Sarriko-On

Análisis Econométrico

ISBN: 978-84-692-1728-3

Mª Victoria Esteban González

04-08

Facultad de Ciencias Económicas y Empresariales
Ekonomia eta Enpresal-Zientzien Fakultatea

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Análisis Econométrico

Mª Victoria Esteban González

Departamento de Economía Aplicada III. Econometría y Estadística
Facultad de Ciencias Económicas y Empresariales
Universidad del País Vasco/Euskal Herriko Unibertsitatea

Preámbulo

Las notas que se desarrollan a continuación no tienen más ambición que servir como apoyo al proceso de aprendizaje de los estudiantes de la asignatura *Econometría* de la Licenciatura en Economía y de la Licenciatura en Administración y Dirección de Empresas. Estas notas se estructuran en cinco capítulos a través de los cuales se van relajando las hipótesis básicas sobre la perturbación aleatoria y sobre la matriz de regresores. El primero de ellos revisa los conceptos de Teoría Asintótica que los alumnos ya han visto en las asignaturas de Estadística. Muestra los diferentes conceptos de convergencia y el Teorema de Mann y Wald adiestrando al alumno en su utilidad para derivar las propiedades en muestras grandes y distribución asintótica de los diferentes estimadores que verán en el curso.

El capítulo dos introduce el concepto de perturbaciones esféricas y muestra las consecuencias en las propiedades del estimador Mínimo Cuadrático Ordinario de que las perturbaciones no cumplan las hipótesis básicas. Asimismo deriva el estimador Mínimo Cuadrático Generalizado. Los capítulos tres y cuatro analizan los problemas de heterocedasticidad y autocorrelación, respectivamente. Muestran como detectar perturbaciones no esféricas y como contrastar la existencia de heterocedasticidad y/o autocorrelación. Aplican el estimador Mínimo Cuadrático Generalizado en el caso de que sea necesario y enseñan cómo estimar cuando la matriz de varianzas y covarianzas de la perturbación es desconocida utilizando el estimador de Mínimos Cuadrados Generalizados Factibles.

En el quinto capítulo se relaja la hipótesis básica sobre la matriz de regresores. Se aborda el escenario en que la matriz de datos es estocástica analizando los diferentes estadios de relación entre los regresores estocásticos y la perturbación aleatoria. Se deriva el estimador de Variables Instrumentales y se muestra la utilidad del contraste de Hausman. El capítulo finaliza analizando cómo actuar cuando además de tener regresores estocásticos la perturbación esta autocorrelada y/o es heterocedástica.

En cada capítulo se muestran ejemplos que ilustran diferentes escenarios de trabajo. Al término de las notas aparece la bibliografía completa.

Como decía anteriormente, estas notas sirven de apoyo al estudio. Analizan los problemas en profundidad y permiten al alumno profundizar en los temas más allá de lo visto en las clases presenciales. En ningún caso deben utilizarse como sustituto de los libros incluidos en la bibliografía. De igual manera recomiendo la realización de ejercicios tanto los recomendados en clase como los que aparecen en la bibliografía. La unión del estudio de los conceptos y la utilización de los mismos en los ejercicios permite adquirir la agilidad necesaria para el dominio de la asignatura.

Contenido

Tema 1

Teoría Asintótica

1.1. Introducción

Escribimos el modelo de regresión lineal en los parámetros como:

$$Y = X\beta + u$$

bajo las hipótesis:

(1) X es una matriz de regresores que no son realizaciones de variables aleatorias, es decir, son regresores fijos, no estocásticos. Este supuesto puede ser válido en experimentos controlados pero es bastante fuerte en econometría. Si consideramos $X_{i1}, X_{i2}, \dots X_{iT}$ realizaciones de una variable aleatoria X_i diremos que el regresor X_i es estocástico. Además, sobre los regresores suponemos que la matriz X es de rango completo por columnas.

(2) $E(u) = 0, \quad E(u_t) = 0 \quad \forall t$

(3) $E(uu') = \sigma^2 I_T \quad (E(u_t^2) = \sigma^2 \quad \forall t \quad y \quad E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s)$

(4) u se distribuye con función de distribución normal multivariante.

Los estimadores que proponemos para estimar los parámetros β y σ^2 del modelo son:

$\hat{\beta}_{MCO} = (X'X)^{-1}(X'Y)$ (vector de variables aleatorias que depende del tamaño muestral)

$\hat{\sigma}^2_{MCO} = \frac{\hat{u}'\hat{u}}{T-k} = \frac{u'Mu}{T-k}$ (vector de variables aleatorias que depende del tamaño muestral)

$\hat{\beta}_{MCO}$ es un estimador lineal, en el sentido de que dado (1) $(X'X)^1 X'$ es una matriz $(k \times T)$ de constantes y

$$\begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_T \end{bmatrix} = \begin{bmatrix} c_{11} & \dots & c_{1T} \\ \vdots & \ddots & \vdots \\ c_{k1} & \dots & c_{kT} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix} \Rightarrow \hat{\beta}_i = c_{i1}Y_1 + \dots + c_{iT}Y_T$$

Las propiedades del estimador $\hat{\beta}_{MCO}$ analizadas para un tamaño de muestra dado, T, son:

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

   $E(\hat{\beta}_{MCO}) = \beta$, a esta propiedad la llamamos insesgadez, es decir, si calculamos $\hat{\beta}_{MCO}$ para cada muestra de (X,Y) con el mismo tamaño muestral T y repitiéramos este proceso para todas las posibles muestras, obteniendo todas las posibles realizaciones de $\hat{\beta}_{MCO}$, la media de todas ellas sería igual al verdadero valor del parámetro β. Para demostrar esta propiedad hemos utilizado dos de las hipótesis básicas, que X es una matriz de variables fijas y que $E(u_t) = 0 \quad \forall t$.

2. Por otro lado la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ (para un tamaño de muestra dado T) es:

   $$Var(\hat{\beta}_{MCO}) = E[(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))']$$

   bajo los supuestos (1) y (2) $\quad E(\hat{\beta}_{MCO}) = \beta$
   y

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

   Se demuestra que bajo estos supuestos (1), (2) y (3) el estimador $\hat{\beta}_{MCO}$ es un estimador lineal, insesgado y eficiente. Es decir, cualquier otro estimador lineal, insesgado de β, $\hat{\beta}^\star = c^\star Y$ tendría una matriz de varianzas y covarianzas $V(\hat{\beta}^\star)$ tal que $V(\hat{\beta}^\star) - V(\hat{\beta}_{MCO})$ es una matriz semidefinida positiva. Insesgadez y eficiencia son propiedades deseables en un estimador, y son propiedades llamadas para muestras finitas, es decir para un tamaño de muestra finito dado T (no variamos el tamaño muestral).

3. Además, bajo el supuesto (4) de normalidad del término de perturbación, como $\hat{\beta}_{MCO}$ es una combinación lineal de variables aleatorias normales, es decir,

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

   entonces podemos conocer la distribución de $\hat{\beta}_{MCO}$ para un tamaño de muestra T finito dado,

   $$(\hat{\beta}_{MCO} - \beta) \sim N(E(cu), E(cuu'c'))$$

   bajo (1) y (2) $E(cu) = 0$ y bajo (3) $E(cuu'c') = \sigma^2 cc' = \sigma^2(X'X)^{-1}$ esto nos permite hacer inferencia y contrastar hipótesis de restricciones lineales sobre β, si σ^2 es conocida.

4. Cuando σ^2 es desconocida proponemos como estimador estimador a $\hat{\sigma}^2_{MCO}$ tal que:

   $$E(\hat{\sigma}^2_{MCO}) = \frac{E(u'Mu)}{T-k} = \frac{1}{T-k}tr(M)E(uu')$$

   Calcular $E(u'Mu)$ bajo el supuesto (1) también es fácil ya que $M = [I - X(X'X)^{-1}X']$ es una matriz de constantes (no realizaciones de variables aleatorias) y por lo tanto esto permite encontrar $E(u'Mu) = tr(M)E(uu')$. Bajo (3) obtenemos $E(\hat{\sigma}^2_{MCO}) = \sigma^2$.

Dado también que bajo estos supuestos (1), (2), (3) y (4)

$$\frac{(T-k)\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{(T-k)}$$

esto nos permite construir estadísticos como la $\hat{t}$ o la $\hat{F}$ habituales cuya distribución para un tamaño de muestra finito T es conocida, es decir bajo $H_0 : R\beta = r$ de q-restricciones lineales.

$\hat{F}_0$ se distribuye como una F de Snedecor con $(q, T-k)$ grados de libertad.

$\hat{t}_0$ se distribuye como una t de Student con $(T-k)$ grados de libertad (cuando q=1).

Las distibuciones tabuladas de estos estadísticos nos permiten, estadísticamente, distinguir a un nivel de significación elegido si aceptar o no la hipótesis nula dado el valor del estadístico obtenido en la muestra.

Hasta ahora, hemos derivado bajo los supuestos (1), (2), (3) y (4) las propiedades para estimadores y estadísticos y sus distribuciones para un tamaño de muestra dado T y finito. En principio T puede ser cualquier tamaño muestral, pero al analizar las propiedades este tamaño muestral está dado y no cambia. Esto se conoce como propiedades para muestras finitas.

La teoría asintótica analiza el comportamiento y las propiedades de variables aleatorias (que, por ejemplo, puede ser un estimador) cuando varía el tamaño de la muestra y permitimos que T aumente hasta infinito (aunque este infinito puede ser algo finito pero suficientemente grande).

Definiremos una serie de conceptos y propiedades deseables en un estimador que serán válidos asintóticamente, es decir, cuando el tamaño muestral sea suficientemente grande ($T \longrightarrow \infty$).

¿Por qué analizamos esto?
Es deseable poder obtener estimadores y estadísticos y que conozcamos sus propiedades, para muestras finitas, de insesgadez, eficiencia y sus distribuciones; pero en muchas ocasiones estas propiedades se conocen a costa de tener que hacer supuestos muy particulares y restrictivos que pueden no satisfacerse y ser difíciles de contrastar.

Otras veces no es posible encontrar estimadores que tengan propiedades deseables para muestras finitas pero que tienen propiedades deseables en muestras grandes, es decir cuando $T \longrightarrow \infty$, cuando el tamaño muestral es suficientemente grande. A estas propiedades las llamaremos propiedades asintóticas.

En este tema pretendemos relajar dos hipótesis básicas:

a) Queremos relajar la hipótesis de normalidad del término de perturbación y no especificamos una función de distribución determinada para u, entonces:

   I) no conocemos, para un tamaño de muestra T dado, como se distribuye $\hat{\beta}_{MCO}$ ó $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$. Por lo tanto, no conocemos la distribución de los estadísticos $\hat{t}$ ó $\hat{F}$ para poder contrastar hipótesis sobre β o σ^2.

   II) veremos resultados de teoría asintótica que nos permitirán bajo algún supuesto pero relajando $u \sim Normal$, conocer la distribución asintótica de los estimadores y estos estadísticos.

b) Si relajamos el supuesto de que los regresores en X son fijos y ahora $X_{i1}, X_{i2}, \ldots, X_{iT}$, son T realizaciones de la variable aleatoria X_{it}, el estimador

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

ahora es una combinación no lineal de las variables aleatorias X y u. En este caso necesitaremos ver qué condiciones necesitamos ahora para que los estimadores sean insesgados.

En general vamos a buscar otro tipo de propiedades que llamaremos asintóticas, veremos consistencia, insesgadez asintótica y eficiencia asintótica. Realizaremos supuestos bajo los cuales podamos determinar cual es la distribución de $\hat{\beta}_{MCO}$ y $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$ y así poder conocer la distribución de los estadísticos $\hat{t}$ y $\hat{F}$ que nos permitan realizar inferencia. Así, veremos el concepto de distribución asintótica y bajo qué condiciones podemos derivar distribuciones asintóticas para los estimadores y los estadísticos que nos permitan hacer inferencia aunque estás sean válidas sólo asintóticamente, es decir, cuando el tamaño muestral sea suficientemente grande.

En resumen, queremos ver cómo se comportan las variables aleatorias en el límite, pero antes necesitamos ciertos conceptos previos.

1.2. Convergencia en probabilidad. Operador plim, propiedades.

Sea Z una variable aleatoria y $Z_1, Z_2, \ldots Z_T = \{Z_{(T)}\}$ una sucesión de variables aleatorias definidas en un espacio de probabilidad $\{\Omega, \forall, Pr\}$

- Definición de **Convergencia en probabilidad o Convergencia en ley débil** de una sucesión de variables aleatorias.

  Sea $Z_1, Z_2, \ldots Z_T$ una sucesión de variables aleatorias denotadas por $\{Z_{(T)}\}$. Se dice que esta sucesión de variables aleatorias converge en probabilidad a Z, donde Z puede ser una variable aleatoria o una constante (no aleatoria) si:

  $$\forall \epsilon > 0, \lim_{T \to \infty} Pr\{|Z_{(T)} - Z| \leq \epsilon\} = 1$$

  o lo que es lo mismo, mirando al suceso contrario:

  $$\forall \epsilon > 0, \lim_{T \to \infty} Pr\{|Z_{(T)} - Z| > \epsilon\} = 0$$

  lo denotamos como :

  $$Z_{(T)} \xrightarrow{p} Z \quad \text{ó} \quad plim Z_T = Z$$

  y se lee el límite en probabilidad de la sucesión de variables aleatorias $\{Z_{(T)}\}$ es Z ó la sucesión de variables aleatorias $\{Z_{(T)}\}$ converge en probabilidad a Z.

- ¿Qué estamos diciendo?
  Si definimos un entorno a Z_0, $Z_0 \pm \epsilon$, indica que la probabilidad de que $Z_{(T)}$ esté dentro de un intervalo estrictamente pequeño en torno al valor Z_0 se puede hacer tan próxima a 1 como queramos, haciendo T suficientemente grande.

  Si $T \longrightarrow \infty$ la probabilidad de que $Z_{(T)}$ salga fuera del entorno $(Z_0 - \epsilon, Z_0 + \epsilon)$ es cero, las variables aleatorias tienen una distribución más ajustada al valor Z_0, es decir, tienden en probabilidad a Z_0 para cualquier valor de ϵ.

- **Ejemplo:** Sea X_t una v.a tal que $X_t \sim (\mu, \sigma^2)$, y definimos $\bar{X} = \frac{\sum X_t}{T}$, entonces:

  $$E(\bar{X}_T) = \mu$$
  $$Var(\bar{X}_T) = \frac{\sigma^2}{T}$$

  si $T \longrightarrow \infty \quad Var(\bar{X}_T) \longrightarrow 0$ y la distribución de $\bar{X}_T$ está más concentrada entorno a μ.

- Comentarios:

a) La convergencia en probabilidad equivale al límite numérico de una sucesión de probabilidades, es decir, $\forall \epsilon > 0$:

$$\begin{array}{rlll} Prob(|Z_1 - Z| \leq \epsilon) = P_1(\epsilon) & es & un & numero \\ Prob(|Z_2 - Z| \leq \epsilon) = P_2(\epsilon) & es & un & numero \\ \vdots & & & \vdots \\ Prob(|Z_{(T)} - Z| \leq \epsilon) = P_T(\epsilon) & es & un & numero \\ \vdots & & & \vdots \end{array}$$

Miramos si $\lim_{T\to\infty} P_t(\epsilon) = 1$
$\Longrightarrow$ la convergencia en probabilidad equivale al límite en los números reales.

b) Si Z es una constante (no aleatoria) o es una variable aleatoria acotada (es decir sus realizaciones están dentro de un rango acotado) podemos escribir:

$$Z_{(T)} - Z \xrightarrow{p} 0 \quad \text{ó} \quad plim(Z_{(T)} - Z) = 0$$

es decir, si Z es una constante, $\{Z_{(T)}\}$ converge a la constante.

c) La convergencia en probabilidad no precisa de la existencia ni el conocimiento de los momentos de la variable aleatoria de la sucesión.

d) **Plim** es un operador similar a **lim** y será equivalente cuando $\{Z_{(T)}\}_{T=1}^{\infty}$ no sean variables aleatorias.

e) Si $\{Z_{(T)}\}_{T=1}^{\infty}$ es una sucesión de vectores aleatorios de dimensión $(s \times 1)$ fija entonces:

$$\begin{array}{rl} |Z_{(T)} - Z| & = \text{norma euclídea del vector } (Z_{(T)} - Z) = \\ & = (\sum_{j=1}^{s}(Z_{Tj} - Z)^2)^{1/2} \end{array}$$

- **Propiedades de los límites en probabilidad:**
Sean $\{Z_{(T)}\}_{T=1}^{\infty}$ y $\{S_{(T)}\}_{T=1}^{\infty}$ dos sucesiones de variables aleatorias:

a) $plimc = c$ donde c es una constante.

b) $plim(Z_{(T)} + S_{(T)}) = plimZ_{(T)} + plimS_{(T)}$

c) $plim(Z_{(T)} \cdot S_{(T)}) = plimZ_{(T)} \cdot plimS_{(T)}$

d) $plim\left(\frac{Z_{(T)}}{S_{(T)}}\right) = \frac{plimZ_{(T)}}{plimS_{(T)}}$ siempre que $plimS_{(T)} \neq 0$

e) $plim\left(\frac{1}{Z_{(T)}}\right) = (plimZ_{(T)})^{-1}$ si $plimZ_{(T)} \neq 0$

f) $plim(Z_{(T)})^2 = (plimZ_{(T)})^2$

En general, tenemos el siguiente resultado:

- **Teorema de Slutsky:**
Si $plimZ_{(T)} = Z$ y $g(\bullet)$ es una función continua entonces:

$$plimg(Z_{(T)}) = g(plimZ_{(T)}) = g(Z)$$

En base a este teorema tenemos los resultados 5 y 6.

$plim(\bullet)$ es un operador matemático más operativo que $E(\bullet)$, sin embargo este último requiere independencia $E(XY) = E(X)E(Y)$ y $plim(\bullet)$ no.

