado utilizando la expresión (2.2) o en el modelo original utilizando la expresión (2.3).

- **Lineal:** dado que X y Ω son no estocásticas el estimador MCG es lineal en la perturbación u, ya que se puede expresar como una combinación lineal del vector de perturbaciones u y la matriz de constantes no estocástica $C = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}$, siendo u lo único aleatorio de su expresión:

$$\hat{\beta}_{MCG} = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}Y = \beta + (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}u = \beta + C\,u$$

- **Insesgado:** Dado que X y Ω son no estocásticas y $E(u) = 0$ el estimador MCG es insesgado:

$$\begin{aligned}
E(\hat{\beta}_{MCG}) &= \beta + E[\,(X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}u\,] = \\
&= \beta + (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}E(u) = \beta
\end{aligned}$$

- **Matriz de varianzas y covarianzas:**

$$\begin{aligned}
V(\hat{\beta}_{MCG}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] \\
&= E[(X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}\,uu'\,\Omega^{-1}X(X^{'}\Omega^{-1}X)^{-1}] \\
&= (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}\,E(uu')\,\Omega^{-1}X(X^{'}\Omega^{-1}X)^{-1} \\
&= \sigma^2(X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}\,\Omega\,\Omega^{-1}X(X^{'}\Omega^{-1}X)^{-1} \\
&= \sigma^2(X^{'}\Omega^{-1}X)^{-1}
\end{aligned}$$

Cuando $E(uu') = \sigma^2\Omega \neq I$ y Ω es conocida, el estimador MCG de β es el de menor varianza entre los estimadores lineales e insesgados.

- **Distribución en muestras finitas:** Bajo el supuesto de normalidad de las perturbaciones, $u \sim N(0,\, \sigma^2\Omega)$, tenemos que

$$\hat{\beta}_{MCG} \sim N(\beta,\, \sigma^2(X'\Omega^{-1}X)^{-1})$$

- **Consistente:** Podemos demostrar la consistencia del estimador mediante las condiciones suficientes de consistencia:
Sea $\operatorname{plim} \dfrac{X'\Omega^{-1}X}{T} = G$, finita, semidefinida positiva y no singular, entonces

$$\begin{aligned} \lim_{T\to\infty} E(\hat{\beta}_{MCG}) &= \beta \\ \lim_{T\to\infty} \mathrm{V}(\hat{\beta}_{MCG}) &= \lim_{T\to\infty} \frac{\sigma^2}{T}\left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} = 0 \times G^{-1} = 0 \end{aligned}$$

luego $\operatorname{plim} \hat{\beta}_{MCG} = \beta$, con lo que el estimador es consistente.

- **Distribución Asintótica:** Dado que

$$\operatorname{plim} \hat{\beta}_{MCG} = \beta \Longrightarrow \hat{\beta}_{MCG} \xrightarrow{p} \beta \Longrightarrow (\hat{\beta}_{MCG} - \beta) \xrightarrow{p} 0$$

el estimador tiene una distribución degenerada en el límite, por lo que buscamos la distribución asintótica para $\sqrt{T}(\hat{\beta}_{MCG} - \beta)$

$$\sqrt{T}(\hat{\beta}_{MCG} - \beta) = \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \frac{X'\Omega^{-1}u}{\sqrt{T}}$$

En el modelo original el teorema de Mann-Wald no se puede aplicar, por lo que demostrar la consistencia es un poco más costoso que si lo hacemos en el modelo transformado. En el modelo transformado las perturbaciones son esféricas y $X^\star$ es no estocástica por lo que podemos aplicar el Teorema de Mann-Wald. Vamos a mostrar como utilizarlo en la obtención de la distribución asintótica de $\hat{\beta}_{MCG}$.

i) Sea $u^\star \sim iid(0, \sigma^2 I)$

ii) Sea $E(X^{\star\prime} u^\star) = X^{\star\prime} E(u^\star) = 0$ ya que $X^\star$ es no estocástica,

iii) Sea $\operatorname{plim} \left(\dfrac{X^{\star\prime} X^\star}{T}\right) = Q^\star = G$ finita y no singular

Entonces se cumplen los dos resultados siguientes:

1. $\operatorname{plim} \left(\dfrac{X^{\star\prime} u^\star}{T}\right) = 0$
2. $\dfrac{X^{\star\prime} u^\star}{\sqrt{T}} \xrightarrow{d} N(0,\, \sigma^2 G)$

por lo que de Mann-Wald, y utilizando el teorema de Cramer, tenemos

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{X^{\star\prime} X^{\star}}{T}\right)^{-1} \frac{X^{\star\prime} u^{\star}}{\sqrt{T}} \quad \xrightarrow{d} \quad N(0,\, \sigma^2 G^{-1})$$

de donde

$$\sqrt{T}(\hat{\beta}_{MCG} - \beta) = \left(\frac{X'\Omega^{-1}X}{T}\right)^{-1} \frac{X'\Omega^{-1}u}{\sqrt{T}} \xrightarrow{d} N(0,\, \sigma^2 G^{-1})$$

con $G = \text{plim}\, \frac{X'\Omega^{-1}X}{T}$.

2.3.2. Estimador de la matriz de varianzas y covarianzas de $\hat{\boldsymbol{\beta}}_{\boldsymbol{MCG}}$

El estimador de MCG es lineal, insesgado y de varianza mínima. Cuando $E(uu') = \sigma^2\Omega$, a pesar de que Ω sea conocida, en general σ^2 será desconocida y su matriz de varianzas y covarianzas habrá de ser estimada. Un estimador de $V(\hat{\beta}_{MCG})$ insesgado y consistente es

$$\widehat{V}(\hat{\beta}_{MCG}) = \hat{\sigma}^2_{MCG}(X'\Omega^{-1}X)^{-1} = \hat{\sigma}^2_{MCG}(X^{\star\prime}X^{\star})^{-1}$$

siendo

$$\hat{\sigma}^2_{MCG} = \frac{\hat{u}'_{MCG}\,\Omega^{-1}\,\hat{u}_{MCG}}{T-K} = \frac{\hat{u}^{\star\prime}\hat{u}^{\star}}{T-K}$$

Puede resultar útil tener las expresiones anteriores en términos de las variables del modelo tal que:

$$\begin{aligned}
\hat{\sigma}^2_{MCG} &= \frac{\hat{u}^{\star\prime}\hat{u}^{\star}}{T-K} = \frac{(Y^\star - X^\star\hat{\beta}_{MCG})'(Y^\star - X^\star\hat{\beta}_{MCG})}{T-K} = \frac{Y^{\star\prime}Y^\star - \hat{\beta}'_{MCG}X^{\star\prime}Y^\star}{T-K} \\
\hat{\sigma}^2_{MCG} &= \frac{\hat{u}'_{MCG}\,\Omega^{-1}\,\hat{u}_{MCG}}{T-K} = \\
&= \frac{(Y - X\hat{\beta}_{MCG})'\Omega^{-1}(Y - X\hat{\beta}_{MCG})}{T-K} = \frac{Y'\Omega^{-1}Y - \hat{\beta}'_{MCG}X'\Omega^{-1}Y}{T-K}
\end{aligned}$$

Ejercicio 2.1 En el modelo $Y = X\beta + u$ donde $E(u) = 0,\ E(uu') = \sigma^2\Omega$ con Ω conocida y X es no estocástica. Demostrar la consistencia del estimador MCG utilizando el modelo transformado y el Teorema de Mann-Wald.

Ejercicio 2.2 En el modelo $Y = X\beta + u$ donde $E(u) = 0, \quad E(uu') = \Sigma,\ \Sigma$ conocida y X es no estocástica. Se pide:

1. Escribir la función objetivo. Obtener las ecuaciones normales y derivar el estimador MCG.
2. Demostrar sus propiedades en muestras finitas y asintóticas.
3. Obtener su distribución en muestras finitas y asintóticas.

Ejercicio 2.3 En el modelo $Y_i = \beta_1 + \beta_2 X_i + u_i \quad i = 1, \ldots, 20$ donde X es no estocástica, $E(u_i) = 0 \, \forall i$, $Var(u_i) = \sigma^2 X_i$, y $Cov(u_i u_j) = 0 \quad \forall i \neq j$. Se pide:

1. Escribir la función objetivo.
2. Obtener las ecuaciones normales y derivar el estimador MCG de β_1 y β_2.

2.4. Método de Mínimos Cuadrados Generalizados Factibles (MCGF)

Hasta ahora hemos supuesto que conocíamos Ω o Σ en $E(uu') = \sigma^2 \Omega = \Sigma$, pero en la práctica la mayoría de las veces Ω (o Σ) son desconocidas. En este caso el estimador MCG no es directamente calculable ya que en su expresión aparecen estas matrices. La solución habitual es sustituir Ω (o Σ) por una estimación suya en la expresión del estimador de MCG. Este es el estimador de Mínimos Cuadrados Generalizados Factibles, MCGF:

$$\hat{\beta}_{MCGF} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}Y \tag{2.4}$$