- **Generalización de los resultados a vectores y matrices:**
La definición de convergencia en probabilidad se generaliza al caso de una sucesión de vectores de variables aleatorias de dimensión k fija.

  a) Sea $A_1, A_2, \ldots, A_T, \ldots$ una sucesión de matrices de orden constante $(k \times k)$ cuyos elementos son variables aleatorias.
  $plimA_T = A$ indica que $\{a_{ijT}\}_{T=1}^{\infty}$ converge en probabilidad a $a_{ij} \quad \forall i, j$ donde a_{ij} es el elemento (i, j) de la matriz A y $a_{ij(T)}$ es el elemento (i, j) de la matriz $A_{(T)}$. (Nota: convergencia elemento a elemento).

  b) Si $plimA_{(T)} = A$ y $plimB_{(T)} = B$

$$\begin{array}{l} plimA_{(T)} \cdot plimB_{(T)} = AB \\ plimA_{(T)} + plimB_{(T)} = A + B \\ plimB_{(T)}^{-1} = (plimB_{(T)})^{-1} = B^{-1} \quad \text{si} \quad B \quad es \quad no \quad singular \end{array}$$

- **Ejemplo:**
h.q.d $\qquad plim\bar{X}_n = \mu$
donde $\bar{X}_n = \frac{X_1+X_2+\ldots+X_n}{n}$, $E(X_i) = \mu \quad \forall i$ y $Var(X_i) = \sigma^2 \quad \forall i.$

Solución: Tenemos la siguiente sucesión de v.a.:

$$\begin{array}{l} \bar{X}_1 = \frac{X_1}{1} \\ \bar{X}_2 = \frac{X_1+X_2}{2} \\ \bar{X}_3 = \frac{X_1+X_2+X_3}{3} \\ \vdots \quad \vdots \\ \bar{X}_n = \frac{X_1+X_2+\ldots+X_n}{n} \end{array}$$

siendo X_i realizaciones de una misma variable.

$$E(\bar{X}_n) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n} = \frac{\mu + \mu + \ldots + \mu}{n} = \frac{n\mu}{n} = \mu$$

$$\begin{aligned} Var(\bar{X}_n) &= Var\left(\tfrac{X_1+X_2+\ldots+X_n}{n}\right) = E\left(\tfrac{X_1+X_2+\ldots+X_n}{n} - E\left(\tfrac{X_1+X_2+\ldots+X_n}{n}\right)\right)^2 \\ &= E\left(\left(\tfrac{X_1-E(X_1)}{n}\right) + \left(\tfrac{X_2-E(X_2)}{n}\right) + \ldots + \left(\tfrac{X_n-E(X_n)}{n}\right)\right)^2 \\ &= E\left(\tfrac{X_1-E(X_1)}{n}\right)^2 + E\left(\tfrac{X_2-E(X_2)}{n}\right)^2 + \ldots + E\left(\tfrac{X_n-E(X_n)}{n}\right)^2 \\ &\quad + Productos \quad cruzados = \\ &= \tfrac{1}{n^2}E(X_1-\mu)^2 + \tfrac{1}{n^2}E(X_2-\mu)^2 + \ldots + \tfrac{1}{n^2}E(X_n-\mu)^2 = \\ &= \tfrac{Var(X_1)}{n^2} + \tfrac{Var(X_2)}{n^2} + \ldots + \tfrac{Var(X_n)}{n^2} = \\ &= \tfrac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \tfrac{n\sigma^2}{n^2} = \tfrac{\sigma^2}{n} \end{aligned}$$

Aplicando la desigualdad de Chebychev:

$$Pr\left[|X - E(X)| > k\sqrt{Var(X)}\right] < \frac{1}{k^2}$$

en nuestro caso tendremos:

$$Pr\left[|\bar{X}_n - \mu| > k\frac{\sigma}{\sqrt{n}}\right] < \frac{1}{k^2} \qquad \forall k > 0$$

llamamos $\epsilon = \frac{\sigma k}{\sqrt{n}} \Longrightarrow k = \frac{\epsilon\sqrt{n}}{\sigma}$

$$Pr\left[|\bar{X}_n - \mu| > \epsilon\right] < \frac{1}{\frac{\epsilon^2 n}{\sigma^2}}$$

$$Pr\left[|\bar{X}_n - \mu| > \epsilon\right] < \frac{\sigma^2}{\epsilon^2 n}$$

$$\lim_{n\to\infty} Pr\left[|\bar{X}_n - \mu| > \epsilon\right] < \lim_{n\to\infty} \frac{\sigma^2}{\epsilon^2 n}$$

así

$$\lim_{n\to\infty} Pr\left[|\bar{X}_n - \mu| > \epsilon\right] = 0 \Longrightarrow plim\bar{X}_n = \mu \quad \text{ó} \quad \bar{X}_n \xrightarrow{p} \mu$$

1.3. Convergencia casi segura o con probabilidad 1. Convergencia en ley fuerte

Se dice que la sucesión $\{Z_{(T)}\}_{T=1}^{\infty}$ de variables aleatorias converge con probabilidad uno o casi segura a la variable aleatoria Z si:

$$Pr\left[\lim_{T\to\infty} Z_{(T)} = Z\right] = 1 \qquad \text{ó} \qquad Pr\left[\lim_{T\to\infty} |Z_{(T)} - Z| \leq \epsilon\right] = 1$$

o mirando al suceso contrario

$$Pr\left[\lim_{T\to\infty} |Z_{(T)} - Z| > \epsilon\right] = 0$$

Convergencia en probabilidad 1 $\Longrightarrow$ convergencia en probabilidad

Lo denotamos: $Z_{(T)} \xrightarrow{a.s} Z$ (nota a.s indica almost sure)

- Observaciones:
  a) Si Z es acotada $\quad Z_{(T)} \xrightarrow{a.s} Z \quad$ equivale a $\quad Z_{(T)} - Z \xrightarrow{a.s} 0$
  b) Si Z es degenerada $Z_{(T)} - Z \xrightarrow{a.s} 0$
  c) El límite en convergencia casi segura no equivale al límite en los números reales.

1.4. Convergencia en media cuadrática

Definición: Una sucesión de variables aleatorias $\{Z_{(T)}\}_{T=1}^{\infty}$ se dice que converge a Z en media cuadrática si $\lim_{T\to\infty} E(Z_{(T)} - Z)^2 = 0$. Se denota: $Z_{(T)} - Z \xrightarrow{m.c} Z$ (Z puede ser una v.a o una constante (no aleatoria)).

- Comentarios:
  a) Este concepto de convergencia exige la existencia del error cuadrático medio de cada variable aleatoria en la sucesión.

b) Convergencia en media cuadrática $\Longrightarrow$ convergencia en probabilidad.

$$Z_{(T)} - Z \xrightarrow{m.c} Z \Longrightarrow Z_{(T)} \xrightarrow{p} Z$$

Demostración:
Consideramos la v.a. $(Z_{(T)} - Z)$. Por la desigualdad de Chebychev podemos escribir:

$$Pr\left[|Z_{(T)} - Z| > k\sqrt{Var(Z)}\right] < \frac{1}{k^2} \Longrightarrow$$

llamamos $\epsilon = k\sqrt{var(Z_{(T)})}$

$$Pr\left[|Z_{(T)} - Z| > \epsilon\right] < \frac{Var(Z_{(T)})}{\epsilon^2} = \frac{E(Z_{(T)} - Z)^2}{\epsilon^2}$$

$$Pr\left[|Z_{(T)} - Z| > \epsilon\right] \leq \frac{E(Z_{(T)} - Z)^2}{\epsilon^2} \qquad \forall \epsilon > 0$$

tomamos límites cuando $T \to \infty$

$$\lim_{T\to\infty}\left[Pr|Z_{(T)} - Z| > \epsilon\right] \leq \frac{1}{\epsilon^2}\lim_{T\to\infty} E(Z_{(T)} - Z)^2$$

si $Z_{(T)} \xrightarrow{m.c} Z$ entonces $\lim_{T\to\infty} E(Z_{(T)} - Z)^2 = 0$, por lo tanto:

$$\lim_{T\to\infty} Pr\left[|Z_{(T)} - Z| > \epsilon\right] = 0 \qquad \forall \epsilon > 0$$

por lo tanto:

$$Z_{(T)} \xrightarrow{p} Z \qquad \text{ó} \qquad plim Z_{(T)} = Z \qquad c.q.d.$$

c) Convergencia en media cuadrática es más fuerte que convergencia en probabilidad. Es decir, convergencia en probabilidad no implica necesariamente convergencia en media cuadrática.

$$Z_{(T)} - Z \xrightarrow{m.c} Z \quad \underset{\not\Longleftarrow}{\Longrightarrow} \quad Z_{(T)} \xrightarrow{p} Z$$

d) Si Z es una constante y

$$\left.\begin{array}{l} (1)\ \lim_{T\to\infty} E(Z_{(T)}) = Z \\ (2)\ \lim_{T\to\infty} Var(Z_{(T)}) = 0 \end{array}\right\} \lim_{T\to\infty} E(Z_{(T)} - Z)^2 = 0$$

$$\Longrightarrow Z_{(T)} \xrightarrow{m.c} Z \Longrightarrow Z_{(T)} \xrightarrow{p} Z$$

Demostración:

$$\begin{aligned} E(Z_{(T)} - Z)^2 &= E(Z_{(T)} - E(Z_{(T)}) + E(Z_{(T)}) - Z)^2 = \\ &= E[(Z_{(T)} - E(Z_{(T)}))^2] + E[(E(Z_{(T)}) - Z)^2] + \\ &+ 2E[(Z_{(T)} - E(Z_{(T)}))(E(Z_{(T)}) - Z)] \end{aligned}$$

como:

$$\begin{aligned} &E[(Z_{(T)} - E(Z_{(T)}))(E(Z_{(T)}) - Z)] \\ &= E[Z_{(T)}E(Z_{(T)}) - Z_{(T)}Z - (E(Z_{(T)}))^2 + ZE(Z_{(T)})] \\ &= [E(Z_{(T)})]^2 - ZE(Z_{(T)}) - (E(Z_{(T)}))^2 + ZE(Z_{(T)}) = 0 \end{aligned}$$

tenemos que:

$$\begin{aligned} E[Z_{(T)} - Z]^2 &= E[(Z_{(T)} - E(Z_{(T)}))^2] + [E[Z_{(T)}] - Z]^2 \\ &= Var(Z_{(T)}) + [E(Z_{(T)}) - Z]^2 \end{aligned}$$

si $\lim_{T\to\infty} E(Z_{(T)}) = Z \Longrightarrow \lim_{T\to\infty} E(Z_{(T)}) - Z = 0$
y $\lim_{T\to\infty} Var(Z_{(T)}) = 0$ entonces

$$\lim_{T\to\infty} E(Z_{(T)} - Z)^2 = 0 \qquad \text{c.q.d}$$

1.5. Insesgadez asintótica y consistencia

- **Consistencia de un estimador:**
  Supongamos que $\hat{\theta}$ es un estimador del parámetro θ. $\hat{\theta}$ es una variable aleatoria que es función de la muestra de un tamaño T. Si consideramos la sucesión de variables aleatorias:
  $\hat{\theta}_1$ estimador función de la muestra tamaño T_1
  $\hat{\theta}_2$ estimador función de la muestra tamaño T_2
  ....
  $\hat{\theta}_N$ estimador función de la muestra tamaño T_N
  y así sucesivamente a medida que consideramos muestras de mayor tamaño ($T \longrightarrow \infty$) obtendríamos una sucesión de variables aleatorias $\{\hat{\theta}_{(T)}\}$.
  Si la sucesión $\{\hat{\theta}_{(T)}\}$ converge en probabilidad al verdadero valor (desconocido) del parámetro θ se dice que el estimador $\hat{\theta}$ es un estimador consistente.

- **Definición formal:**
  Se dice que un estimador $\hat{\theta}$ es un estimador consistente del parámetro desconocido θ si la sucesión $\{\hat{\theta}_{(T)}\} = \{\hat{\theta}_1, \hat{\theta}_2, \ldots\}$ converge en probabilidad a θ y lo denotamos.

$$plim\hat{\theta} = \theta$$

$$\Longleftrightarrow \lim_{T\to\infty} Pr\{|\hat{\theta}_{(T)} - \theta| \leq \epsilon\} = 1 \quad \forall \epsilon > 0$$

$$\Longleftrightarrow \lim_{T\to\infty} Pr\{|\hat{\theta}_{(T)} - \theta| > \epsilon\} = 0 \quad \forall \epsilon > 0$$

- Comentarios:
  a) Es una propiedad asintótica deseable en un estimador ya que analizamos si la sucesión de estimadores $\{\hat{\theta}_{(T)}\}$ converge en probabilidad al verdadero valor del parámetro.
  b) Para que un estimador sea consistente no necesitamos conocer los momentos de $\hat{\theta}_T$, es decir $E(\hat{\theta}_T)$ ó $E(\hat{\theta}_T^2)$, ó $E[\hat{\theta}_T^2 - E(\hat{\theta}_t)]$ etc ... ni necesitamos que existan para cada tamaño muestral.

- **Condiciones suficientes de consistencia:**(no necesarias)
  Sea $\{E(\hat{\theta}_{(T)})\}$ una sucesión del momento de primer orden de $\hat{\theta}_T$ cuando el tamaño muestral T aumenta hasta infinito.
  Sea $\{Var(\hat{\theta}_{(T)})\}$ una sucesión del momento centrado de orden dos de $\hat{\theta}_T$ cuando el tamaño muestral T aumenta hasta infinito. Entonces si:

$$\left.\begin{array}{l} (1)\, \lim_{T\to\infty} E(\hat{\theta}_{(T)}) = \theta \\ (2)\, \lim_{T\to\infty} Var(\hat{\theta}_{(T)}) = 0 \end{array}\right\} \hat{\theta}_{(T)} \xrightarrow{p} \theta \Longrightarrow plim\hat{\theta}_{(T)} = \theta$$

  - Demostración: como hemos demostrado

$$(1) + (2) \Longrightarrow \lim_{T\to\infty} E(\hat{\theta}_{(T)} - \theta)^2 = 0 \Longrightarrow \hat{\theta}_{(T)} \xrightarrow{m.c} \theta \Longrightarrow \hat{\theta}_{(T)} \xrightarrow{p} \theta \Longrightarrow plim\hat{\theta} = \theta$$

  - Comentarios:
    a) Estas dos condiciones son suficientes pero no necesarias.
    b) Para verificar estas dos condiciones suficientes para que $\hat{\theta}$ sea consistente necesitamos conocer y que existan:

$$E(\hat{\theta}_{(T)}) \quad \forall T \quad \text{y} \quad E(\hat{\theta}_{(T)} - E(\hat{\theta}_{(T)}))^2 \quad \forall T$$

cosa que en principio no es necesaria para covergencia en probabilidad. Esto es debido a que estas dos condiciones son suficientes para convergencia en media cuadrática de $\hat{\theta}_{(T)}$ a θ y que implica convergencia en probabilidad pero éste último es un concepto de convergencia más débil.

1.5.1. Insesgadez Asintótica

Definición:
Diremos que el estimador $\hat{\theta}$ de θ es un estimador insesgado asintóticamente si

$$\lim_{T\to\infty} E(\hat{\theta}_{(T)}) = \theta \Longleftrightarrow \lim_{T\to\infty} [E(\hat{\theta}_{(T)} - \theta] = 0$$

- Comentarios:

  a) Insesgadez asintótica no implica consistencia. Además se requiere como otra condición suficiente aunque no necesaria, que

  $$\lim_{T\to\infty} Var(\hat{\theta}_{(T)}) = 0$$

  b) Insesgadez asintótica no requiere que el estimador sea insesgado, es decir que $E(\hat{\theta}_{(T)}) = \theta \quad \forall T$. Ahora si $E(\hat{\theta}_{(T)}) = \theta \quad \forall T$

  $$\Longrightarrow \lim_{T\to\infty} E(\hat{\theta}_{(T)}) = \theta$$

  c) Consistencia no implica insesgadez asintótica.

  d) Insesgadez asintótica requiere el conocer y la existencia del momento de primer orden $E(\hat{\theta}_{(T)}) \quad \forall T$.

1.5.2. Eficiencia Asintótica

Si nos limitamos a la clase de estimadores consistentes, asintóticamente insesgados y asintóticamente normales, diremos que un estimador de esa clase es eficiente asintóticamente si y sólo si su varianza asintótica es la menor de todas las varianzas asintóticas de los estimadores de esa clase.

1.6. Convergencia en distribución

- **Definición de Convergencia en Distribución:**
  Sea $Z_1, Z_2, Z_3 \ldots, Z_T$ una sucesión de variables aleatorias definidas conjuntamente con sus respectivas funciones de distribución $F_1(Z_1), F_2(Z_2), F_3(Z_3) \ldots, F_T(Z_T)$ diremos que la sucesión $\{Z_{(T)}\}_{T=1}^{\infty}$ converge en distribución a la variable aleatoria Z con función de distribución F si y sólo si para todos los puntos de continuidad de $F(Z)$ se cumple:

  $$\lim_{T\to\infty} F_T(Z_{(T)}) = F(Z)$$

  y lo denotamos

  $$Z_{(T)} \xrightarrow{d} Z \qquad \text{ó} \qquad Z_{(T)} \xrightarrow{d} F(Z)$$

  la distribución $F(Z)$ se conoce como Distribución Asintótica o Distribución Límite.

- Ejemplo:
  Si tenemos una v.a. $Z_{(T)} : Pr[a < Z_T < b] = \int_a^b f(Z_T)dZ_t = F_T(b) - F_T(a)$ la convergencia en ley débil nos dice:

$$\lim_{T\to\infty} Pr[a < Z_T < b] = F_T(b) - F_T(a) = \int_a^b f(Z)dZ = F(b) - F(a)$$

  A medida que aumenta el tamaño muestral T, la función de distribución tiende a identificarse con $F(Z)$.

- Comentarios:

  a) La utilidad de este concepto está en la posibilidad de aproximar una función de distribución $F_{(T)}$, que podemos no conocer, o tener que realizar supuestos que no sean satisfactorios para conocerla, por una función de distribución F que sea conocida y que si T es suficientemente grande esa aproximación sea válida o buena para $F_{(T)}$. En Econometría muchas veces la distribución para un T dado de un estimador o estadístico para la realización de un contraste no es conocida o es difícil de evaluar, pero todavía es posible derivar la forma de la distribución asintótica de ese estimador o estadístico. La distribución asintótica puede servirnos como una aproximación a la distribución para un tamaño muestral T dado o muestra finitas. Esta aproximación será mejor cuanto mayor sea el tamaño muestral T dado que es un resultado para $T \longrightarrow \infty$.

  b) Convergencia en distribución es una forma de convergencia más débil que convergencia en probabilidad, esto es.

$$Z_{(T)} \xrightarrow{p} Z \Longrightarrow Z_{(T)} \xrightarrow{d} Z$$

  pero la implicación inversa no es necesariamente cierta. $Z_{(T)} \xrightarrow{p} Z$ implica que para un T suficientemente grande la probabilidad de que Z_T difiera de Z es muy pequeña y por lo tanto esperaremos que tengan aproximadamente la misma distribución. En cambio, es posible que una sucesión de variables aleatorias $Z_1, Z_2, \ldots$ independientes, con la misma distribución, por ejemplo $N(0, \sigma^2)$ si $Z_{(T)} \xrightarrow{d} Z$ entonces cualquier v.a. de la sucesión $\{Z_{(T)}\}_{T=1}^{\infty}$ tendrá aproximadamente la misma distribución que Z, es decir $N(0, \sigma^2)$. Pero cualquier realización de $Z_{(T)}$ puede que no tenga relación con una realización de la variable aleatoria Z.

  c) Si Z no es una v.a. entonces si $Z_{(T)}$ converge en distribución a la constante $Z \Longrightarrow Z_{(T)}$ converge en probabilidad a Z.

  El que $Z_{(T)}$ converja en distribución a una constante implica que la distribución asintótica es una distribución degenerada, es decir la varianza de la distribución asintótica es cero y la media de la distribución asintótica es la constante Z.

  En Econometría, lo habitual va a ser encontrarnos con distribuciones centradas en una constante (β, σ^2) a las que llamaremos degeneradas, en estos casos:

$$Z_{(T)} \xrightarrow{d} Z \Longrightarrow Z_{(T)} \xrightarrow{p} Z$$

  Ejemplo:
  Si $X \sim N(\mu, \sigma^2)$ y tenemos $X_1, X_2, \ldots, X_T$, su media aritmética muestral es:

$$\bar{X}_T = \frac{\sum_1^T X_t}{T}$$

y sabemos que $\bar{X}_T \sim N(\mu, \frac{\sigma^2}{T}) \quad \forall t$
Además la varianza $Var(\bar{X}_T) = \frac{\sigma^2}{T} \longrightarrow 0$ cuando $T \longrightarrow \infty$, por tanto $F_T(\bar{X}_T)$ está concentrada entorno a μ, es decir en el límite es degenerada. Para evitar este problema hacemos transformaciones que nos dan distribuciones límite no degeneradas.
Así formaremos:

$$Z_T = \sqrt{T}(\bar{X}_T - \mu)$$
$$E(Z_T) = \sqrt{T}[E(\bar{X}_T) - E(\mu)] = 0$$
$$V(Z_T) = E[\sqrt{T}(\bar{X}_t - \mu)]^2 = T\frac{\sigma^2}{T} = \sigma^2$$

por tanto

$$Z_t \sim N(0, \sigma^2) \quad \forall t$$
$$\lim_{T\to\infty} F_T(Z) = F(Z)$$

d) Los momentos de la distribución asintótica no son necesariamente igual a los límites de los momentos de las distribuciones de la sucesión de variables aleatorias. Es decir si

$$Z_{(T)} \xrightarrow{d} Z$$

donde Z se distribuye con función de distribución F, $E(Z)$ no tiene por qué ser igual a $\lim_{T\to\infty} E(Z_{(T)})$ ó lo mismo para cualquier otro momento. Puede incluso ocurrir que algunos de los momentos de las variables aleatorias en la sucesión $\{Z_{(T)}\}_{T=1}^{\infty}$ no existan, (es decir, no sean finitos) y en cambio los de la distribución asintótica existan y estén bien definidos (es decir sean finitos).
Es decir, en general:

$$E(\bar{X}_T) = \mu \neq \lim_{T\to\infty} E(\bar{X}_T)$$
$$E(\bar{X}_T - E(\bar{X}_T))^2 = \frac{\sigma^2}{T} \neq \lim_{T\to\infty} E(\bar{X}_T - E(\bar{X}_T))^2$$

Sin embargo, en algunos casos, bajo determinadas condiciones es posible que coincidan.
Si la sucesión $F_1(Z_1), F_2(Z_2), \ldots, F_T(Z_T)$ converge a la distribución límite y si todos los momentos de la sucesión existen y convergen a límites finitos se puede demostrar que la distribución límite tiene momentos finitos de todo orden y que son idénticos a los momentos correspondientes a la sucesión.

e) Llamamos media asintótica al momento no centrado de orden uno de la variable aleatoria a la que converge en distribución la sucesión. Llamamos varianza asintótica al momento centrado de segundo orden de la variable aleatoria a la que converge en distribución la sucesión.

- **Teorema de Cramer**
  Si tenemos dos sucesiones de variables aleatorias $\{Z_{(T)}\}$ y $\{A_{(T)}\}$ donde: $\{Z_{(T)}\}$ es una sucesión de vectores de v.a. de dimensión fija.
  $\{A_{(T)}\}$ es una sucesión de matrices de v.a. de dimensión fija.

$$\begin{array}{llll} Si & Z_{(T)} \xrightarrow{d} Z & Z & vector \ \ de \ \ v.a. \\ & A_{(T)} \xrightarrow{p} A & A & matriz \ \ de \ \ constantes. \end{array}$$

  Entonces:

$$A_{(T)} Z_{(T)} \xrightarrow{d} AZ$$

  Este teorema será de gran utilidad para derivar distribuciones asintóticas de estadísticos para contrastes.

1.7. Teorema de Mann y Wald

Sea X una matriz $(T \times k)$ tal que:

1) Sean $u_1, u_2, \ldots, u_t$ una sucesión de v.a. independientes tal que $\{u_{(T)}\}$ cumple:

   i) $E(u_t) = 0 \quad \forall t$

   ii) $E(u_t^2) = \sigma^2 \quad \forall t$

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

2) $E(X'_{it} u_t) = 0 \quad \forall i = 1, 2, \ldots k$ donde X_i es la columna i-ésima de la matriz X.

3) $plim \left(\frac{1}{T} X'X \right) = Q$ finita, simétrica y definida (+).

Entonces se cumple:

a) $plim \left(\frac{1}{T} X'u \right) = 0$ b) $\frac{1}{\sqrt{T}} X'u \xrightarrow{d} N(0, \sigma^2 Q)$

En este teorema los elementos de la matriz X son variables aleatorias, así en vez de X deberíamos poner $\{X_{(T)}\}$. Si los elementos de X no fueran v.a. entonces: 2) se satisface por i) y 3) sería equivalente a $\lim_{T \to \infty} \left(\frac{1}{T} X'X \right) = Q$ pero se tendrían los mismos resultados a) y b).

Este teorema es condición suficiente de consistencia y existencia de distribución asintótica.

1.7.1. Distribuciones asintóticas

Distribución asintótica normal

Diremos que una sucesión de variables aleatorias Z_T es asintóticamente normal con media m_T y varianza $\sigma^2 \longrightarrow AN(m_T, \sigma^2)$ si la función de distribución de las variables tipificadas $\frac{Z_T - m_T}{\sigma} \xrightarrow{d} N(0,1)$ cuando $T \longrightarrow \infty$.

$$\{Z_{(T)}\}_{T=1}^{\infty} \sim AN(m_T, \sigma^2) \Longleftrightarrow \frac{Z_T - m_T}{\sigma} \xrightarrow{d} N(0,1)$$

$$\Longleftrightarrow \lim_{T \to \infty} \left[Pr \left(\frac{Z_T - m_T}{\sigma} \leq x \right) \right] = \Phi(x)$$

Teorema Central del Límite. (Lindeberg y Levy)

- Caso A:
  Sea $Z_1, Z_2, \ldots, Z_T$ v.a. distribuidas idéntica e independientemente tal que $E(Z_i) = \mu$ y $Var(Z_i) = \sigma^2$, entonces:

  $$\mathbf{Z_T} = \sum_{i=1}^{T} Z_i \xrightarrow{d} N(T\mu, T\sigma^2)$$

  esto se puede expresar:

  $$\frac{\mathbf{Z_T} - m_T}{\sigma} \xrightarrow{d} N(0,1)$$

donde llamamos $m_T = T\mu$

$$\frac{\mathbf{Z_T} - m_T}{\sigma} = \frac{\sum_{i=1}^T Z_i - T\mu}{\sqrt{T}\sigma} = \frac{\bar{\mathbf{Z}}_\mathbf{T} - \mu}{\sigma/\sqrt{T}} = \frac{\sqrt{T}(\bar{\mathbf{Z}}_\mathbf{T} - \mu)}{\sigma}$$

o sea, el Teorema Central del Límite en este caso nos dice:

$$\frac{\sqrt{T}(\bar{\mathbf{Z}}_\mathbf{T} - \mu)}{\sigma} = \frac{1}{\sqrt{T}} \sum_{i=1}^T \left(\frac{Z_i - \mu}{\sigma} \right) \xrightarrow{d} N(0,1)$$

- Caso B:
  Sean $Z_1, Z_2, \ldots, Z_T$ v.a. independientes pero no idénticamente distribuidas, es decir: $E(Z_i) = \mu_i$ y $Var(Z_i) = \sigma_i^2$, $Z_i \sim (\mu_i, \sigma_i^2)$. En este caso el Teorema Central del Límite nos dice:

$$\sum_{i=1}^T Z_i \xrightarrow{d} N(\sum_{i=1}^T \mu_i, \sum_{i=1}^T \sigma_i^2)$$

  o lo que es igual:

$$\frac{\sum_{i=1}^T Z_i - \sum_{i=1}^T \mu_i}{\sqrt{\sum_{i=1}^T \sigma_i^2}} \xrightarrow{d} N(0,1)$$

Observaciones:

a) El Teorema Central del Límite nos da una distribución límite normal para $\sum_{i=1}^T Z_i = \mathbf{Z_T}$ con independencia de la forma de $f(Z_i)$ (función de densidad de Z_i). Sólo imponemos la condición de independencia.

$$\sqrt{T}(\bar{\mathbf{Z}}_\mathbf{T} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

b) Si Z_i es idéntica e independientemente distribuida pero no imponemos la distribución normal el Teorema Central del Límite nos dice:

$$\bar{\mathbf{Z}}_\mathbf{T} \sim N\left(\mu, \frac{\sigma^2}{T}\right) \quad \forall t$$

y por tanto

$$\mathbf{Z_T} = \sqrt{T}(\bar{\mathbf{Z}}_\mathbf{T} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

1.8. Propiedades Asintóticas del estimador MCO en el MRLG

En el Modelo de Regresión Lineal General:

$$Y = X\beta + u$$

bajo los supuestos:

a) X matriz de regresores no estocásticos tal que $rg(X) = k < T$

b) $\text{lím}_{T\to\infty}\left(\frac{1}{T}X'X\right) = Q$, donde Q es una matriz simétrica, definida positiva, finita y no singular.

$$\lim_{T\to\infty}\left(\frac{1}{T}X'X\right) = \begin{bmatrix} 1 & \lim_{T\to\infty}\frac{\sum_{t=1}^{T}X_{2t}}{T} & \dots & \lim_{T\to\infty}\frac{\sum_{t=1}^{T}X_{kt}}{T} \\ & \lim_{T\to\infty}\frac{\sum_{t=1}^{T}X_{2t}^2}{T} & \dots & \lim_{T\to\infty}\frac{\sum_{t=1}^{T}X_{2t}X_{kt}}{T} \\ & & \ddots & \vdots \\ & & & \lim_{T\to\infty}\frac{\sum_{t=1}^{T}X_{kt}^2}{T} \end{bmatrix}$$

c) $E(u) = 0$

d) $E(uu') = \sigma^2 I_T$, σ^2 constante finita.

Se demuestra que muestras finitas:

i) bajo 1) y 3), $E(\hat{\beta}_{MCO}) = \beta$

ii) bajo 1), 3) y 4), $\hat{\beta}_{MCO}$ tiene menor varianza entre los estimadores lineales e insesgados en muestras finitas. $Var(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}$

iii) el estimador de σ^2 definido como $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$ es insesgado bajo 1), 3) y 4), $E(\hat{\sigma}^2) = \sigma^2$

Además vamos a demostrar los siguientes resultados asintóticos.

- **RESULTADO 1:**
  $\hat{\beta}_{MCO}$ es un estimador consistente.
  Bajo 1), 2), 3) y 4) $\hat{\beta}_{MCO}$ es un estimador consistente, es decir, $plim\hat{\beta}_{(T)MCO} = \beta$ para cada parámetro estimado $\hat{\beta}_i$, $plim\hat{\beta}_{(T)i,MCO} = \beta_i \quad \forall i = 1,\dots,k$ podemos demostrarlo de varias formas:

  a) ver si se satisfacen las condiciones suficientes de consistencia, ya que bajo estos supuestos podemos conocer $E(\hat{\beta}_{MCO})$ y $V(\hat{\beta}_{MCO})$.
  b) Ver si $plim\hat{\beta}_{(T)MCO} = \beta$ bajo estos supuestos.

  A) Demostración de la consistencia del estimador MCO por las condiciones suficientes de consistencia.

$$\left.\begin{array}{l} \text{lím}_{T\to\infty} E(\hat{\beta}_{MCO}) = \beta \\ \text{lím}_{T\to\infty} V(\hat{\beta}_{MCO}) = 0 \end{array}\right\}$$

  condiciones suficientes a demostrar que se cumplen.

  a.1) $E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'E(u) = \beta$

$$\Longrightarrow \lim_{T\to\infty} E(\hat{\beta}) = \lim_{T\to\infty}(\beta) = \beta$$

  luego bajo (1) y (3) $\hat{\beta}_{MCO}$ es un estimador asintóticamente insesgado.

  a.2) $Var(\hat{\beta}) = \sigma^2(X'X)^{-1} = \frac{\sigma^2}{T}\left(\frac{1}{T}X'X\right)^{-1}$ y se satisface para cualquier T dado.

$$\lim_{T\to\infty} Var(\hat{\beta}) = \lim_{T\to\infty}\left[\frac{\sigma^2}{T}\left(\frac{1}{T}X'X\right)^{-1}\right]$$

$$= \lim_{T\to\infty}\frac{\sigma^2}{T}\lim_{T\to\infty}\left(\frac{1}{T}X'X\right)^{-1} = 0 \times Q^{-1} = 0$$

luego por las condiciones suficientes

$$plim\hat{\beta}_{(T)MCO} = \beta$$

y $\hat{\beta}_{MCO}$ es un estimador consistente.

B) Demostración de la consistencia de $\hat{\beta}_{MCO}$ utilizando la definición de consistencia.
Para demostrarlo aplicando $plim\hat{\beta}_{MCO}$ no necesitamos conocer $E(\hat{\beta}_{(T)MCO})$ ni $V(\hat{\beta}_{(T)MCO})$.

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

$$plim\hat{\beta}_{MCO} = plim\beta + plim\left[\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)\right]$$

$$= \beta + plim\left(\frac{1}{T}X'X\right)^{-1} plim\left(\frac{1}{T}X'u\right)$$

Sabemos que:

- $plim\left(\frac{1}{T}X'X\right) = Q \Longrightarrow plim\left(\frac{1}{T}X'X\right)^{-1} = Q^{-1}$
- y necesitamos buscar:
$$plim\left(\frac{1}{T}X'u\right) = \begin{bmatrix} plim\frac{1}{T}\sum_{t=1}^{T} u_t \\ plim\frac{1}{T}\sum_{t=1}^{T} X'_{2t}u_t \\ \vdots \\ plim\frac{1}{T}\sum_{t=1}^{T} X'_{kt}u_t \end{bmatrix} = ?$$
- Para buscar $plim\left(\frac{1}{T}X'u\right)$ aplicamos las condiciones suficientes:

$$E\left(\frac{1}{T}X'u\right) = \frac{1}{T}X'E(u) = 0 \quad \forall t$$

(1) $\Longrightarrow \lim_{T\to\infty} E\left(\frac{1}{T}X'u\right) = 0$

$$Var\left(\frac{1}{T}X'u\right) = E\left(\frac{1}{T^2}X'uu'X\right) = \frac{\sigma^2}{T}\frac{X'X}{T}$$

(2) $\lim_{T\to\infty} Var\left(\frac{1}{T}X'u\right) = \lim_{T\to\infty}\frac{\sigma^2}{T}\lim_{T\to\infty}\frac{X'X}{T} = 0 \times Q = 0$

por (1) + (2) $\quad plim\left(\frac{1}{T}X'u\right) = 0$

y por tanto $plim\hat{\beta}_{(T)MCO} = \beta + Q^{-1} \times 0 = \beta$

$\Longrightarrow \hat{\beta}_{MCO}$ es un estimador consistente.

Para buscar $plim\left(\frac{1}{T}X'u\right)$ también podríamos haber hecho:

$$plim\left(\frac{1}{T}X'u\right) = \begin{bmatrix} plim\frac{1}{T}\sum_{t=1}^{T} u_t \\ plim\frac{1}{T}\sum_{t=1}^{T} X'_{2t}u_t \\ \vdots \\ plim\frac{1}{T}\sum_{t=1}^{T} X'_{kt}u_t \end{bmatrix}$$

y buscar cada uno de los límites en probabilidad de la matriz anterior.

a) Por (3) y (4) tenemos que las variables aleatorias $u_1, u_2, \ldots, u_T$ son v.a. tal que $E(u_t) = 0 \quad \forall t$, $E(u_t^2) = \sigma^2 \quad \forall t$ y $E(u_t u_s) = 0 \quad \forall t, s, \quad t \neq s$ luego aplicando las condiciones suficientes de consistencia tenemos:

$$\bar{u}_T = \frac{1}{T}\sum_{t=1}^{T} u_t \xrightarrow{p} 0$$

de donde

$$plim\left(\frac{1}{T}\sum_{t=1}^{T} u_t\right) = plim\left(\bar{u}_{(T)}\right) = 0$$

b) para $j = 2, \ldots, k \qquad plim\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}u_t\right) = 0 \quad$ ya que:

i) por (1), (3) y (4):

$$Var\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}u_t\right) = E\left(\frac{1}{T^2}(\sum_{t=1}^{T} X_{jt}u_t)^2\right) =$$

$$= \frac{1}{T^2}E\left(\sum_{t=1}^{T} X_{jt}^2 u_t^2\right) + \frac{1}{T^2}E\left(\sum\sum_{\neq s} X_{jt}X_{js}u_t u_s\right) =$$

$$= \frac{\sigma^2}{T^2}\sum_{t=1}^{T} X_{jt}^2 = \frac{\sigma^2}{T}\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}^2\right)$$

luego

ii)

$$\lim_{T\to\infty}\frac{\sigma^2}{T}\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}^2\right) = \lim_{T\to\infty}\frac{\sigma^2}{T}\lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}^2\right) = 0 \times q_{jj} = 0$$

ya que hemos supuesto que $\lim_{T\to\infty}\left(\frac{1}{T}X'X\right) = Q$ finito.

$$(a) + (b) \Longrightarrow plim\left(\frac{1}{T}\sum_{t=1}^{T} X_{jt}u_t\right) = 0$$

luego bajo los supuestos (1), (2), (3) y (4)

$$plim\left(\frac{1}{T}X'u\right) = 0$$

$$\text{y} \qquad plim\hat{\beta}_{(T)MCO} = \beta + 0 \times Q^{-1} = \beta$$

- **RESULTADO 2:**
Sea $\hat{\sigma}^2_{MCO} = \frac{\hat{u}'\hat{u}}{T-k}$ donde $\hat{u} = Y - X\hat{\beta}_{MCO}$ es un estimador consistente de σ^2 bajo (1), (2), (3) y (4).