Notar que sustituir Ω por una estimación suya implica hacer un supuesto de comportamiento, es decir, suponer una forma funcional para los elementos de Ω que son desconocidos. Una vez supuesta una forma funcional, que en general dependerá de un conjunto de parámetros desconocidos y variables exógenas, es necesario estimar esos parámetros desconocidos[4]. Cómo proponer una forma funcional para los elementos de Ω y cómo estimar los parámetros desconocidos de esta forma funcional es algo que se verá en los temas siguientes, con cada caso en concreto, heterocedasticidad y/o autocorrelación.

Ahora vamos a limitarnos a estudiar las propiedades del estimador MCGF una vez supuesto que hemos sido capaces de proponer una forma funcional sensata a la información disponible sobre el comportamiento de las varianzas y covarianzas del vector de perturbaciones. Además, hemos sido capaces de estimar de forma consistente los parámetros desconocidos de los que dependen estos momentos. Es decir, disponemos de $\widehat{\Omega}$ estimador consistente de Ω.

2.4.1. Propiedades del estimador de MCGF

Bajo el supuesto de que las varianzas de las perturbaciones se han modelado correctamente, el estimador MCGF en muestras finitas es un estimador no lineal y sesgado en general, su distribución en muestras finitas no es conocida. En muestras grandes, bajo ciertas condiciones de regularidad, y si $\widehat{\Omega}$ es un estimador consistente de Ω, el estimador MCGF es consistente, asintóticamente eficiente y tiene distribución asintótica conocida.

- **Linealidad:** El estimador MCGF no es lineal en u.

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

[4] El número de parámetros desconocidos es limitado ya que en otro caso no se podrían estimar los K coeficientes β del modelo y los $T(T+1)/T$ elementos desconocidos en la matriz de varianzas y covarianzas de la perturbación con T observaciones.

En la expresión del estimador MCGF aparecen u y $\widehat{\Omega}$, ambas son variables aleatorias, ya que $\widehat{\Omega}$ que ahora es un estimador, luego es una matriz estocástica. Por tanto, $\hat{\beta}_{MCGF}$ es una función no lineal de u y $\widehat{\Omega}$.

- **Insesgadez y matriz de varianzas y covarianzas:** En general el estimador de MCGF es sesgado:

$$E(\hat{\beta}_{MCGF}) = \beta + E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)] \neq \beta$$

Para determinar $E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)]$ necesitamos conocer la distribución conjunta de las variables aleatorias en $\widehat{\Omega}$ y en u. En general $E[(X'\widehat{\Omega}^{-1}X)^{-1}(X'\widehat{\Omega}^{-1}u)] \neq 0$ y por tanto $E(\hat{\beta}_{MCGF}) \neq \beta$. Además, $V(\hat{\beta}_{MCGF}) = E[\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF})][\hat{\beta}_{MCGF} - E(\hat{\beta}_{MCGF})]'$ es una expresión difícil de obtener analíticamente. Las propiedades en muestras finitas y la distribución exacta del estimador MCGF son desconocidas. Por lo tanto, en muestras finitas su comportamiento es difícil de comparar con el de MCO e incluso puede ser peor.

- **Consistencia:** Bajo ciertas condiciones de regularidad, y si $\widehat{\Omega}$ es un estimador consistente de Ω, el estimador MCGF es consistente, $\text{plim}\, \hat{\beta}_{MCGF} = \beta$.

- **Distribución asintótica:** En general si las condiciones anteriores se satisfacen el estimador MCGF es asintóticamente equivalente al estimador MCG y su distribución asintótica coincide y por tanto es asintóticamente eficiente dentro de esta clase. Así, se puede demostrar que

$$\sqrt{T}(\hat{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, \sigma^2 G^{-1})$$

con $G = \text{plim}\, \frac{X'\Omega^{-1}X}{T}$.

2.4.2. Estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCGF}$

Cuando $E(uu') = \sigma^2\Omega$ y decimos que Ω es desconocida en realidad nos estamos refiriendo a que $\sigma^2\Omega$ es desconocido. Parece complicado ser capaces de establecer que $E(uu')$, que es desconocida, tiene un elemento común, o factor de escala, a todos sus elementos que también es desconocido y distinto de Ω. Lo lógico es pensar en $E(uu') = \Sigma$, Σ desconocida. Un estimador consistente de $V(\hat{\beta}_{MCGF})$ es $\widehat{V}(\hat{\beta}_{MCGF}) = (X'\widehat{\Sigma}^{-1}X)^{-1}$.

Sin embargo, y por si fuera el caso de que, siendo Ω desconocida existe este factor de escala σ^2, también desconocido en $E(uu')$, un estimador consistente de la matriz de varianzas y covarianzas asintótica del estimador de MCGF es $\widehat{V}(\hat{\beta}_{MCGF}) = \hat{\sigma}^2_{MCGF}(X'\widehat{\Omega}^{-1}X)^{-1}$ siendo

$$\hat{\sigma}^2_{MCGF} = \frac{\hat{u}'_{MCGF}\,\widehat{\Omega}^{-1}\,\hat{u}_{MCGF}}{T-K} \quad \text{y} \quad \hat{u}_{MCGF} = Y - X\hat{\beta}_{MCGF}$$

Es importante recalcar que la propiedad de consistencia es invariante a tener en el denominador $(T-K)$ o T.

Ejercicio 2.4 En el modelo $Y = X\beta + u$ donde $E(u) = 0, \quad E(uu') = \Sigma,\ \Sigma$ desconocida y X es no estocástica. Se pide:

1. Escribir la expresión del estimador MCGF.
2. Demostrar sus propiedades en muestras finitas y asintóticas.
3. Obtener su distribución asintótica.

2.5. Contrastes de restricciones lineales

Una vez hemos estimado los coeficientes del modelo de manera apropiada estaremos interesados en realizar contraste de hipótesis. Vamos a ver cómo realizar contrastes de restricciones lineales sobre el vector β en el MRLG con perturbaciones no esféricas.

Sean las hipótesis nula y alternativa para el contraste de q restricciones lineales:

$$H_0 : R\beta = r$$
$$H_a : R\beta \neq r$$

donde R es una matriz $(q \times K)$ y r es un vector de dimensión $(q \times 1)$, siendo q el número de restricciones lineales a contrastar.

• **Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCG.**
En este caso

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \quad \text{y} \quad \Omega \quad \text{es conocida}$$

el estimador MCG es lineal en u, insesgado y de varianza mínima, su distribución en muestras finitas es:

$$\hat{\beta}_{MCG} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$$

Estadísticos de contraste y distribución asociada:

Para $q = 1$

$$\frac{R\hat{\beta}_{MCG} - r}{\sqrt{R\,\widehat{V}(\hat{\beta}_{MCG})\,R'}} \overset{H_0}{\sim} t_{(T-K)}$$

luego

$$\frac{R\hat{\beta}_{MCG} - r}{\hat{\sigma}\sqrt{R\,(X'\Omega^{-1}X)^{-1}\,R'}} \overset{H_0}{\sim} t_{(T-K)}$$

Si el valor muestral del estadístico, t, es tal que t $< t_{(T-K)|\frac{\alpha}{2}}$ no rechazamos la H_0 para un nivel de significación α.

Para $q > 1$

$$(R\hat{\beta} - r)' \, [R \, \widehat{V}(\hat{\beta}_{MCG}) \, R' \,]^{-1} \, (R\hat{\beta} - r)\frac{1}{q} \overset{H_0}{\sim} F_{(q,T-K)}$$

luego

$$\frac{(R\hat{\beta}_{MCG} - r)' \, [R \, (X'\Omega^{-1}X)^{-1} \, R' \,]^{-1} \, (R\hat{\beta}_{MCG} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} F_{(q,T-K)}$$

Si el valor muestral del estadístico, F, es tal que F $< F_{(q,T-K)|\alpha}$ no rechazamos la H_0 para un nivel de significación α.

Ejercicio 2.5 Sea el modelo:

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \qquad \Omega \text{ conocida} \quad \text{y} \quad X \text{ no} \quad \text{estocástica}$$

Escribe el estadístico general y distribución asociada para el contraste de q restricciones lineales en el modelo transformado. Detalla cada uno de sus elementos y la regla de decisión.

Ejercicio 2.6 En el modelo $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad i = 1, \ldots, 500$ donde X_2 y X_3 son no estocásticas, $E(u_i) = 0 \,\forall i$, $Var(u_i) = \sigma^2 X_{2i}$, y $Cov(u_i u_j) = 0 \,\forall i \neq j$. Se pide:

1. ¿Cómo estimarías eficientemente β_1, β_2 y β_3? Describe claramente el método de estimación y escribe la forma matricial del estimador completando todos los elementos de las matrices implicadas.