a) Demostración aplicando la definición de consistencia:

$$\hat{\sigma}^2_{MCO} = \frac{\hat{u}'\hat{u}}{T-k} = \frac{u'Mu}{T-k}$$

siendo $M = I - X(X'X)^{-1}X'$

$$\begin{aligned}\hat{\sigma}^2_{MCO} &= \tfrac{1}{T-k}\left[u'[I - X(X'X)^{-1}X']u\right]\\ &= \tfrac{1}{T-k}\left[u'u - u'X(X'X)^{-1}X'u\right]\\ &= \tfrac{T}{T-k}\left[\tfrac{u'u}{T} - \left(\tfrac{1}{T}u'X\right)\left(\tfrac{1}{T}X'X\right)^{-1}\left(\tfrac{1}{T}X'u\right)\right]\end{aligned}$$

$$\begin{aligned} plim\hat{\sigma}^2_{MCO} = & \ plim\left[\frac{T}{T-k}\left(\frac{1}{T}u'u\right)\right] - plim\left[\frac{T}{T-k}\left(\frac{1}{T}u'X\right)\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'u\right)\right] = \\ = & \ \text{lím}_{T\to\infty}\frac{T}{T-k}plim\left(\frac{1}{T}u'u\right) - \\ & \ \text{lím}_{T\to\infty}\frac{T}{T-k}\left[plim\left(\frac{1}{T}u'X\right)plim\left(\frac{1}{T}X'X\right)^{-1}plim\left(\frac{1}{T}X'u\right)\right] \end{aligned}$$

- $\text{lím}_{T\to\infty}\frac{T}{T-k} = \text{lím}_{T\to\infty}\frac{1}{1-k/T} = 1$
- $plim\left(\frac{1}{T}u'X\right) = plim\left(\frac{1}{T}X'u\right) = 0$ demostrado anteriormente.
- $plim\left(\frac{1}{T}X'X\right)^{-1} = Q^{-1}$ por el supuesto (2).
- $plim\frac{1}{T}u'u = plim\frac{1}{T}\sum_{t=1}^{T}u_t^2$ vamos a buscarlo utilizando el siguiente teorema o ley débil de los grandes números:

- **Teorema de Khinchine**:
  Si $Z_t \quad \forall t$ son variables aleatorias independientes, distribuidas con media finita μ, entonces se cumple que:

$$plim\frac{1}{T}\sum_{t=1}^{T}Z_t = \mu$$

  ya que si $E(Z_t) = \mu$ y Z_t son independientes:

$$\frac{1}{T}\sum_{t=1}^{T}Z_t \xrightarrow{p} \mu \qquad \text{ó} \qquad plim\left(\frac{1}{T}\sum_{t=1}^{T}Z_t\right) = \mu$$

  En nuestro caso $Z_t = u_t^2$, $E(u_t^2) = \sigma^2 \quad \forall t$. Dado que $u_1, u_2, \ldots, u_T$ son independientes:

$$\frac{1}{T}\sum_{t=1}^{T}u_t^2 \xrightarrow{p} \sigma^2 \quad \Longleftrightarrow \quad plim\left(\frac{1}{T}\sum_{t=1}^{T}u_t^2\right) = \sigma^2$$

  luego

$$plim\hat{\sigma}^2_{MCO} = \sigma^2 - 0 \times Q^{-1} \times 0 = \sigma^2$$

b) Demostración por las condiciones suficientes de consistencia.
Necesitamos conocer $E(\hat{\sigma}^2_{T,MCO})$ y $V(\hat{\sigma}^2_{T,MCO})$:
Si $u \sim N(0, \sigma^2 I)$ entonces $\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{(T-k)}$

$$\frac{(T-k)\sum_{t=1}^{T}\hat{u}_t^2/(T-k)}{\sigma^2} = \frac{(T-k)\hat{\sigma}^2_{MCO}}{\sigma^2} \sim \chi^2_{(T-k)}$$

ya que $E(\chi^2_{(T-k)}) = T-k$:

$$E\left[(T-k)\frac{\hat{\sigma}^2_{T,MCO}}{\sigma^2}\right] = T-k \implies \frac{(T-k)E(\hat{\sigma}^2_{MCO})}{\sigma^2} = T-k$$
$$\implies E(\hat{\sigma}^2_{MCO}) = \frac{(T-k)\sigma^2}{(T-k)} = \sigma^2$$

ya que $V(\chi^2_n) = 2n$

$$V\left((T-k)\frac{\hat{\sigma}^2_{MCO}}{\sigma^2}\right) = 2(T-k)$$

$$\frac{(T-k)^2}{\sigma^2}V(\hat{\sigma}^2_{MCO}) = 2(T-k) \implies V(\hat{\sigma}^2_{MCO}) = \frac{2(T-k)(\sigma^2)}{(T-k)^2} \implies V(\hat{\sigma}^2) = \frac{2(\sigma^2)}{T-k}$$

$$\left.\begin{array}{l} \lim_{T\to\infty} E(\hat{\sigma}^2_{MCO}) = \sigma^2 \\ \lim_{T\to\infty} V(\hat{\sigma}^2_{MCO}) = \lim_{T\to\infty} \frac{2(\sigma^2)}{T-k} = 0 \end{array}\right\} plim\hat{\sigma}^2_{MCO} = \sigma^2$$

- **RESULTADO 3:**
  - $\hat{\beta}_{(T)MV}$ es un estimador consistente.
    $\hat{\beta}_{MV} = (X'X)^{-1}X'Y = \hat{\beta}_{MCO}$ si u sigue una distribución normal, luego $plim\hat{\beta}_{MV} = \beta$
  - $\hat{\sigma}^2_{MV}$ es consistente.
    $\hat{\sigma}^2_{MV} = \frac{\hat{u}'\hat{u}}{T}$ donde $\hat{u} = Y - X\hat{\beta}_{MV}$
    $E(\hat{\sigma}^2_{MV}) = \frac{1}{T}E(\hat{u}'\hat{u}) = \frac{1}{T}\sigma^2(T-k) \neq \sigma^2$ sesgado en muestras finitas.

$$\begin{aligned} plim\sigma^2_{MV} &= \frac{1}{T}plim(u'Mu) \\ &= \frac{T}{T}\left[plim\left(\frac{1}{T}u'u\right) - plim\left(\frac{1}{T}u'X\right)\left(plim\left(\frac{1}{T}X'X\right)\right)^{-1} plim\left(\frac{1}{T}X'u\right)\right] \\ &= \sigma^2 - 0 \times Q^{-1} \times 0 = \sigma^2 \end{aligned}$$

luego consistente $plim\hat{\sigma}^2_{MV} = \sigma^2$

- **RESULTADO 4:**
  Sea $Y = X\beta + u$ en el modelo de regresión lineal bajo las hipótesis:

  (1) X matriz de regresores fijos $rg(X) = k < T$
  (2) $\lim_{T\to\infty}\left(\frac{1}{T}X'X\right) = Q$
  (3) $E(u) = 0$
  (4) $E(uu') = \sigma^2 I_T$
  (5) u se distribuye con función de distribución normal multivariante.

  bajo (1), (2), (3), (4) y (5) tenemos los siguientes resultados para muestras finitas (es decir para un tamaño de muestra dado):

  (i) $(\hat{\beta}_{MCO} - \beta) \sim N(0, \sigma^2(X'X)^{-1})$
  (ii) $\frac{(T-k)\hat{\sigma}^2_{MCO}}{\sigma^2} \sim \chi^2_{(T-k)}$
  También bajo estos supuestos tenemos el siguiente resultado asintótico:
  (iii) $\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

  Hemos visto que bajo (1), (2), (3) y (4) $\hat{\beta}_{MCO} \xrightarrow{p} \beta$ donde β es un vector de constantes que son valores desconocidos de los parámetros poblacionales.

  Entonces $(\hat{\beta}_{MCO} - \beta) \xrightarrow{p} 0 \Longrightarrow (\hat{\beta}_{MCO} - \beta) \xrightarrow{d} 0$ y por tanto la distribución asintótica de $(\hat{\beta}_{MCO} - \beta)$ es una distribución degenerada, es decir, la matriz de varianzas y covarianzas asintótica es cero y toda la masa de probabilidad está concentrada en el punto cero. Dada esta característica podemos pensar que esta distribución asintótica no es muy interesante por lo que realizaremos la transformación $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ para obtener una distribución no degenerada.

  Si $u \sim N(0, \sigma^2 I)$ distribución exacta para cualquier T dado.
  $(\hat{\beta}_{MCO} - \beta) \sim N(0, \sigma^2(X'X)^{-1})$ distribución exacta para cualquier T dado.
  $\sqrt{T}(\hat{\beta}_{MCO} - \beta) \sim N\left(0, \sigma^2\left(\frac{1}{T}X'X\right)^{-1}\right)$ distribución exacta para cualquier T dado.
  ya que:

$$E(\sqrt{T}(\hat{\beta}_{MCO} - \beta)) = \sqrt{T}E[(\hat{\beta}_{MCO} - \beta)] = 0$$

$$V(\sqrt{T}(\hat{\beta}_{MCO}-\beta)) = E[T(\hat{\beta}_{MCO}-\beta)(\hat{\beta}_{MCO}-\beta)'] = \sigma^2 T(X'X)^{-1} = \sigma^2 \left(\frac{1}{T}X'X\right)^{-1}$$

Entonces si miramos a la sucesión de vectores de variables aleatorias con sus funciones de distribución asociadas:

$$\sqrt{T_1}(\hat{\beta}_{T_1,MCO}-\beta) \sim N\left(0, \sigma^2\left(\frac{1}{T_1}X'X\right)^{-1}\right)$$

$$\sqrt{T_2}(\hat{\beta}_{T_2,MCO}-\beta) \sim N\left(0, \sigma^2\left(\frac{1}{T_2}X'X\right)^{-1}\right) \quad T_1 < T_2$$

$$\sqrt{T_3}(\hat{\beta}_{T_3,MCO}-\beta) \sim N\left(0, \sigma^2\left(\frac{1}{T_3}X'X\right)^{-1}\right) \quad T_2 < T_3$$

$$\vdots$$

vemos que $\{\sqrt{T}(\hat{\beta}_{MCO}-\beta)\}$ converge en distribución a un vector de v.a. con función de distribución $N\left(0, \sigma^2\left(\lim_{T\to\infty}\left(\frac{1}{T}X'X\right)\right)^{-1}\right)$ y lo denotamos como:

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

Comentarios:

a) Dado que $u \sim N(,)$ para cualquier T dado, podemos conocer la función de distribución exacta para cada uno de esos tamaños de muestra.

b) Si consideramos la sucesión:

$$(\hat{\beta}_{T_1,MCO}-\beta) \sim N\left(0, \frac{\sigma^2}{T_1}\left(\frac{1}{T_1}X'X\right)^{-1}\right)$$

$$(\hat{\beta}_{T_2,MCO}-\beta) \sim N\left(0, \frac{\sigma^2}{T_2}\left(\frac{1}{T_2}X'X\right)^{-1}\right) \quad T_1 < T_2$$

$$(\hat{\beta}_{T_3,MCO}-\beta) \sim N\left(0, \frac{\sigma^2}{T_3}\left(\frac{1}{T_3}X'X\right)^{-1}\right) \quad T_2 < T_3$$

$$\vdots$$

cuando $T \to \infty \quad (\hat{\beta}_{MCO}-\beta) \xrightarrow{d} 0$

- **RESULTADO 5**: (Relajación del supuesto de normalidad)
Si relajamos el supuesto (5), bajo (1), (2), (3) y (4) sin el supuesto de que u se distribuye como función de distribución normal (no especificamos su función de distribución) entonces no podemos conocer la distribución exacta para un tamaño muestral dado T, de $\hat{\beta}_{MCO}$, es decir, no tenemos el resultado (i). Tendremos el siguiente resultado asintótico:

$$(iii) \qquad \sqrt{T}(\hat{\beta}_{MCO}-\beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

siendo (i) $(\hat{\beta}_{MCO}-\beta) \sim N(0, \sigma^2(X'X)^{-1})$
El vector de variables aleatorias $\sqrt{T}(\hat{\beta}_{MCO}-\beta)$ converge en distribución a un vector de v.a que se distribuye normal de media cero y matriz de varianzas y covarianzas $\sigma^2 Q^{-1}$. Luego si el tamaño muestral es suficientemente grande podemos considerar esta distribución

asintótica como una buena aproximación a la distribución exacta de $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ para ese tamaño muestral T. Demostración de (iii):
Bajo los supuestos (1), (2), (3) y (4):

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

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

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

Aplicando el Teorema de Mann y Wald y considerando X matriz de regresores fijos, tenemos que:

$$\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, \sigma^2 Q)$$

es decir, converge en distribución a un vector de v.a. Z que se distribuye $N(0, \sigma^2 Q)$ donde $Q = \lim_{T\to\infty}\left(\frac{1}{T}X'X\right)$.

Dado que:

$$\lim_{T\to\infty}\left(\frac{1}{T}X'X\right)^{-1} = Q^{-1} \Longleftrightarrow \left(\frac{1}{T}X'X\right)^{-1} \Longrightarrow Q^{-1}$$

y

$$\frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma^2 Q)$$

Aplicando el Teorema de Cramer tenemos que la sucesión de vectores de variables aleatorias:

$$\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} Q^{-1}Z$$

donde $Z \sim N(0, \sigma^2 Q)$, luego $Q^{-1}Z \sim N(0, \sigma^2 Q^{-1})$ ya que:

$$E(Q^{-1}Z) = Q^{-1}E(Z) = 0$$

$$V(Q^{-1}Z) = E((Q^{-1}Z)(Q^{-1}Z)') = Q^{-1}E(ZZ')Q^{-1} =$$
$$= Q^{-1}(\sigma^2 Q)Q^{-1} = \sigma^2 Q^{-1}QQ^{-1} = \sigma^2 Q^{-1}$$

luego podemos demostrar también que:

$$\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

y dado que:

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

entonces

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

Este resultado nos va a permitir hacer inferencia sobre β basándonos en esta distribución asintótica de $\sqrt{T}(\hat{\beta}_{MCO} - \beta)$ cuando no conozcamos la distribución exacta para un tamaño de muestra dado (del que nosotros disponemos) y que será una aproximación mejor cuanto mayor sea el tamaño muestral.

- **RESULTADO 6:**

$$\sqrt{T}(\hat{\sigma}^2_{T,MCO} - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$$

donde μ_4 es el cuarto momento finito de u. Si u es normal $\mu_4 - \sigma^4 = 2\sigma^4$ de donde si u es normal

$$\sqrt{T}(\hat{\sigma}^2_{T,MCO} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$$

Demostración:

$$\frac{\hat{\sigma}^2_{T,MCO}}{\sigma^2}(T-k) \sim \chi^2_{(T-k)}$$

$$(T-k)\frac{\hat{\sigma}^2_{T,MCO}}{\sigma^2} \xrightarrow{d} N((T-k), 2(T-K))$$

trabajamos con la transformación $\frac{\sigma^2}{T-k}\frac{T-k}{\sigma^2}\hat{\sigma}^2$ donde $\frac{T-k}{\sigma^2}$ es una constante.

$$\frac{\sigma^2}{T-k}\frac{T-k}{\sigma^2}\hat{\sigma}^2 \xrightarrow{d} N\left(\frac{\sigma^2}{T-k}(T-k), 2(T-k)\frac{\sigma^4}{(T-K)^2}\right)$$

$$\hat{\sigma}^2_{T,MCO} \xrightarrow{d} N\left(\sigma^2, \frac{2\sigma^4}{T-k}\right)$$

$$\sqrt{T-k}(\hat{\sigma}^2_{T,MCO} - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$$

en el límite, si $T \to \infty$ $\quad \sqrt{T} \simeq \sqrt{T-k}$ por tanto

$$\sqrt{T}(\hat{\sigma}^2_{T,MCO} - \sigma^2) \xrightarrow{p} N(0, 2\sigma^4) \quad \Longleftrightarrow \quad \hat{\sigma}^2_{T,MCO} \overset{a}{\sim} N\left(\sigma^2, \frac{2\sigma^4}{T}\right)$$

1.9. Contraste de hipótesis

Contraste de hipótesis de la forma $H_0 : R\beta = r$ en el modelo de regresión lineal $Y = X\beta + u$ bajo los supuestos: $E(u) = 0; E(uu') = \sigma^2 I_T, \text{lím}_{T\to\infty}\left(\frac{1}{T}X'X\right) = Q$ finita, simétrica, definida positiva y no singular, pero no especificamos la función de distribución de u (en particular no suponemos normalidad).

- Caso 1: Una restricción lineal
  - R es una matriz de constantes conocidas $(q \times k)$.
  - r es un vector de constantes conocidas $(q \times 1)$.

  Un caso particular de $H_0 : R\beta = r$ es $H_0 : \beta_i = 0$. Si no suponemos normalidad, no conocemos la distribución exacta. Bajo la hipótesis nula el estadístico:

$$\frac{R\hat{\beta}_{MCO} - r}{\hat{\sigma}\sqrt{R(X'X)^{-1}R'}}$$

  en general no se distribuirá como una t-Student, pero vamos a demostrar que la sucesión de este estadístico cuando el tamaño muestral tiende a infinito converge en distribución a una v.a. con distribución $N(0,1)$.

$$\frac{R\hat{\beta}_{MCO} - r}{\hat{\sigma}\sqrt{R(X'X)^{-1}R'}} \xrightarrow{d} N(0,1)$$

Luego para un tamaño de muestra T dado si éste es suficientemente grande podemos aproximar la distribución exacta de este estadístico por la distribución asintótica $N(0,1)$.

Por tanto, para un nivel de significación elegido α, no aceptaremos la hipótesis nula $H_0 : R\beta = r$ si el valor obtenido dada nuestra muestra de este estadístico es mayor que el valor crítico mirando a las tablas de $N(0,1)$. Para $H_0 : \beta_i = 0$ el estadístico de contraste sería:

$$t_c = \frac{\hat{\beta}_{i,MCO}}{\hat{\sigma}\sqrt{a_{ii}}}$$

donde a_{ii} es el elemento i-ésimo de la matriz $(X'X)^{-1}$ es decir $\hat{\sigma}_{\hat{\beta}_i}$
Demostración:

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

como R es un vector de constantes de dimensión fija:

$$\sqrt{T}(R\hat{\beta}_{MCO} - R\beta) \xrightarrow{d} N(0, \sigma^2 RQ^{-1}R')$$

donde $RQ^{-1}R'$ es un escalar. Dado que, R es un vector de constantes que no depende de T

$$R\left(\frac{1}{T}X'X\right)^{-1} R' \xrightarrow{T\to\infty} RQ^{-1}R'$$

Bajo estos supuestos $\hat{\sigma}^2_{MCO}$ es un estimador consistente de σ^2, así: $\hat{\sigma}^2_{MCO} \xrightarrow{p} \sigma^2$. Por las propiedades del operador plim:

$$\hat{\sigma}^2_{MCO} R\left(\frac{1}{T}X'X\right)^{-1} R' \xrightarrow{p} \sigma^2 RQ^{-1}R'$$

y por el teorema de Slustky:

$$\sqrt{\hat{\sigma}^2_{MCO} R\left(\frac{1}{T}X'X\right)^{-1} R'} \xrightarrow{p} \sqrt{\sigma^2 RQ^{-1}R'}$$

y dado que bajo $H_0 : R\beta = r$ $\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} Z \sim N(0, \sigma^2 Q^{-1})$ aplicando el teorema de Cramer

$$\frac{1}{\hat{\sigma}^2_{MCO} R\left(\frac{1}{T}X'X\right)^{-1} R'} \sqrt{T}(R\hat{\beta}_{MCO} - r) \xrightarrow{d} \frac{1}{a}Z$$

donde $\frac{1}{a}Z$ se distribuye $N(0,1)$, dado que a es una constante y $Z \sim N(0, a^2)$

$$E\left(\frac{1}{a}Z\right) = \frac{1}{a}E(Z) = 0$$

$$V\left(\frac{1}{a}Z\right) = E\left(\frac{1}{a^2}ZZ'\right) = \frac{1}{a^2}E(ZZ') = \sigma^2 \frac{1}{a^2}(RQ^{-1}R')$$

$$= \sigma^2 \frac{1}{\sigma^2(RQ^{-1}R')}(RQ^{-1}R') = 1$$

luego el estadístico:

$$\frac{\sqrt{T}(R\hat{\beta}_{MCO} - r)}{\sqrt{T}\hat{\sigma}_{MCO}\sqrt{R(X'X)^{-1}R'}} \xrightarrow{d} N(0,1)$$

bajo $H_0 : R\beta = r$.