2. ¿Como contrastarías la significatividad individual de la variable X_{2i}? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.

3. ¿Como contrastarías $H_0 : \beta_2 + \beta_3 = 2$? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.

4. ¿Como contrastarías la hipótesis compuesta $H_0 : \beta_2 = 1$ y $\beta_3 = 2$? Escribe la hipótesis nula, la alternativa y el estadístico de contraste junto con su distribución.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCGF.

En este caso

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \text{ desconocida, pero estimable consistentemente}$$

El estimador de MCGF, $\hat{\beta}_{MCGF} = (X'\widehat{\Sigma}^{-1}X)^{-1}(X'\widehat{\Sigma}^{-1}Y)$, es un estimador no lineal y en general sesgado, cuya distribución en muestras finitas no es conocida. En muestras grandes es un estimador consistente si $\widehat{\Sigma}$ es un estimador consistente de Σ. Su distribución asintótica es:

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

Los estadísticos de contraste y distribución asociada utilizando como estimador de $G^{-1} = \text{plim}\left[\frac{1}{T}X'\Sigma X\right]^{-1}$ a $T\widehat{V}(\hat{\beta}_{MCGF}) = T(X'\widehat{\Sigma}^{-1}X)^{-1}$ son,

para $q = 1$

$$\frac{R\hat{\beta}_{MCGF} - r}{\sqrt{R\ (X'\widehat{\Sigma}^{-1}X)^{-1}\ R'}} \xrightarrow{d,H_0} N(0,1)$$

para $q > 1$

$$(R\hat{\beta}_{MCGF} - r)'[\ R\ (X'\widehat{\Sigma}^{-1}X)^{-1}\ R'\ ]^{-1}(R\hat{\beta}_{MCGF} - r) \xrightarrow{d,H_0} \chi^2_q$$

Las reglas de decisión son las habituales.

• Estadísticos basados en el estimador de $\boldsymbol{\beta}$ por MCO y un estimador robusto a heterocedasticidad y/o autocorrelación de $\boldsymbol{V(\hat{\beta}_{MCO})}$

En presencia de perturbaciones no esféricas, el estimador Mínimo Cuadrático Ordinario es lineal, insesgado y consistente, pero no es de varianza mínima. Su matriz de varianzas y covarianzas se define $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Un estimador de la matriz de varianzas y covarianzas del estimador MCO es $\widehat{V}(\hat{\beta}_{MCO}) = \hat{\sigma}^2(X'X)^{-1}$. Utilizar este estimador para hacer inferencia cuando hay heterocedasticidad y/o autocorrelación no es adecuado. Los estadísticos t y F habituales para hacer inferencia sobre β definidos en base a este estimador de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ son inapropiados ya que es un estimador sesgado e inconsistente.

La dificultad que entraña el conocimiento de Ω, o Σ en su caso, hace interesante el poder contar con un estimador de $V(\hat{\beta}_{MCO})$ consistente y robusto a la posible existencia de heterocedasticidad y/o autocorrelación para de esta forma derivar estadísticos válidos, al menos asintóticamente, y contrastar hipótesis sobre el vector de coeficientes β utilizando $\hat{\beta}_{MCO}$.

Supongamos que $\widehat{V}(\hat{\beta}_{MCO})_{Consistente}$ es la estimación consistente de $V(\hat{\beta}_{MCO})$ de la que hablamos. Dado que el estimador MCO es consistente podemos utilizarlos conjuntamente para hacer inferencia asintótica válida. En este caso el estadístico válido para realizar contraste de hipótesis de la forma $H_0 : R\beta = r$ y su distribución asintótica asociada bajo la hipótesis nula son:

$$(R\hat{\beta}_{MCO} - r)'(R\ \widehat{V}(\hat{\beta}_{MCO})_{Consistente}\ R')^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_q$$

Siendo q el número de restricciones de contraste. Para $q = 1$ podemos escribir el estadístico anterior como:

$$\frac{R\hat{\beta}_{MCO} - r}{\sqrt{R\ \widehat{V}(\hat{\beta}_{MCO})_{Consistente}\ R'}} \xrightarrow{d,H_0} N(0,1)$$

Las reglas de decisión son las habituales. En los temas siguientes veremos que esta matriz robusta a heterocedasticidad y/o autocorrelación existe y puede ser calculada.