Por tanto si no especificamos la distribución de u, en particular no especificamos que sea normal, no conocemos la distribución exacta de este estadístico para un tamaño de muestra dado, pero si la muestra es suficientemente grande podemos aproximarla por una $N(0,1)$.

Para un nivel de significación α elegiremos el valor crítico con el que comparar el valor obtenido del estadístico, mirando las tablas de la normal $N(0,1)$. Por ejemplo, a un nivel $\alpha = 5\,\% \quad \alpha = 0{,}05) \quad \frac{\alpha}{2} = 0{,}025$ contraste a dos colas $\Phi(1{,}96) = 1 - 0{,}025 = 0{,}975$.

No aceptamos H_0 al nivel de significación del 5 % si el valor absoluto del estadístico obtenido con la muestra utilizada es mayor que el valor crítico 1.96.

- Caso 2: En general, q restricciones lineales. R es una matriz de constantes conocidas $(q \times k)$. r es un vector de constantes conocidas $(q \times 1)$. Si no suponemos normalidad de u, el estadístico bajo la $H_0 : R\beta = r$ es:

$$F_c = \frac{(R\hat{\beta}_{MCO} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2_{MCO}}$$

no tiene porqué distribuirse como una F de Snedecor con $(q, T-k)$ grados de libertad, ni tampoco sabemos como se distribuye para un tamaño de muestra dado T.

$$\frac{(R\hat{\beta}_{MCO} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2_{MCO}} \xrightarrow{d} \chi^2_{(q)}$$

donde $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$ es un estimador consistente de σ^2, es decir $\hat{\sigma}^2_{MCO} \xrightarrow{p} \sigma^2$, luego puedo considerar, bajo estos supuestos

$$\hat{\sigma}^2_{MCO} = \frac{\hat{u}'\hat{u}}{T-k} \qquad \text{ó} \qquad \hat{\sigma}^2_{MV} = \frac{\hat{u}'\hat{u}}{T}$$

Demostración: Bajo $H_0 : R\beta = r$:

$$\sqrt{T}(R\hat{\beta}_{MCO} - r) \xrightarrow{d} N(0, \sigma^2(RQ^{-1}R'))$$

donde $RQ^{-1}R'$ es una matriz $(q \times q)$. Por el teorema de Cramer:

$$\sqrt{T}(R\hat{\beta}_{MCO} - r)' \left[\sigma^2 R \left(\frac{1}{T}X'X\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{MCO} - r) \xrightarrow{d} \chi^2_{(q)}$$

si $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-k}$ es un estimador consistente de σ^2 por el Teorema de Cramer:

$$\sqrt{T}(R\hat{\beta}_{MCO} - r)' \left[\hat{\sigma}^2_{MCO} R \left(\frac{1}{T}X'X\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{MCO} - r) \xrightarrow{d} \chi^2_{(q)}$$

o lo que es igual

$$\frac{(R\hat{\beta}_{MCO} - r)' \left[R \left(\frac{1}{T}X'X\right)^{-1} R'\right]^{-1} (R\hat{\beta}_{MCO} - r)}{\hat{\sigma}^2_{MCO}} \xrightarrow{d} \chi^2_{(q)}$$

Por lo tanto si el tamaño de muestra es suficientemente grande podemos utilizar este estadístico y aproximar su distribución por la distribución asintótica $\chi^2_{(q)}$.

No aceptaremos la hipótesis nula si el valor del estadístico obtenido para la muestra utilizada es mayor que un valor crítico, elegido un valor de significación α; mirando a las tablas de $\chi^2_{(q)}$, este valor crítico $\bar{Z}$ será aquel que.

$$Pr\{Z \leq \bar{Z}\} = 1 - \alpha \qquad \text{donde} \quad Z \sim \chi^2_{(q)}$$

Tema 2

Generalización del Modelo de Regresión Lineal

2.1. Modelo de regresión con perturbaciones no esféricas

En el modelo de regresión lineal general

$$Y = X\beta + u \tag{2.1}$$

donde X es no estocástica y sobre las perturbaciones suponemos:

a) Homocedasticidad: $var(u_t) = \sigma^2$, $\forall\ t$.

b) No autocorrelación: $cov(u_t, u_s) = 0$, $\forall\ t \neq s$.

lo que se conoce como perturbaciones esféricas, podemos escribir la matriz de varianzas y covarianzas de la perturbación como:

$$E(uu') = \quad \sigma^2 I = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

En este tema relajamos estos supuestos permitiendo que exista heterocedasticidad:

$$var(u_t) = \sigma_t^2$$

y/o autocorrelación:

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

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.

En general permitimos que las perturbaciones tengan una matriz de varianzas y covarianzas no escalar:

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

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

donde

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

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

• **Ejemplo 1:** Supongamos una muestra de observaciones relativas a gastos de consumo familiares, C_i, y renta disponible, R_i, 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 y σ^2 será grande mientras que para las rentas bajas σ^2 será pequeña. En este supuesto la varianza de la perturbación cambia según la renta de las familias, existe heterocedasticidad y podemos escribirla:

$$E(u_i^2) = \sigma_i^2 \quad \text{sección cruzada}$$

$$E(u_t^2) = \sigma_t^2 \quad \text{serie temporal}$$

La matriz de varianzas y covarianzas de la perturbación sería:

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

donde suponemos $E(u_i) = 0 \ \forall i, \quad E(u_i^2) = \sigma_i^2, \quad E(u_i u_j) = 0 \ \ \forall i, j \ \ i \neq j$

• **Ejemplo 2:** Supongamos que investigamos la relación entre la tasa de inflación, π_t, y el incremento en el nivel de salarios, W_t, para un conjunto de años T. La indiciación salarial nos indica que el nivel de salarios fijado para el periodo t dependerá del nivel de inflación del periodo anterior. Así, lo que ocurre en un período actual depende de lo ocurrido en el periodo pasado y será difícil mantener $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$. Ocurrirá lo contrario, que las perturbaciones están correladas, así:

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

En este caso, y suponiendo que la varianza es constante, escribimos la matriz de varianzas y covarianzas de la perturbación como:

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

• **Ejemplo 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:

$$Y_i = \alpha_1 + \beta_1 X_i + u_i \qquad u_i \sim N(0, \sigma_1^2 I_{N_1})$$

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

$$Y_j = \alpha_2 + \beta_2 X_j + u_j \qquad u_j \sim N(0, \sigma_2^2 I_{N_2})$$

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:

$$\begin{bmatrix} Y_i \\ Y_j \end{bmatrix} = \begin{bmatrix} i_{N_1} & 0 & X_i \\ 0 & i_{N_2} & X_j \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} + \begin{bmatrix} u_i \\ u_j \end{bmatrix} \Longleftrightarrow Y = X\beta + u$$

y la matriz de varianzas y covarianzas del sistema de ecuaciones a estimar es:

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

donde al ser $\sigma_1^2 \neq \sigma_2^2$ es heterocedástica. Además estamos suponiendo que u_i y u_j son independientes, pero también podemos suponer que son dependientes.

2.2. Propiedades del estimador MCO

Sea el MRLG, $Y = X\beta + u$, donde se mantienen las hipótesis básicas salvo que:

$$E(uu') = \Sigma = \sigma^2 \Omega, \text{ donde } \Omega \neq I \tag{2.2}$$

Propiedades de $\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$:

a) Lineal: dado que X es no estocástica el estimador MCO es lineal en u.

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

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

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

c) Matriz de varianzas y covarianzas:

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

tal que

$$\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}$$

El estimador MCO **no** es el estimador con varianza mínima entre los estimadores lineales e insesgados.

d) Distribución en muestras finitas:
Si las perturbaciones tienen una distribución normal $u \sim N(0, \Sigma)$

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

Si las perturbaciones tienen una distribución normal $u \sim N(0, \sigma^2\Omega)$

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

e) Consistente. Vamos a demostrar la consistencia del estimador por las condiciones suficientes:

Sean

$$\lim_{T\to\infty} \frac{X'X}{T} = Q \quad \text{finita, semidefinida positiva y no singular}$$

$$\lim_{T\to\infty} \frac{X'\Omega X}{T} = Z \quad \text{finita, semidefinida positiva y no singular}$$

Entonces:

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

y así, por las condiciones suficientes de consistencia:

$$\text{plim}\hat{\beta}_{MCO} = \beta$$

Resumiendo, el estimador de MCO bajo perturbaciones no esféricas es lineal en la perturbación, insesgado y consistente pero no es de varianza mínima.

2.2.1. Estimador de σ^2 e inferencia

En este caso el estimador de σ^2, $\hat{\sigma}^2 = \dfrac{\hat{u}\hat{u}'}{T-K}$ es sesgado:

$$E(\hat{\sigma}^2) = \frac{E(\hat{u}\hat{u}')}{T-K} = \frac{\sigma^2 tr(M\Omega)}{T-K} \neq \sigma^2$$

Este estimador sesgado no es válido para hacer inferencia.

Consecuencias para la inferencia: Los estadísticos t y F habituales ahora no tienen las distribuciones t-student y F-Snedecor habituales por lo tanto la inferencia en base a estos estadísticos no es válida.

Por todo ello parece más adecuado buscar un nuevo estimador que tuviera una matriz de varianzas y covarianzas menor que $\sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$. En particular podemos encontrar un estimador que en circunstancias donde $E(uu') = \sigma^2\Omega$ (bien heterocedasticidad, bien autocorrelación, bien ambos) sea lineal, insesgado, de varianza mínima y consistente. Este estimador es el de Mínimos Cuadrados Generalizados, y a su vez permitiría proponer un estimador insesgado para σ^2 y realizar inferencia válida.

2.3. Método de Mínimos Cuadrados Generalizados (MCG)

Supongamos que en el MRLG $Y = X\beta + u$ conocemos $E(uu') = \Sigma = \sigma^2\Omega$. Si lo que queremos es estimar los coeficientes β desconocidos de forma que el estimador sea eficiente, una manera sensata de proceder es transformar el modelo (2.1) en otro con perturbaciones esféricas, de forma que podamos proceder como lo hacíamos hasta ahora.

• *Resultado 1:* Dado que $\Sigma = \sigma^2\Omega$ es simétrica y semidefinida positiva, existe una matriz no singular P tal que $\Omega = PP'$.

La inversa de la matriz P se utiliza como matriz de transformación del modelo original dado que

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

Premultiplicando el modelo (2.1) por P^{-1} obtenemos el siguiente modelo transformado:

$$\underbrace{P^{-1}Y}_{Y_*} = \underbrace{P^{-1}X}_{X_*}\beta + \underbrace{P^{-1}u}_{u_*}$$

$$Y_* = X_*\beta + u_* \tag{2.3}$$

Este modelo transformado tiene perturbaciones esféricas, u_*:

$$\begin{aligned} E(u_*) &= E(P^{-1}u) = P^{-1}E(u) = 0 \\ E(u_*u_*') &= E(P^{-1}uu\mathrm{tr}(P^{-1})\mathrm{tr}) = P^{-1}E(uu\mathrm{tr})(P^{-1})\mathrm{tr} = \\ &= \sigma^2P^{-1}\Omega P^{-1}\mathrm{tr} = \sigma^2P^{-1}PP\mathrm{tr}(P\mathrm{tr})^{-1} = \sigma^2 I. \end{aligned}$$

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.

$$\begin{aligned}
\hat{\beta}_{MCO} &= (X_*'X_*)^{-1}X_*'Y_* = &(2.4)\\
&= (X^{'}(P\mathrm{tr})^{-1}P^{-1}X)^{-1}X^{'}(P\mathrm{tr})^{-1}P^{-1}Y = &\\
&= (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}Y = &(2.5)\\
&= (X^{'}\Sigma^{-1}X)^{-1}X^{'}\Sigma^{-1}Y = \tilde{\beta}_{MCG} &(2.6)
\end{aligned}$$

Si Ω (o Σ) es conocida el estimador es inmediatamente calculable. En resumen, tenemos dos alternativas para estimar los coeficientes de un modelo por MCG:

- Aplicar el estimador MCG (ecuación (2.5) o (2.6)) al modelo original.
- Aplicar el estimador MCO al modelo transformado (ecuación (2.4)).

2.3.1. Propiedades de los estimadores MCG

Las propiedades del estimador MCG podemos demostrarlas alternativamente en el modelo transformado utilizando la expresión (2.4) o en el modelo original utilizando (2.5) o (2.6).

a) **Lineal** en la perturbación u dado que X es no estocástica:

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

b) **Insesgado:** Dado que X es no estocástica, Ω es conocida y $E(u) = 0$ el estimador MCG es insesgado:

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

c) **Matriz de varianzas y covarianzas:**

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

El **Teorema de Gauss-Markov** aplicado en el modelo transformado, garantiza que el estimador MCO es el estimador con varianza mínima entre los estimadores lineales e insesgados. Del mismo modo, por el **Teorema de Aitken** aplicado al modelo original, garantiza que el estimador MCG es el estimador con varianza mínima entre los estimadores lineales e insesgados.

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

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

Si $u \sim N(0,\, \Sigma)$, tenemos que

$$\tilde{\beta}_{MCO} \sim N(\beta,\, (X^{'}\Sigma^{-1}X)^{-1})$$

e) **Consistencia**:
Sea plim$\frac{X'\Omega^{-1}X}{T} = G$ finita, semidefinida positiva y no singular

Por las condiciones suficientes de consistencia: plim$\tilde{\beta}_{MCG} = \beta$, con lo que el estimador es consistente.

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

- Si buscamos las propiedades en el modelo transformado tenemos:

a) Linealidad en u_*, dado que X_* es no estocástica el estimador es lineal en u_*:

$$\hat{\beta} = \beta + (X_*'X_*)^{-1}(X_*'u_*)$$

b) Insesgado: Dado que X_* es no estocástica y $E(u_*) = 0$ el estimador es insesgado:

$$\begin{aligned}
E(\hat{\beta}) &= \beta + E[(X_*'X_*)^{-1}X_*'u_*] = \\
&= \beta + (X_*'X_*)^{-1}X_*'E(u_*) = \beta
\end{aligned}$$

c) **Matriz de varianzas y covarianzas:**

$$\begin{aligned}
V(\hat{\beta}) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] \\
&= E[(X_*'X_*)^{-1}X_*'u_*u_*'X_*(X_*'X_*)^{-1}] = \\
&= (X_*'X_*)^{-1}X_*'E(u_*u_*')X_*(X_*'X_*)^{-1} = \\
&= \sigma^2(X_*'X_*)^{-1} \\
&= \sigma^2(X'\Omega^{-1}X)^{-1} = (X'\Sigma^{-1}X)^{-1}
\end{aligned}$$

El **Teorema de Gauss-Markov** aplicado en el modelo transformado, garantiza que el estimador MCO es el estimador con varianza mínima entre los estimadores lineales e insesgados.

d) Bajo el supuesto de normalidad de las perturbaciones, $u_* \sim N(0, \sigma^2 I)$, tenemos que

$$\hat{\beta} \sim N(\beta, \sigma^2(X_*'X_*)^{-1})$$

e) Consistente: Sea plim$\frac{X_*'X_*}{T} = Q_* = G$ finita, semidefinida positiva y no singular por las condiciones suficientes de consistencia: plim$\hat{\beta}_{MCG} = \beta$.

$$\begin{aligned}
\lim_{T\to\infty} E(\tilde{\beta}_{MCG}) &= \beta \\
\lim_{T\to\infty} \text{Var}(\tilde{\beta}_{MCG}) &= \lim_{T\to\infty} \frac{\sigma^2}{T}\left(\frac{X_*'X_*}{T}\right)^{-1} = 0 \times G^{-1} = 0
\end{aligned}$$

2.3.2. Distribución Asintótica

Dado que

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

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

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

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_* es no estocástica por lo que podemos aplicar Mann-Wald:

i) Sea $u_* \sim iid(0, \sigma^2 I)$

ii) Sea $E(X_*'u_*) = X_*'E(u_*) = 0$

iii) Sea $plim\left(\frac{X_*'X_*}{T}\right) = Q_* = G$ finita y no singular

Entonces se cumplen los dos resultados siguientes:

1. $plim\left(\frac{X_*'u_*}{T}\right) = 0$
2. $\frac{X_*'u_*}{\sqrt{T}} \xrightarrow{d} N(0,\, \sigma^2 G)$

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

$$\begin{aligned}
\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{X_*'X_*}{T}\right)^{-1} \frac{X_*'u_*}{\sqrt{T}} \quad &\xrightarrow{d} \quad N\left(0, \sigma^2 plim\left(\frac{X_*'X_*}{T}\right)^{-1} G\, plim\left(\frac{X_*'X_*}{T}\right)^{-1}\right) \\
&\xrightarrow{d} \quad N(0,\, \sigma^2 G^{-1}\, G\, G^{-1}) \\
&\xrightarrow{d} \quad N(0,\, \sigma^2 G^{-1})
\end{aligned}$$

por lo que

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

2.3.3. Estimador de σ^2

Si $E(uu') = \sigma^2\Omega$ en general σ^2 será desconocida y habremos de estimarla. Al igual que los coeficientes del modelo podremos estimarla en el modelo original o en el transformado.

- En el modelo transformado, un estimador insesgado y consistente sería:

$$\tilde{\sigma}^2 \;=\; \frac{\hat{u}_*'\hat{u}_*}{T-K} = \frac{(Y_* - X_*\tilde{\beta}_{MCG})'(Y_* - X_*\tilde{\beta}_{MCG})}{T-K} = \frac{Y_*'Y_* - \tilde{\beta}_{MCG}'X_*'Y_*}{T-K}$$

- En el modelo original un estimador insesgado y consistente de σ^2 bajo el supuesto de que Ω es conocida sería:

$$
\begin{aligned}
\tilde{\sigma}^2_{MCO} &= \frac{\tilde{u}^{\text{tr}}_{MCG}\Omega^{-1}\tilde{u}_{MCG}}{T-K} = \\
&= \frac{(Y - X\tilde{\beta}_{MCG})'\Omega^{-1}(Y - X\tilde{\beta}_{MCG})}{T-K} = \frac{Y'\Omega^{-1}Y - \tilde{\beta}'_{MCG}X'\Omega^{-1}Y}{T-K}
\end{aligned}
$$

• **Función objetivo:** Notar que para el estimador de MCG la función objetivo en un marco donde $Y = X\beta + u \quad E(u) = 0 \quad E(uu') = \sigma^2\Omega \quad X$ fija sería:

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

$$\Longleftrightarrow Min_\beta\, Y'\Omega^{-1}Y - 2\tilde{\beta}'_{MCG}X'\Omega^{-1}Y + \tilde{\beta}'_{MCG}X'\Omega^{-1}X\tilde{\beta}_{MCG}$$

$$\frac{\partial \hat{u}'\Omega^{-1}\hat{u}}{\partial \tilde{\beta}_{MCG}}|_{=0}; \qquad -2X'\Omega^{-1}Y + 2X'\Omega^{-1}X\tilde{\beta}_{MCG} = 0$$

las ecuaciones normales son:

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

de donde

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

y podemos estimar la varianza de la perturbación como:

$$\tilde{\sigma}^2_{MCG} = \frac{\hat{u}'\Omega^{-1}\hat{u}}{T-K} = \frac{Y'\Omega^{-1}Y - \tilde{\beta}'_{MCG}X'\Omega^{-1}Y}{T-K}$$

2.4. Método de Mínimos Cuadrados Generalizados Factibles (MCGF)

Hasta ahora hemos supuesto que conocíamos $E(uu') = \Sigma = \sigma^2\Omega$ ó al menos Ω. El estimador de MCG en este caso es lineal, insesgado y de varianza mínima. Pero en la práctica la mayoría de las veces Ω o Σ son desconocidas. En este caso el estimador MCG no es directamente calculable. La solución habitual es sustituir Ω (o Σ) por una estimación suya en la expresión del estimador de MCG dando lugar al estimador MCGF:

$$
\begin{aligned}
\tilde{\beta}_{MCGF} &= (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}Y \\
&= (X'\widehat{\Sigma}^{-1}X)^{-1}X'\widehat{\Sigma}^{-1}Y
\end{aligned} \tag{2.7}
$$

2.4.1. Propiedades del estimador de MCGF

Bajo el supuesto de que las varianzas de las perturbaciones se han modelado correctamente, tenemos las siguientes propiedades:

Propiedades en muestras finitas:

a) $\tilde{\beta}_{MCGF}$ no es lineal en u.

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

donde u y $\widehat{\Omega}$ son variables aleatorias y por tanto $\tilde{\beta}_{MCGF}$ es una combinación no lineal de variables aleatorias.

b) En general $\tilde{\beta}_{MCGF}$ es sesgado:

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

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: $E[\cdot] \neq 0$ y por tanto $E(\tilde{\beta}_{MCGF}) \neq \beta$.

c) Matriz de varianzas y covarianzas de $\tilde{\beta}_{MCGF}$:

$$V(\tilde{\beta}_{MCGF}) = E[\tilde{\beta}_{MCGF} - E(\tilde{\beta}_{MCGF})][\tilde{\beta}_{MCGF} - E(\tilde{\beta}_{MCGF})]'$$

expresión difícil de obtener.

Propiedades asintóticas: Dado que generalmente no podemos decir nada de las propiedades en muestras finitas buscaremos propiedades en muestras grandes:

a) Consistencia. Necesitamos que:

$$\text{plim}\tilde{\beta}_{MCGF} = \text{plim}\tilde{\beta}_{MCG} = \beta$$

Tenemos que demostrar que $\text{plim}[\tilde{\beta}_{MCGF} - \tilde{\beta}_{MCG}] = 0$

Dado que:

$$\begin{aligned}
\text{plim}\tilde{\beta}_{MCGF} &= \beta + \text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}u\right)\\
\text{plim}\tilde{\beta}_{MCG} &= \beta + \text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)
\end{aligned}$$

entonces

$$\text{plim}[\tilde{\beta}_{MCGF} - \tilde{\beta}_{MCG}] =$$
$$\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}u\right) - \text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)$$

Sumamos y restamos la expresión:

$$\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)$$

de forma que:

$$\begin{aligned}
& \text{plim}[\tilde{\beta}_{MCGF} - \tilde{\beta}_{MCG}] = \\
= \quad & \text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}u\right) - \text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)\\
& +\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\left(\frac{1}{T}X'\Omega^{-1}u\right) - \text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1}\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right) = \\
= \quad & \text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1}\left[\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}u\right) - \text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)\right]\\
& +\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)\left[\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1} - \text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1}\right]
\end{aligned}$$

Para que:

$$\text{plim}\tilde{\beta}_{MCGF} = \text{plim}\tilde{\beta}_{MCG} = \beta$$

es suficiente (no necesario) que:

I) $\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}u\right) = \text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right)$

II) $\text{plim}\left(\frac{1}{T}X'\widehat{\Omega}^{-1}X\right)^{-1} = \text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1}$

Las condiciones suficientes de consistencia podemos encontrarlas como:

i. $\text{plim}\left(\frac{1}{T}X'\Omega^{-1}u\right) = 0$
ii. $\text{plim}\left(\frac{1}{T}X'\Omega^{-1}X\right)^{-1} = G^{-1}$
iii. $plim\widehat{\Omega} = \Omega$

b) Distribución asintótica. Para obtener la distribución asintótica de $\tilde{\beta}_{MCGF}$ debemos hacer uso de algunas propiedades asintóticas. Sabemos que:

$$X_T \xrightarrow{p} X \Rightarrow X_T \xrightarrow{d} X$$

por tanto si:

$$\sqrt{T}(\tilde{\beta}_{MCGF} - \beta) \xrightarrow{p} \sqrt{T}(\tilde{\beta}_{MCG} - \beta)$$

entonces:

$$\sqrt{T}(\tilde{\beta}_{MCGF} - \beta) \xrightarrow{d} \sqrt{T}(\tilde{\beta}_{MCG} - \beta)$$

y en consecuencia ambos estimadores tendrían idéntica distribución asintótica:

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

2.4.2. Estimador de σ^2

Si especificamos la matriz de varianzas y covarianzas de la perturbación $E(uu') = \sigma^2\Omega$, con σ^2 y Ω desconocidas, estimamos σ^2 de forma habitual pero teniendo en cuenta que las perturbaciones no son esféricas:

$$\tilde{\sigma}^2_{MCGF} = \frac{\tilde{u}'_{MCGF}\widehat{\Omega}^{-1}\tilde{u}_{MCGF}}{T-K}$$

Si $\widehat{\Omega}$ es un estimador consistente de Ω:

$$\text{plim}\left(\frac{1}{T}\tilde{u}'(\widehat{\Omega}^{-1} - \Omega^{-1})\tilde{u}\right) = 0$$

y $\tilde{\sigma}^2_{MCGF}$ es un estimador consistente de σ^2:

$$\tilde{\sigma}^2_{MCGF} = \frac{(Y - X\tilde{\beta}_{MCGF})'\widehat{\Omega}^{-1}(Y - X\tilde{\beta}_{MCGF})}{T-K} = \frac{Y'\widehat{\Omega}^{-1}Y - \tilde{\beta}'_{MCGF}X'\widehat{\Omega}^{-1}Y}{T-K}$$

2.5. Contrastes de restricciones lineales

Vamos a ver cómo realizar contrastes de restricciones lineales sobre el vector β en el MRLG con perturbaciones no esféricas pero normales.

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

$$\begin{aligned} H_0 &: R\beta = r \\ H_a &: R\beta \neq r \end{aligned}$$

donde R es una matriz $(q\times K)$ y r es un vector de dimensión q, siendo q el número de restricciones lineales a contrastar. Podemos optar por realizar los contrastes en el modelo transformado para lo cual podemos aplicar el estadístico de diferencias en las sumas de cuadrados.

Suponiendo $E(uu') = \sigma^2\Omega \quad \Omega$ conocida y σ^2 desconocida tendríamos:

$$F = \frac{(R\tilde{\beta}_{MCG} - r)'[R(X_*'X_*)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

o equivalentemente

$$\frac{(SCR_r - SCR)/q}{SCR/T-K} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)}$$

que implicaría estimar dos modelos el restringido y el no restringido bajo la hipótesis de contraste correspondiente. La regla de decisión es la habitual.

Si optamos por trabajar en el modelo original podemos distinguir los siguientes casos:

a) $E(uu') = \Sigma$, Σ conocida.

b) $E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas.

c) $E(uu') = \sigma^2\Omega$, Ω conocida pero σ^2 desconocida.

d) $E(uu') = \Sigma$, Σ desconocida.

e) $E(uu') = \sigma^2\Omega$, Ω y σ^2 desconocidas.

y aplicar el estadístico general o el correspondiente estadístico de diferencias en las sumas de cuadrados.

Caso 1: $E(uu') = \Sigma$, Σ conocida.

Sea $Y = X\beta + u$ con $E(uu') = \Sigma$, Σ conocida. En este caso estimamos por MCG

$$\tilde{\beta}_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

y si las perturbaciones tienen una distribución Normal tenemos,

$$\tilde{\beta}_{MCG} \sim N(\beta,\, (X'\Sigma^{-1}X)^{-1})$$

de donde

$$R\tilde{\beta}_{MCG} \sim N(R\beta,\, R(X'\Sigma^{-1}X)^{-1}R') \tag{2.8}$$

con lo que si la hipótesis nula es cierta

$$R\tilde{\beta}_{MCG} \overset{H_0}{\sim} N(r,\, R(X'\Sigma^{-1}X)^{-1}R')$$

y con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

donde q es el número de restricciones. Rechazamos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\sqrt{R(X'\Sigma^{-1}X)^{-1}R'}} \overset{H_0}{\sim} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{des(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} N(0,1)$$

Caso 2: $E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas.

Sea $Y = X\beta + u \quad$ con $\quad E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas. En este caso estimamos por MCG

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

y si las perturbaciones tienen una distribución Normal tenemos,

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

con lo que el estadístico de contraste y su distribución bajo H_0 son:

$$(R\tilde{\beta}_{MCG} - r)'[\sigma^2R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r) \overset{H_0}{\sim} \chi^2_q$$

Rechazaremos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\sigma\sqrt{R(X'\Omega^{-1}X)^{-1}R'}} \overset{H_0}{\sim} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{des(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} N(0,1)$$

Caso 3: $E(uu') = \sigma^2\Omega$, Ω conocida y σ^2 desconocida.

Sea $Y = X\beta + u \quad$ con $\quad E(uu') = \sigma^2\Omega$, Ω conocida y σ^2 desconocida. En este caso estimamos por MCG

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

y si las perturbaciones tienen una distribución Normal tenemos,

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

con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

Rechazaremos la hipótesis nula si $F > \mathcal{F}_{(q,T-K)\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\tilde{\sigma}_{MCG}\sqrt{R(X'\Omega^{-1}X)^{-1}R'}} \overset{H_0}{\sim} \mathrm{t}_{(T-K)}$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{\widehat{des}(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} \mathrm{t}_{(T-K)}$$

Rechazaremos la hipótesis nula si $t > \mathrm{t}_{(T-K)\,\frac{\alpha}{2}}$ para un nivel de significación α.

Caso 4: $\boldsymbol{E(uu') = \Sigma,\ \Sigma}$ desconocida.

Sea $Y = X\beta + u \quad \text{con} \quad E(uu') = \Sigma$, Σ desconocida. En este caso estimamos por MCGF

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

El estimador de MCGF es no lineal, en general es sesgado y consistente si $\widehat{\Sigma}$ es consistente, además es eficiente asintóticamente y con distribución asintótica conocida. Por tanto podemos hacer inferencia asintótica con este estimador:

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

$$\sqrt{T}(R\tilde{\beta}_{MCGF} - R\beta) \xrightarrow{d} N(0,\ RG^{-1}R')$$

con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

donde q es el número de restricciones. Rechazamos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCGF} - r}{\sqrt{R(X'\widehat{\Sigma}^{-1}X)^{-1}R'}} \xrightarrow{d,H_0} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCGF}}{des(\tilde{\beta}_{i,MCGF})} \xrightarrow{d,H_0} N(0,1)$$

Caso 5: $\boldsymbol{E(uu') = \sigma^2\Omega,\ \Omega}$ y $\boldsymbol{\sigma^2}$ desconocidas.

Sea $Y = X\beta + u \quad \text{con} \quad E(uu') = \sigma^2\Omega$, Ω y σ^2 desconocidas. En este caso estimamos por MCGF

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

El estimador de MCGF es no lineal, en general es sesgado y consistente si $\widehat{\Sigma}$ es consistente, además es eficiente asintóticamente y con distribución asintótica conocida. Por tanto podemos hacer inferencia asintótica con este estimador:

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

$$\sqrt{T}(R\tilde{\beta}_{MCGF} - R\beta) \xrightarrow{d} N(0,\ RG^{-1}R')$$

con lo que el estadístico de contraste y su distribución bajo H_0 son:

$$F = (R\tilde{\beta}_{MCGF} - r)'[\tilde{\sigma}^2_{MCGF}R(X'\widehat{\Omega}^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCGF} - r) \xrightarrow{d,H_0} \chi^2_q$$

donde q es el número de restricciones. Rechazamos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCGF} - r}{\tilde{\sigma}_{MCGF}\sqrt{R(X'\widehat{\Omega}^{-1}X)^{-1}R'}} \xrightarrow{d,H_0} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCGF}}{\widehat{des}(\tilde{\beta}_{i,MCGF})} \xrightarrow{d,H_0} N(0,1)$$

2.6. Ejemplo: Sistemas de Ecuaciones

En ocasiones necesitamos estimar un sistema de ecuaciones. Nosotros 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.

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:

$$Y_{1i} = a_1 + b_1X_{1i} + u_{1i} \quad \Longleftrightarrow \quad Y_1 = X_1\beta_1 + u_1 \quad \text{siendo} \qquad u_1 \sim N(0, \sigma_1^2 I_{N_1})$$

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

$$Y_{2i} = a_2 + b_2X_{2i} + u_{2i} \quad \Longleftrightarrow \quad Y_2 = X_2\beta_2 + u_2 \quad \text{siendo} \qquad u_2 \sim N(0, \sigma_2^2 I_{N_2})$$

Podemos escribir el siguiente sistema de ecuaciones:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \Longleftrightarrow \quad Y = X\beta + u \tag{2.9}$$

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

Sobre X suponemos que es no estocástica y

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

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

Notar:

- 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
  a) $E(u_1u_1') = \sigma_1^2 I_{N_1}, E(u_2u_2') = \sigma_2^2 I_{N_2}$ 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.
  b) $E(u_1u_1') = \sigma_1^2 I_{N_1}, E(u_2u_2') = \sigma_2^2 I_{N_2}$ con $\sigma_1^2 \neq \sigma_2^2$ es decir heterocedasticidad entre ecuaciones y $E(u_1u_2') = E(u_2u_1') = 0$.
  c) Ecuaciones aparentemente no relacionadas $E(u_1u_2') = E(u_2u_1') \neq 0$.

Notar que dado que X es no estocástica 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 y 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, 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 por separado que es más sencillo.

Además

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{T-K} \quad \text{donde} \quad N = N_1 + N_2 \quad \text{y} \quad K = K_1 + K_2$$

es insesgado y consistente.

• **Contraste de cambio estructural o de Chow:** Se llama contraste de cambio estructural al contraste de que todos o algunos de los parámetros 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:

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

Hay dos formas alternativas de realizar el contraste:

- Alternativa 1: Con el estadístico:

$$\frac{(R\hat{\beta}_{MCO} - r)' \left[R(X'X)^{-1}R' \right]^{-1} (R\hat{\beta}_{MCO} - r)/q}{\hat{\sigma}^2} \overset{H_0}{\sim} \mathcal{F}(q, T-K)$$

  donde

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

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K} = \frac{\hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2}{T-K}$$

  Regla de decisión:

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

- Alternativa 2: Con el estadístico

$$\frac{(\hat{u}_r'\hat{u}_r - \hat{u}'\hat{u})/q}{\hat{u}'\hat{u}/T-K} \overset{H_0}{\sim} \mathcal{F}(q, T-K)$$

  donde $\hat{u}'\hat{u}$ es la SCR del modelo no restringido (2.9) tal que $\hat{u}'\hat{u} = \hat{u}_1'\hat{u}_1 + \hat{u}_2'\hat{u}_2$;

  $\hat{u}_r'\hat{u}_r$ es la SCR del modelo restringido siguiente

$$\left[\begin{array}{c} Y_1 \\ Y_2 \end{array} \right] = \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] \beta + \left[\begin{array}{c} u_1 \\ u_2 \end{array} \right] \qquad (2.10)$$

  Dado que hemos supuesto que $E(uu') = \sigma^2 I_{N_1+N_2}$ el modelo restringido se estima por MCO y no podemos hacerlo equivalentemente a 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

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

Suponiendo σ_1^2, σ_2^2 conocidas, el modelo debe ser estimado por MCG que coincide con MCO ecuación por ecuación ya que X es diagonal por bloques, $E(uu')$ también lo es y hay homocedasticidad dentro de cada ecuación.

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

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

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 por separado que es más sencillo. Además el resultado es independiente de que conozcamos o no σ_1^2, σ_2^2.

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

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

Hay dos formas alternativas de realizar el contraste:

- Alternativa 1: Con 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

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

  con la regla de decisión habitual.

- Alternativa 2: Si queremos utilizar el estadístico de diferencias en las sumas residuales de cuadrados debemos estimar el modelo restringido (2.10) por MCG, dado la estructura de $E(uu')$, y no podemos hacerlo equivalentemente a MCO ecuación por ecuación ya que su matriz de regresores no es diagonal por bloques.

Si σ_1^2 y σ_2^2 son desconocidas tendremos que estimarlas. El estimador de los parámetros 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_1^2 y σ_2^2; un estimador consistente de $\sigma_i^2 \quad i = 1, 2$ sería:

$$\hat{\sigma}_1^2 = \frac{\hat{u}_1'\hat{u}_1}{N_1} \qquad \hat{u}_1'\hat{u}_1 = Y_1'Y_1 - \hat{\beta}_1^{MCO'}X_1'Y_1$$

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

Notar que en cada ecuación por separado hay homocedasticidad y no autocorrelación. En este caso el contraste de cambio estructural podríamos hacerlo alternativamente por el estadístico de diferencias en las sumas de cuadrados o con el estadístico:

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

con la regla de decisión habitual.