2.6. Ejemplo: Sistemas de Ecuaciones

En ocasiones necesitamos estimar un sistema de ecuaciones. Vamos a ver diferentes posibilidades de estimación de un sistema como ilustración del tema de perturbaciones no esféricas y a la vez, mostraremos como realizar el contraste de cambio estructural o de Chow.

Ilustraremos esta sección utilizando uno de los ejemplos vistos al comienzo del tema. Supongamos que queremos estimar la demanda de automóviles Y como una función de la renta X, utilizando datos microeconómicos sobre gastos de las familias en dos núcleos geográficos distintos, núcleo urbano y núcleo rural.

La función de demanda para las familias del núcleo urbano es:

$$Y_{1i} = a_1 + b_1 X_{1i} + u_{1i} \quad \Longleftrightarrow \quad Y_1 = X_1\beta_1 + u_1 \quad \text{siendo} \quad u_1 \sim N(0, \sigma_1^2 I_{N_1}) \quad i = 1, \ldots, N_1$$

La función de demanda para las familias del núcleo rural es:

$$Y_{2j} = a_2 + b_2 X_{2j} + u_{2j} \quad \Longleftrightarrow \quad Y_2 = X_2\beta_2 + u_2 \quad \text{siendo} \quad u_2 \sim N(0, \sigma_2^2 I_{N_2}) \quad j = 1, \ldots, N_2$$

Podemos escribir el siguiente sistema de ecuaciones:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \Longleftrightarrow \quad Y = X\beta + u \tag{2.5}$$

$$\underset{((N_1+N_2)\times 1)}{Y} = \underset{((N_1+N_2)\times 4)}{X} \; \underset{(4\times 1)}{\beta} + \underset{((N_1+N_2)\times 1)}{u}$$

Suponemos que X es no estocástica. El vector de medias y la matriz de varianzas y covarianzas del vector de perturbaciones u son:

$$\underset{((N_1+N_2)\times 1)}{E(u)} = \begin{bmatrix} E(u_1) \\ E(u_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\underset{((N_1+N_2)\times (N_1+N_2))}{E(uu')} = \begin{bmatrix} E(u_1u_1') & E(u_1u_2') \\ E(u_2u_1') & E(u_2u_2') \end{bmatrix}$$

En el sistema de ecuaciones anterior no hay relación entre los coeficientes de las dos ecuaciones. En cuanto a la estructura de $E(uu')$, podemos distinguir tres situaciones:

1. $E(u_1u_1') = \sigma_1^2 I_{N_1}$, $E(u_2u_2') = \sigma_2^2 I_{N_2}$ con $\sigma_1^2 = \sigma_2^2$, es decir, homocedasticidad entre ecuaciones y $E(u_1u_2') = E(u_2u_1') = 0$ lo que implica que no hay relación entre las perturbaciones de las dos ecuaciones.
2. $E(u_1u_1') = \sigma_1^2 I_{N_1}$, $E(u_2u_2') = \sigma_2^2 I_{N_2}$ con $\sigma_1^2 \neq \sigma_2^2$, es decir, heterocedasticidad entre ecuaciones, y $E(u_1u_2') = E(u_2u_1') = 0$.
3. Ecuaciones aparentemente no relacionadas $E(u_1u_2') = E(u_2u_1') \neq 0$.

Dado que no existe relación entre los coeficientes de las dos ecuaciones, el método más adecuado para estimar un modelo $Y = X\beta + u$ depende de la estructura de $E(uu')$.

2.6.1. Ecuaciones no relacionadas con varianza común

Sean las ecuaciones:

$$Y_1 = X_1\beta_1 + u_1 \quad N_1 \text{ obs.}$$

$$Y_2 = X_2\beta_2 + u_2 \quad N_2 \text{ obs.}$$

En principio β_1 y β_2 son distintos, como nada relaciona a las dos ecuaciones podríamos pensar que ganamos en eficiencia utilizando toda la información conjuntamente y por ello deberíamos estimar conjuntamente el modelo. Dado que $\sigma_1^2 = \sigma_2^2$ y $E(u_1u_2') = E(u_2u_1') = 0$ tenemos que

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \begin{bmatrix} \sigma^2 I_{N_1} & 0 \\ 0 & \sigma^2 I_{N_2} \end{bmatrix} = \sigma^2 I_{N_1+N_2}$$

El modelo puede ser estimado por MCO que será lineal, insesgado y de varianza mínima. Se puede probar que dado que no hay información común entre las ecuaciones, es decir, dado que X es diagonal por bloques y $E(u_1u_2') = E(u_2u_1') = 0$, la estimación del modelo conjunto por MCO es equivalente a estimar cada ecuación por separado por MCO. La estimación conjunta no gana en eficiencia, por tanto estimaremos las ecuaciones por separado por MCO.

Además, $\hat{\sigma}^2$ es insesgado y consistente:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{N-K} \qquad \text{donde} \quad N = N_1 + N_2 \quad \text{y} \quad K = K_1 + K_2$$

• **Contraste de cambio estructural o de Chow:** Se llama contraste de cambio estructural al contraste de que todos o algunos de los coeficientes que corresponden a las mismas variables en las dos ecuaciones son iguales. Supongamos que queremos contrastar la igualdad de ordenadas y pendientes o lo que es igual cambio estructural total:

$$\begin{cases} H_0: & a_1 = a_2 \text{ y } b_1 = b_2 \\ H_a: & a_1 \neq a_2 \text{ y/o } b_1 \neq b_2 \end{cases} \Longleftrightarrow \begin{cases} H_0: & \beta_1 = \beta_2 \\ H_a: & \beta_1 \neq \beta_2 \end{cases} \longrightarrow q = 2$$

Hay dos formas alternativas de realizar el contraste:

- Alternativa 1: Con el estadístico:

$$\text{F} = \frac{(R\hat{\beta}_{MCO} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}(q, N-K)$$

  donde

$$R\beta - r = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad q = 2 \qquad K = 4$$

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{N-K}$$

  Regla de decisión:

- Si F $< F_{(q,N-K)|\alpha}$ no rechazamos la H_0 para un nivel de significación α y concluimos que no existe cambio estructural.
- Si F $> F_{(q,N-K)|\alpha}$ rechazamos la H_0 para un nivel de significación α y concluimos que existe cambio estructural.

- Alternativa 2: Con el estadístico

$$\frac{(\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/N-K} \overset{H_0}{\sim} \mathcal{F}(q, N-K)$$

donde $\hat{u}'\hat{u}$ es la SCR del modelo no restringido (2.5) tal que $\hat{u}'\hat{u} = \hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2$ y $\hat{u}_r'\hat{u}_r$ es la SCR del modelo restringido siguiente:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{2.6}$$

Dado que hemos supuesto que $E(uu') = \sigma^2 I_{N_1+N_2}$ el modelo restringido se estima por MCO. Esta estimación no es equivalentemente a estimar por MCO ecuación por ecuación, ya que su matriz de regresores no es diagonal por bloques.

2.6.2. Ecuaciones no relacionadas con varianzas distintas

Sean las ecuaciones:

$$Y_1 = X_1\beta_1 + u_1 \quad N_1 \text{ obs.}$$

$$Y_2 = X_2\beta_2 + u_2 \quad N_2 \text{ obs.}$$

Dado que $\sigma_1^2 \neq \sigma_2^2$ y $E(u_1u_2') = E(u_2u_1') = 0$ tenemos que

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_{N_1} & 0 \\ 0 & \sigma_2^2 I_{N_2} \end{bmatrix} = \Sigma_{(N_1+N_2)\times(N_1+N_2)}$$

Suponiendo σ_1^2 y σ_2^2 conocidas, el modelo debiera ser estimado por MCG. Sin embargo, este procedimiento coincide con estimar por MCO ecuación por ecuación, ya que no existe relación entre los coeficientes de ambas ecuaciones, X es diagonal por bloques, $E(uu')$ también lo es y hay homocedasticidad dentro de cada ecuación.

$$\hat{\beta}_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y = \begin{bmatrix} (X_1'X_1)^{-1}X_1'Y_1 \\ (X_2'X_2)^{-1}X_2'Y_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}_{MCO}$$

$$V(\hat{\beta}_{MCG}) = (X'\Sigma^{-1}X)^{-1} = \begin{bmatrix} \sigma_1^2(X_1'X_1)^{-1} & 0 \\ 0 & \sigma_2^2(X_2'X_2)^{-1} \end{bmatrix}$$

En este caso la estimación del modelo conjunto por MCG no mejora la eficiencia con respecto a estimar cada ecuación por separado por MCO, por tanto estimaremos las ecuaciones por separado. Además, el resultado de estimar β es independiente de que conozcamos o no σ_1^2 y σ_2^2.