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

$$\begin{cases} Y_1 &= X_1\beta_1 + u_1 \quad N\text{obs} \\ Y_2 &= X_2\beta_2 + u_2 \quad N\text{obs} \end{cases}$$

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

- $E(u_1u_1') = \sigma_1^2 I_N \quad E(u_2u_2') = \sigma_2^2 I_N$, 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.

$$E(uu') = \begin{bmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{12}I & \sigma_{22}I \end{bmatrix}$$

En este caso necesariamente debemos estimar el modelo conjunto (2.9) por MCG si $\sigma_1^2, \sigma_2^2, \sigma_{12}$ son conocidas. Si son desconocidas el modelo conjunto debe estimarse por MCGF. Podemos encontrar estimadores consistentes de estos parámetros utilizando los residuos MCO de estimar cada ecuación por separado:

$$\hat{\sigma}_{11} = \frac{\hat{u}_1'\hat{u}_1}{N} \qquad \hat{u}_1'\hat{u}_1 = Y_1'Y_1 - \hat{\beta}_1^{MCO'}X_1'Y_1$$

$$\hat{\sigma}_{22} = \frac{\hat{u}_2'\hat{u}_2}{N} \qquad \hat{u}_2'\hat{u}_2 = Y_2'Y_2 - \hat{\beta}_2^{MCO'}X_2'Y_2$$

$$\hat{\sigma}_{12} = \frac{\hat{u}_1'\hat{u}_2}{N} \qquad \hat{u}_1 = Y_1 - X_1\hat{\beta}_1 \qquad \hat{u}_2 = Y_2 - X_2\hat{\beta}_2$$

2.7. Contrastes de restricciones lineales

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:

$$\begin{aligned} H_0 &: R\beta = r \\ H_a &: R\beta \neq r \end{aligned}$$

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

Vamos a distinguir los siguientes casos:

a) $E(uu') = \Sigma$, Σ conocida.

b) $E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas.

c) $E(uu') = \sigma^2\Omega$, Ω conocida pero σ^2 desconocida.

d) $E(uu') = \Sigma$, Σ desconocida.

Caso 1: $E(uu') = \Sigma$, Σ conocida.

Sea $Y = X\beta + u \quad \text{con} \quad E(uu') = \Sigma$, Σ conocida. En este caso estimamos por MCG

$$\tilde{\beta}_{MCG} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

y si las perturbaciones tienen una distribución Normal tenemos,

$$\tilde{\beta}_{MCG} \sim N(\beta,\, (X'\Sigma^{-1}X)^{-1})$$

de donde

$$R\tilde{\beta}_{MCG} \sim N(R\beta,\, R(X'\Sigma^{-1}X)^{-1}R') \tag{2.11}$$

con lo que si la hipótesis nula es cierta

$$R\tilde{\beta}_{MCG} \overset{H_0}{\sim} N(r,\, R(X'\Sigma^{-1}X)^{-1}R')$$

y con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

donde q es el número de restricciones. Rechazamos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\sqrt{R(X'\Sigma^{-1}X)^{-1}R'}} \overset{H_0}{\sim} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{des(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} N(0,1)$$

Caso 2: $E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas.

Sea $Y = X\beta + u \quad \text{con} \quad E(uu') = \sigma^2\Omega$, Ω y σ^2 conocidas. En este caso estimamos por MCG

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

y si las perturbaciones tienen una distribución Normal tenemos,

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

con lo que el estadístico de contraste y su distribución bajo H_0 son:

$$(R\tilde{\beta}_{MCG} - r)'[\sigma^2 R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r) \overset{H_0}{\sim} \chi^2_q$$

Rechazaremos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\sigma\sqrt{R(X'\Omega^{-1}X)^{-1}R'}} \overset{H_0}{\sim} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{des(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} N(0,1)$$

Caso 3: $E(uu') = \sigma^2\Omega$, Ω conocida y σ^2 desconocida.

Sea $Y = X\beta + u$ con $E(uu') = \sigma^2\Omega$, Ω conocida y σ^2 desconocida. En este caso estimamos por MCG

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

y si las perturbaciones tienen una distribución Normal tenemos,

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

con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

Rechazaremos la hipótesis nula si $F > \mathcal{F}_{(q,\,T-K)\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCG} - r}{\tilde{\sigma}_{MCG}\sqrt{R(X'\Omega^{-1}X)^{-1}R'}} \overset{H_0}{\sim} \mathrm{t}_{(T-K)}$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCG}}{\widehat{des}(\tilde{\beta}_{i,MCG})} \overset{H_o}{\sim} \mathrm{t}_{(T-K)}$$

Rechazaremos la hipótesis nula si $t > \mathrm{t}_{(T-K)\,\frac{\alpha}{2}}$ para un nivel de significación α.

Caso 4: $E(uu') = \Sigma$, Σ desconocida.

Sea $Y = X\beta + u$ con $E(uu') = \Sigma$, Σ desconocida. En este caso estimamos por MCGF

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

El estimador de MCGF es no lineal, en general es sesgado y consistente si $\widehat{\Sigma}$ es consistente, además es eficiente asintóticamente y con distribución asintótica conocida. Por tanto podemos hacer inferencia asintótica con este estimador:

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

$$\sqrt{T}(R\tilde{\beta}_{MCGF} - R\beta) \xrightarrow{d} N(0,\ RG^{-1}R')$$

con lo que el estadístico de contraste y su distribución bajo H_0 son:

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

donde q es el número de restricciones. Rechazamos la hipótesis nula si $F > \chi^2_{(q)\,\alpha}$ para un nivel de significación α.

Si $q = 1$ el estadístico anterior se puede escribir como:

$$t = \frac{R\tilde{\beta}_{MCGF} - r}{\sqrt{R(X'\widehat{\Sigma}^{-1}X)^{-1}R'}} \xrightarrow{d,H_0} N(0,1)$$

por ejemplo si $H_0 : \beta_i = 0$, es decir, contrastamos la significatividad de una de las variables exógenas, podemos escribir el estadístico anterior de la manera habitual:

$$t = \frac{\tilde{\beta}_{i,MCGF}}{des(\tilde{\beta}_{i,MCGF})} \xrightarrow{d,H_0} N(0,1)$$

2.7.1. Estadístico de diferencias en las sumas de cuadrados

Cuando un conjunto de hipótesis lineales $H_0 : R\beta = r$ se acepta tras un contraste para un nivel de significatividad dado deberíamos estimar sujeto a estas restricciones. El problema objetivo ahora sería:

$$Min\ \hat{u}'_{MCG}\Omega^{-1}\hat{u}_{MCG}$$
$$s.a.\ R\beta = r$$

de donde

$$\tilde{\beta}_{MCG_r} = \tilde{\beta}_{MCG} - (X'\Omega^{-1}X)^{-1}R[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r)$$

donde $\tilde{\beta}_{MCG_r}$ es el estimador sujeto a restricciones. Por tanto el estadístico F de contraste puede escribirse como:

$$\frac{(SCR_r - SCR)/q}{SCR/T-K} \overset{H_0}{\sim} F_{(q,T-K)}$$

o lo que es igual

$$\frac{(\hat{u}'_{MCG_r}\Omega^{-1}\hat{u}_{MCG_r} - \hat{u}'_{MCG}\Omega^{-1}\hat{u}_{MCG})/q}{(\hat{u}'_{MCG}\Omega^{-1}\hat{u}_{MCG})/T-K} \overset{H_0}{\sim} F_{(q,T-K)}$$

Tema 3

Heterocedasticidad

3.1. Definición y causas

Hasta el momento uno de los supuestos básicos del modelo de regresión lineal es que la varianza de cada término de perturbación u_t condicionada a los valores de las variables explicativas, es constante e igual a σ^2. Llamábamos a este supuesto homocedasticidad y lo denotábamos: $E(u_t^2) = \sigma^2 \quad \forall t$. En este tema vamos a relajar este supuesto y consideraremos el modelo de regresión lineal bajo heterocedasticidad.

Llamamos heterocedasticidad al caso en que la matriz de varianzas y covarianzas de la perturbación de un modelo econométrico sigue siendo diagonal, pero los elementos de ésta diagonal ya no son todos iguales. Es decir, la varianza del término de error varía a través del tiempo si miramos a series temporales, o cambia de un individuo a otro si miramos datos de sección cruzada, (familias, países, etc.).
El caso más sencillo de heterocedasticidad es aquel en que la matriz de varianzas y covarianzas tiene la forma siguiente:

$$E(uu') = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_T^2 \end{pmatrix}$$

donde la varianza de la perturbación, $V(u_t) = \sigma_t^2$, no es constante porque varía a lo largo del tiempo. Seguimos suponiendo que no existe autocorrelación entre perturbaciones de distinto momento del tiempo, es decir, $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$ por lo que sólo consideramos la existencia de heterocedasticidad.

Para entender la diferencia entre el concepto de homocedasticidad y el concepto de heterocedasticidad podemos considerar el modelo de regresión simple en los dos gráficos siguientes.

En la Figura 3.1 se puede observar que la varianza condicional de Y_t a las X_t permanece igual sin importar los valores que tome la variable X. Recordar que la varianza condicional de Y_t es la misma que la de u_t, por tanto, en el gráfico estamos observando cómo la varianza de la perturbación permanece constante independientemente del valor que tome el regresor. En la Figura 3.2 se puede observar que la varianza de Y_t aumenta a medida que X_t aumenta y por

Gráfico 3.1: Perturbaciones homocedásticas

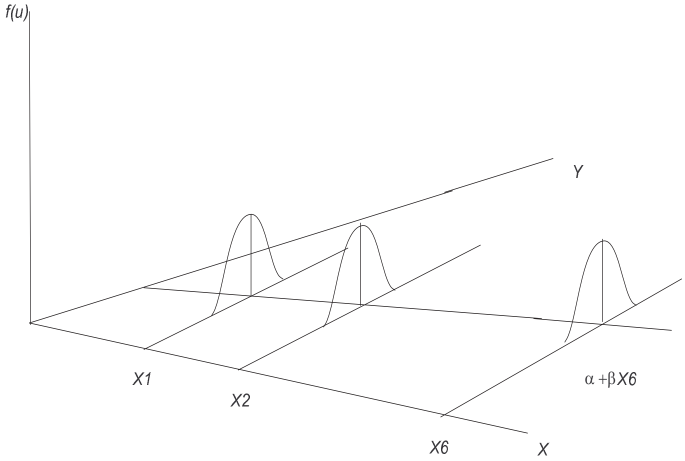
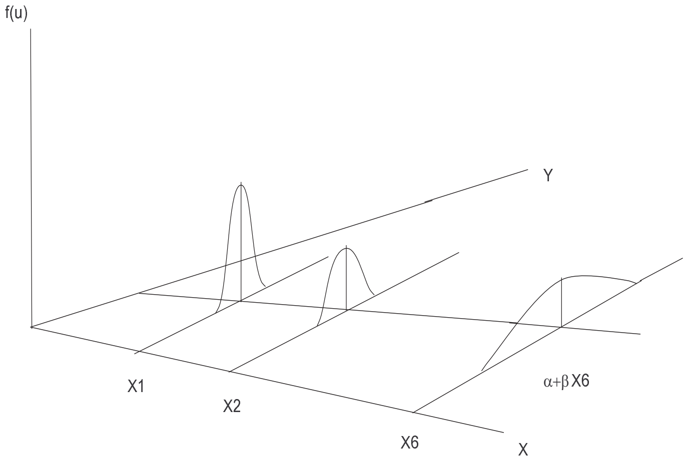


Gráfico 3.2: Perturbaciones heterocedásticas



tanto hay heterocedasticidad:

$$E(u_t^2) = \sigma_t^2$$

Hay diversas razones por las cuales las varianzas de u_t pueden no ser constantes:

- Modelos que tengan en cuenta expectativas: una expectativa no es más que una medida de lo que un agente espera que ocurra, la formación de esa medida conlleva un proceso de aprendizaje. Es de esperar que los agentes aprendan de sus errores y según avance el tiempo se confundan menos, en este caso σ_i^2 se reducirá.

- Si estamos analizando la relación entre consumo y renta podemos esperar que a medida que aumente la renta aumente σ_i^2. Una familia con mayor renta tiene mayores posibilidades de consumo, no sólo consumir más variedad de productos, sino que aumentará el valor del consumo real. Si la renta es suficientemente grande, podrá diferir consumo entre periodos y podrá ahorrar.

- Por razonamientos parecidos a los anteriores las empresas con mayores beneficios podrán presentar mayor variabilidad en sus políticas de dividendos. Si las ganancias son muy bajas simplemente no podrán repartir dividendos.

- Otra causa de heterocedasticidad puede encontrarse en la mala especificación de un modelo. Si en un modelo se ha omitido una variable relevante su exclusión puede llevar a pensar que existe heterocedasticidad en las perturbaciones del modelo. Por ejemplo, si consideramos la función de demanda de un producto y excluimos los precios de los bienes complementarios a él o de sus competidores, los estimadores MCO serán sesgados y el estudio de los residuos minimocuadráticos del modelo puede dar la impresión de que la varianza de la perturbación no es constante. Si incluimos la variable o variables omitidas la impresión puede desaparecer. En este caso la solución al problema pasa por especificar correctamente el modelo.

El problema de heterocedasticidad es más frecuente en datos de sección cruzada. En datos de sección cruzada disponemos de datos sobre diferentes unidades económicas en el mismo momento del tiempo. Las unidades generalmente son consumidores individuales, familias, empresas, industrias, países, estados, provincias, etc., con diverso tamaño dentro de la misma población. En estos casos es más adecuado aplicar el subíndice i que es el habitual en datos de sección cruzada. En este caso denotamos la existencia de heterocedasticidad como:

$$E(u_i^2) = \sigma_i^2 \qquad i = 1, 2, \ldots, N$$

y la matriz de varianzas y covarianzas de la perturbación sería:

$$E(uu') = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & 0 & \ldots & 0 \\ 0 & 0 & \sigma_3^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \sigma_N^2 \end{pmatrix}$$

En presencia de heterocedasticidad el número de parámetros a estimar en el modelo crece con el número de observaciones ya que con cada observación aparece un nuevo parámetro σ_i^2 (σ_t^2). Además hay que recordar que se deben estimar los K-betas desconocidos del modelo. Esta colección de parámetros desconocidos no es estimable con una muestra de tamaño N (T) si lo que queremos son estimaciones fiables. Es preciso, por tanto, establecer algún supuesto acerca del modo en que σ_i^2 varía a través de las distintas familias, individuos o países que integran la muestra (σ_i^2 a lo largo del tiempo), de forma que consigamos reducir el número de parámetros desconocidos.
Esta es una restricción importante porque tanto la detección de la heterocedasticidad como la estimación del modelo en presencia de ésta se ven condicionados por el supuesto específico que se haya establecido acerca del modo en que la varianza de la perturbación varía entre observaciones muestrales.

Cuando un modelo presenta una situación de heterocedasticidad, hay varias cuestiones de importancia:

a) ¿Cómo puede detectarse la presencia de heterocedasticidad?

b) ¿Cuáles son las consecuencias de la heterocedasticidad sobre el estimador de MCO y su matriz de varianzas y covarianzas? Sabemos que en presencia de heterocedasticidad el estimador MCO es lineal, insesgado y consistente pero ineficiente. Su matriz de varianzas y covarianzas se define:

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

c) ¿Cómo debe estimarse un modelo que presenta heterocedasticidad? Si nuestro objetivo es obtener estimadores lineales, insesgados, eficientes y consistentes estimaremos por MCG: $\tilde{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$

   Este estimador puede obtenerse por dos vías alternativas:

   I) Aplicando directamente el estimador MCG, definido en la expresión anterior, al modelo.

   II) Transformando el modelo hasta obtener perturbaciones esféricas y aplicando MCO al modelo transformado.

d) ¿Cuál es la forma correcta de hacer contraste de hipótesis lineales en un modelo con heterocedasticidad?

e) ¿Cómo se elaboran las predicciones del modelo econométrico en tal situación?

A todas estas preguntas intentaremos contestar en los puntos siguientes.

3.2. Contrastes de heterocedasticidad

3.2.1. Detección

El hecho de que las perturbaciones de un modelo sean heterocedásticas no es una razón para rechazarlo. Lo importante es tenerlo en cuenta. Sabemos que en presencia de heterocedasticidad el estimador MCO es ineficiente, mientras que si conocemos Ω o lo que es lo mismo, la forma funcional de la heterocedasticidad, el estimador de MCG es ELIO. El mensaje parece claro, en presencia de heterocedasticidad con Ω conocida debemos estimar el modelo por MCG. Sin embargo, sólo estaremos dispuestos a estimar por MCG cuando verdaderamente exista heterocedasticidad. Así que el primer paso para nosotros será detectar la posible existencia de heterocedasticidad.

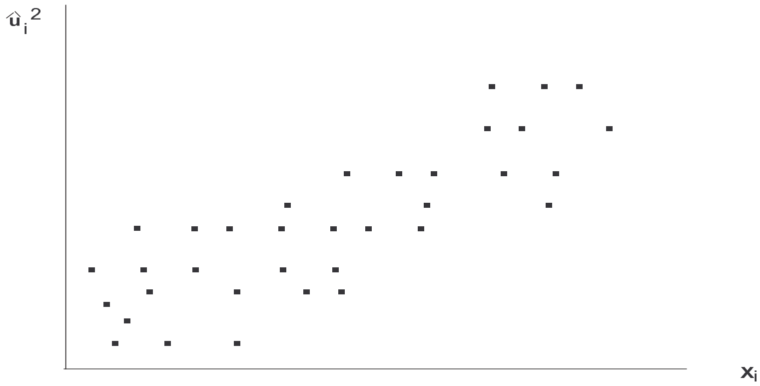
La determinación de la existencia de heterocedasticidad sólo podremos conseguirla aplicando un test de heterocedasticidad, pero en ocasiones para aplicar este test necesitamos conocer la forma funcional de la misma. Salvo en casos puntuales en los que la heterocedasticidad venga provocada por una transformación de los datos realizada por el investigador, en el resto de los casos el investigador generalmente, no conocerá su existencia a priori. Si maneja datos de sección cruzada estará sobre aviso ya que, como ya hemos dicho anteriormente, la existencia de heterocedasticidad en datos de sección cruzada es más una norma que una excepción. Llegados a este punto podemos pensar ¿qué instrumento me va a proporcionar información sobre mi problema? muy fácil si pensamos en cuál es el problema. Nuestro problema es que la varianza de la perturbación no es constante, la varianza poblacional es desconocida, a su vez la perturbación no es observable, ¿qué conocemos próximo a ella que nos sea de utilidad? el residuo, ¿cuál? el de mínimos cuadrados ordinarios ya que no conocemos otro. ¿Podemos utilizar el residuo como aproximación a la perturbación? sí, es un estimador consistente aunque ineficiente de la perturbación. Por la misma razón usaremos el residuo al cuadrado como aproximación al comportamiento de la varianza de la perturbación.