• **Contraste de cambio estructural o de Chow:**

$$\left\{\begin{array}{ll} H_0: & a_1 = a_2 \text{ y } b_1 = b_2 \\ H_a: & a_1 \neq a_2 \text{ y/o } b_1 \neq b_2 \end{array}\right. \Longleftrightarrow \left\{\begin{array}{ll} H_0: & \beta_1 = \beta_2 \\ H_a: & \beta_1 \neq \beta_2 \end{array}\right. \longrightarrow q = 2$$

Una forma de realizar el contraste es utilizar el estadístico:

$$(R\hat{\beta}_{MCG} - r)' \left[R(X'\Sigma^{-1}X)^{-1}R'\right]^{-1} (R\hat{\beta}_{MCG} - r) \overset{H_0}{\sim} \chi^2_q$$

donde

$$R\beta - r = \left[\begin{array}{rrrr} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{array}\right] \left[\begin{array}{c} a_1 \\ b_1 \\ a_2 \\ b_2 \end{array}\right] - \left[\begin{array}{c} 0 \\ 0 \end{array}\right] \qquad q = 2 \qquad K = 4$$

con la regla de decisión habitual.

• Si σ^2_1 y σ^2_2 son desconocidas es necesario estimarlas. El estimador de los coeficientes en el modelo conjunto sería el de MCGF tal que:

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

Estimar Σ implica estimar σ^2_1 y σ^2_2; un estimador consistente de $\sigma^2_i \quad i = 1, 2$ sería:

$$\hat{\sigma}^2_1 = \frac{\hat{u}'_1\hat{u}_1}{N_1} \qquad \hat{u}'_1\hat{u}_1 = Y'_1Y_1 - \hat{\beta}'_{1,MCO}X'_1Y_1 \qquad \hat{u}_1 = Y_1 - X_1\hat{\beta}_{1,MCO}$$

$$\hat{\sigma}^2_2 = \frac{\hat{u}'_2\hat{u}_2}{N_2} \qquad \hat{u}'_2\hat{u}_2 = Y'_2Y_2 - \hat{\beta}'_{2,MCO}X'_2Y_2 \qquad \hat{u}_2 = Y_2 - X_2\hat{\beta}_{2,MCO}$$

Notar que en cada ecuación por separado hay homocedasticidad y no autocorrelación.

2.6.3. Ecuaciones aparentemente no relacionadas

Si un conjunto de ecuaciones se relacionan únicamente por los términos de perturbación reciben el nombre de Ecuaciones Aparentemente no Relacionadas. Sea el sistema de ecuaciones

$$\left\{\begin{array}{lll} Y_1 & = X_1\beta_1 + u_1 & N \text{ obs} \\ Y_2 & = X_2\beta_2 + u_2 & N \text{ obs} \end{array}\right.$$

Un supuesto sencillo acerca de la estructura de la matriz de varianzas y covarianzas sería:

- $E(u_1u'_1) = \sigma^2_1 I_N \quad E(u_2u'_2) = \sigma^2_2 I_N$, homocedasticidad y no autocorrelación dentro de cada ecuación.

- $E(u_1u'_2) = E(u_2u'_1) = \sigma_{12}I_N$, correlación contemporánea entre las perturbaciones de las ecuaciones.

$$E(uu') = \begin{bmatrix} \sigma_1^2 I_N & \sigma_{12} I_N \\ \sigma_{12} I_N & \sigma_2^2 I_N \end{bmatrix} = \Sigma_{2N\times 2N}$$

En este caso ganamos en eficiencia si estimamos el modelo conjunto (2.5). Si σ_1^2, σ_2^2, σ_{12} son conocidas estimaremos por MCG. Si son desconocidas el modelo conjunto debiera estimarse por MCGF. Podemos encontrar estimadores consistentes de estos parámetros utilizando los residuos MCO de estimar cada ecuación por separado:

$$\begin{aligned}
\hat{\sigma}_1^2 &= \frac{\hat{u}_1'\hat{u}_1}{N} \qquad & \hat{u}_1'\hat{u}_1 &= Y_1'Y_1 - \hat{\beta}_{1,MCO}'X_1'Y_1 \qquad & \hat{u}_1 &= Y_1 - X_1\hat{\beta}_1 \\
\hat{\sigma}_2^2 &= \frac{\hat{u}_2'\hat{u}_2}{N} \qquad & \hat{u}_2'\hat{u}_2 &= Y_2'Y_2 - \hat{\beta}_{2,MCO}'X_2'Y_2 \qquad & \hat{u}_2 &= Y_2 - X_2\hat{\beta}_2 \\
\hat{\sigma}_{12} &= \frac{\hat{u}_1'\hat{u}_2}{N}
\end{aligned}$$