Por ejemplo si en el modelo:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i \quad i = 1, 2, \ldots, N \tag{3.1}$$

donde $E(u_i) = 0 \quad \forall i$ y $E(u_i u_j) = 0 \quad \forall i, j \quad i \neq j$
sospechamos que u_i es heterocedástica debido a la variable X_i, por ejemplo, su varianza es creciente con X_i. La forma correcta de proceder para detectar la existencia de heterocedasticidad en las perturbaciones del modelo sería estimar éste por MCO y estudiar el gráfico de los residuos MCO, ($\hat{u}_{MCO,i}$), y X_i. Si el gráfico es como el de la Figura 3.3

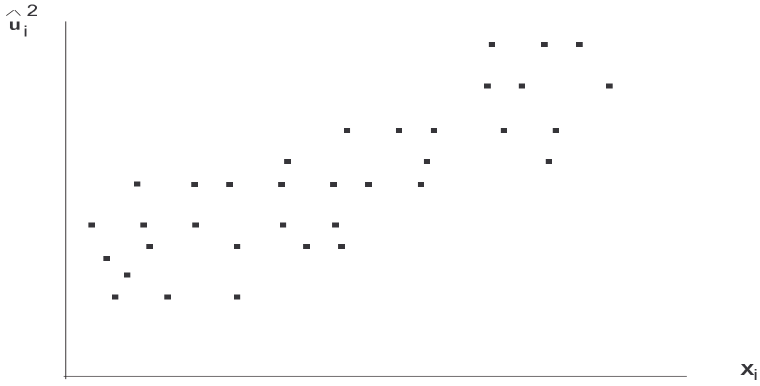
Gráfico 3.3: Perturbaciones heterocedásticas



pensaremos que los residuos $\hat{u}_{MCO,i}$ se incrementan con X_i y que el incremento es proporcional. Dado que el residuo es una estimación de la perturbación propondremos, por ejemplo:

$$V(u_i) = E(u_i^2) = \sigma^2 X_i$$

Gráfico 3.4: Perturbaciones heterocedásticas

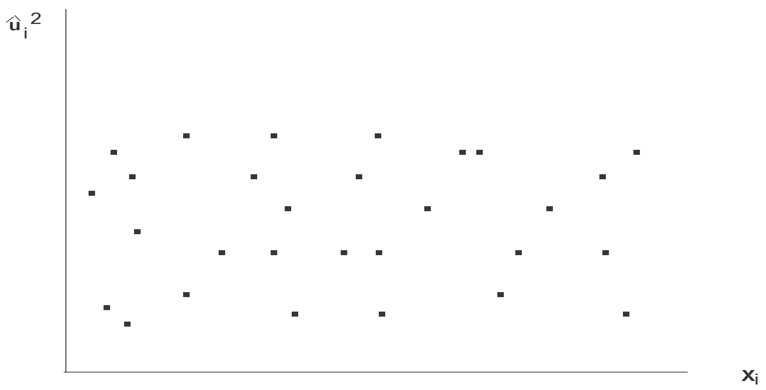


Si el gráfico de los residuos $\hat{u}_{MCO,i}$ y X_i fuera como en la Figura 3.4 supondríamos que el aumento en la varianza de u_i es lineal a X_i y propondríamos:

$$E(u_i^2) = a + bX_i$$

En el caso de que no conozcamos cuál de las variables exógenas genera heterocedasticidad tendremos que estudiar los gráficos de los residuos de MCO, contraponiéndolos a cada una de las variables exógenas restantes. Así, si la gráfica entre $\hat{u}_{MCO,i}$ y X_i resultara como la de la Figura 3.5, en la que no se aprecia ningún patrón de comportamiento y parece que hay una distribución aleatoria de los pares $(X_i, \hat{u}_i^2)$, procederíamos a analizar los residuos frente a Z_i.

Gráfico 3.5: Perturbaciones homocedásticas



En la Figura 3.6 podemos observar otros patrones de comportamiento en los residuos que pueden indicar la existencia de heterocedasticidad en las perturbaciones.

Sin embargo el estudio gráfico de los residuos no es determinativo. Para determinar si existe o no heterocedasticidad tendremos que realizar un contraste de existencia de heterocedasticidad con un estadístico adecuado. El análisis gráfico no es una pérdida de tiempo ya que la relación entre X_{ki} y $\hat{u}_{MCO,i}$ nos indicará una posible forma funcional (de heterocedasticidad) para la varianza de la perturbación y puede indicarnos cuál es el test de contraste más adecuado.

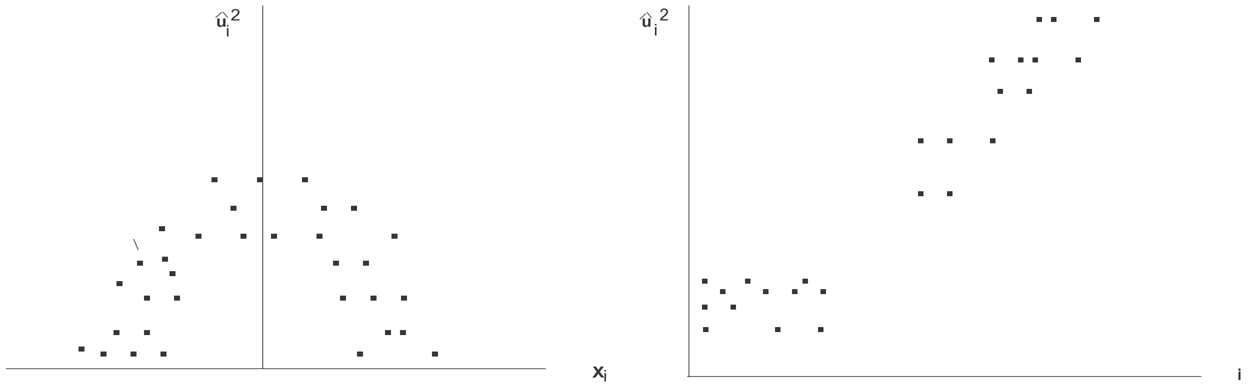
3.2.2. Contrastes de heterocedasticidad

A continuación veremos algunos de los test de contraste para heterocedasticidad más importantes. Todos ellos contrastan la existencia de heterocedasticidad suponiendo:
H_0: ausencia de heterocedasticidad.
H_a: existencia de heterocedasticidad.
Algunos de ellos necesitan conocer la forma funcional de heterocedasticidad y otros no. Algunos de ellos sugieren la forma funcional de la heterocedasticidad cuando se rechaza la H_0, por lo que la transformación de variables necesaria para estimar por MCG es inmediata, otros en cambio no.

Test de Goldfeld y Quandt

El test de Goldfeld y Quandt es un contraste paramétrico que depende de la forma de heterocedasticidad supuesta. El contraste fue propuesto por Goldfeld y Quandt en 1965 y parte del

Gráfico 3.6: Perturbaciones heterocedásticas



supuesto de que la magnitud de σ_i^2 depende monótonamente de los valores de una variable Z_i. Por ejemplo, en el análisis del gasto familiar podemos suponer que la varianza del gasto depende del nivel de renta de cada familia y proponer $\sigma_i^2 = \sigma^2 g(R_i)$, donde $g(\cdot)$ es una función creciente con la renta familiar y σ^2 un factor de escala. La variable Z_i generalmente suele ser una de las variables explicativas del modelo, aunque no es preciso que lo sea para llevar a cabo el contraste. En todo caso, necesitamos disponer de información muestral acerca de dicha variable. Para contrastar la hipótesis nula de ausencia de heterocedasticidad:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2$$

contra la alternativa de existencia de heterocedasticidad:

$$H_a : \sigma_i^2 = \sigma^2 g(Z_i, \gamma)$$

donde $g(\cdot)$ es una función monótona creciente con Z_i y γ un parámetro desconocido. Se procede de la siguiente manera:

a) Ordenar las observaciones de la muestra correspondiéndose con los valores de Z_i de menor a mayor.

b) Dividir la muestra en dos bloques de tamaño muestral N_1 y N_2 respectivamente, dejando fuera p observaciones centrales para hacer más independientes los dos grupos. El número de observaciones de cada grupo ha de ser similar y mayor que el número de parámetros a estimar: $\frac{N-p}{2} = N_1 = N_2$

c) Estimar, por MCO, el modelo de regresión separadamente para cada grupo de observaciones y calcular la Suma de Cuadrados Residual correspondiente.

d) Construir el siguiente estadístico de contraste, que bajo la hipótesis nula de ausencia de heterocedasticidad y suponiendo que la perturbación sigue una distribución normal y no está serialmente correlacionada, sigue una distribución F-Snedecor.

$$GQ = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} = \frac{\hat{u}_2'\hat{u}_2}{\hat{u}_1'\hat{u}_1}\frac{N_1 - K}{N_2 - K} \overset{H_0}{\sim} F_{(N_2-K, N_1-K)}$$

donde:
$\hat{u}_2'\hat{u}_2$ es la SCR de la regresión de Y sobre X en el segundo grupo de observaciones.

$\hat{u}_1'\hat{u}_1$ es la SCR de la regresión de Y sobre X en el primer grupo de observaciones.

$$Y_1 = X_1\beta + u_1 \longrightarrow \hat{u}_1'\hat{u}_1 \longrightarrow \frac{\hat{u}_1'\hat{u}_1}{\sigma_1^2} \sim \chi^2_{N_1-K}$$

$$Y_2 = X_2\beta + u_2 \longrightarrow \hat{u}_2'\hat{u}_2 \longrightarrow \frac{\hat{u}_2'\hat{u}_2}{\sigma_2^2} \sim \chi^2_{N_2-K}$$

La idea del contraste es la siguiente: si existe homocedasticidad las varianzas han de ser iguales, $\hat{u}_1'\hat{u}_1 \simeq \hat{u}_2'\hat{u}_2$ y $GQ \simeq 1$. Pero si existe heterocedasticidad del tipo propuesto, con la ordenación de la muestra de menor a mayor, la varianza del término de error será mayor al final de la muestra. Como el cuadrado de los residuos está asociado a la varianza de u_i, entonces $\hat{u}_2'\hat{u}_2$ debería ser sensiblemente mayor que $\hat{u}_1'\hat{u}_1$. Cuanto más diverjan las sumas de cuadrados, mayor será el valor del estadístico y mayor será la evidencia contra la H_0. Rechazaremos H_0, a un nivel de significación α si:

$$GQ > F_{(N_1-K,N_2-K)\alpha}$$

Observaciones:

- Si se sospecha que la varianza del término de error depende inversamente de los valores que toma una variable Z_i, entonces se ordena la muestra de acuerdo a los valores decrecientes de dicha variable y se procede del modo descrito anteriormente.
- ¿Cómo elegir p?
  Anteriormente se ha propuesto dividir la muestra en dos partes. Elegir el valor de p es relevante ya que cuanto mayor sea p más grados de libertad se pierden y por tanto, perdemos potencia del contraste. Si p es demasiado pequeño no habrá independencia entre grupos y se prima la homocedasticidad frente a la posibilidad de heterocedasticidad. Harvey y Phillips (1974) sugieren fijar p a un tercio de la muestra.
- En principio, el contraste se puede utilizar para detectar heterocedasticidad de forma general, aunque está pensado para alternativas específicas donde se supone un crecimiento de las varianzas en función de una determinada variable. Si en realidad existe otra forma de heterocedasticidad, el estadístico puede no detectarla.
- Por otro lado si no se rechaza la H_0 también puede deberse a una mala especificación de σ_i^2, que puede depender de una variable diferente a la considerada. Por ello puede ser necesario repetir el contraste para otras variables de las que podamos sospechar a priori.

Contraste de White

El contraste de heterocedasticidad propuesto por White en 1980 es un contraste paramétrico, de carácter general, que no precisa especificar la forma que puede adoptar la heterocedasticidad. En este sentido puede calificarse de robusto. Se procede de la forma siguiente:

a) Estimamos por MCO el modelo original y calculamos los residuos de MCO, $\hat{u}_{MCO,i}$.

b) Estimamos la regresión auxiliar: el cuadrado de los residuos mínimo-cuadráticos de la regresión anterior, sobre una constante, los regresores del modelo original, sus cuadrados y productos cruzados de segundo orden, evitando los redundantes:

$$\hat{u}^2_{MCO,i} = \delta_0 + \sum_{j=1}^{K}\gamma_j X_j + \sum_{j=1}^{K}\sum_{\ell=j}^{K}\delta_{j\ell}X_{ji}X_{\ell i} + v_t \qquad i = 1, 2, \ldots, N \tag{3.2}$$

Contrastar la hipótesis nula de homocedasticidad es equivalente a contrastar que todos los coeficientes de esta regresión, exceptuando el término independiente son cero. Es decir:

$$H_0 : \delta_{j\ell} = \gamma_j = 0 \quad \forall j, \ell$$

c) El estadístico de contraste es $\lambda = NR^2$ donde R^2 es el coeficiente de determinación de la regresión auxiliar (3.2). Se puede demostrar que bajo la H_0:

$$\lambda = NR^2 \xrightarrow{H_0,d} \chi^2_{(p)}$$

donde p es el número de regresores, sin incluir el intercepto, en la regresión auxiliar (3.2).

Rechazamos la H_0 si el valor muestral del estadístico excede del valor crítico de las tablas elegido para un nivel de significatividad α dado.

Observaciones:

a) Este contraste es muy flexible ya que no especifica la forma funcional de heterocedasticidad, pero por otro lado, si se rechaza la hipótesis nula de homocedasticidad no indica cuál puede ser la dirección a seguir.

b) El contraste de White puede recoger otro tipo de problemas de mala especificación de la parte sistemática, omisión de variables relevantes, mala forma funcional etc. Esto es positivo si se identifica cuál es el problema, en caso contrario, la solución que se tome puede estar equivocada.

Contraste de Breusch y Pagan

Breusch y Pagan en 1979 derivan un contraste de heterocedasticidad donde la hipótesis alternativa es bastante general:

$$\begin{array}{l} H_0 : \sigma_i^2 = \sigma^2 \quad \forall i \\ H_a : \sigma_i^2 = \sigma^2 g(\alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \ldots + \alpha_p Z_{pi}) = \sigma^2 g(Z_j'\alpha) \quad j = 1, \ldots, p \end{array}$$

donde las variables Z_{pi} pueden ser variables explicativas del modelo y $g(\cdot)$ no se especifica. Si todos los coeficientes de la combinación lineal $Z_j'\alpha$ fuesen cero, excepto α_0, la varianza sería homocedástica, $\sigma_i^2 = \sigma^2 g(\alpha_0)$. La hipótesis nula de homocedasticidad equivale a la siguiente hipótesis:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$$

donde se contrastan p restricciones lineales. El proceso de contraste es el siguiente:

a) Estimar por MCO el modelo original obteniendo los residuos correspondientes, $\hat{u}_{MCO,i}$.

b) Obtener la siguiente serie de residuos normalizados:

$$\hat{e}_i^2 = \frac{\hat{u}^2_{MCO,i}}{\hat{u}'\hat{u}/N} \qquad i = 1, 2, \ldots, N$$

donde: $\frac{\hat{u}'\hat{u}}{N} = \hat{\sigma}^2_{MV}$ es un estimador consistente, aunque sesgado, de la varianza de la perturbación.

c) Calcular la Suma de Cuadrados Explicada de la siguiente regresión realizada por MCO:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \ldots + \alpha_p Z_{pi} + \eta_i \quad i = 1, 2, \ldots, N$$

d) Se utiliza como estadístico de contraste el siguiente:

$$\frac{SCE}{2} \xrightarrow{H_0,d} \chi_p^2$$

siendo p los grados de libertad (el número de variables Z_j en la regresión auxiliar). Rechazamos a un nivel de significatividad α, si el valor muestral del estadístico excede del cuantil $\chi^2_{(p)\alpha}$.

Observaciones:

- Interpretación del contraste: Si los residuos fuesen homocedásticos, las variables $\{Z_j\}_{j=1}^p$ no deberían tener poder explicativo acerca de los residuos transformados y por tanto la SCE debería ser pequeña. Si $SCE/2$ es grande rechazaremos la H_0 y existiría heterocedasticidad.
- En el caso de que el contraste rechace la H_0 se podría dividir cada observación por $\sqrt{Z_j'\alpha}$ como una aproximación a la desviación típica de cada período. La estimación por MCO de este modelo transformado es equivalente a hacer MCG en el original.
- Candidatos a formar parte del vector Z: las variables explicativas o sus cuadrados, variables ficticias (grupos, estacionalidad, etc).

3.3. MCG: Mínimos Cuadrados Ponderados

En este tema nos estamos ocupando de relajar el supuesto clásico de homocedasticidad de la varianza de las perturbaciones. Como ya hemos visto si $E(uu') = \sigma^2\Omega$ el estimador MCO será lineal, insesgado y consistente pero no será eficiente. El problema radica en que nuestra matriz de varianzas y covarianzas de la perturbación ya no es de la forma $E(uu') = \sigma^2 I$. Esta hipótesis básica (en realidad recoge dos hipótesis básicas, homocedasticidad y no autocorrelación en las perturbaciones) es necesaria para obtener la siguiente matriz de varianzas y covarianzas del estimador MCO: $V(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}$, que bajo las hipótesis básicas proporciona varianzas $(\text{Var}(\hat{\beta}_{MCO,i}))$ mínimas. Incumplida esta hipótesis básica no podremos obtener las propiedades del estimador para las cuales esta hipótesis es necesaria. En concreto no obtendremos la misma matriz de varianzas y covarianzas para el estimador MCO y por lo tanto la obtenida no garantiza que las varianzas sean mínimas. El resto de propiedades se mantienen como ya vimos en el tema anterior.

En el criterio de estimación de Mínimos Cuadrados Ordinarios la función objetivo es:

$$\min_\beta(\hat{u}'\hat{u}) = \min_\beta \sum_{i=1}^{N} \hat{u}_i^2$$

esta función objetivo trata a todas las observaciones por igual ya que supone que la varianza de la perturbación es constante. Minimiza la distancia de la Función de Regresión Poblacional a los puntos de la muestra.

En presencia de heterocedasticidad, las varianzas de las perturbaciones serán distintas. Cuanto mayor sea la varianza de la perturbación, mayor será el peso de la misma dentro de la muestra. Si en estas circunstancias aplicamos el criterio MCO, que concede a todas las observaciones el mismo peso, el estimador que alcancemos no será apropiado. Lo indicado sería ponderar cada observación inversamente a su peso. Esto es lo que hace el criterio de MCG, cuya función objetivo es:

$$\min_\beta(\hat{u}'\Omega^{-1}\hat{u}) = \min_\beta \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\hat{u}_i^2$$

siendo $E(uu') = \sigma^2\Omega$. Dada la función objetivo lo que hace el criterio MCG es reconocer que es más importante que la Función de Regresión Muestral esté más cerca de los puntos con menor varianza, aunque sea a costa de quedar más alejada de otras observaciones, las de mayor varianza.

Para el caso particular de heterocedasticidad, la matriz Ω es diagonal y por tanto la función objetivo es una suma ponderada de residuos al cuadrado, donde se pondera cada observación inversamente a su varianza. Por ello en la literatura econométrica el estimador de MCG bajo heterocedasticidad aparece nombrado como estimador de Mínimos Cuadrados Ponderados.

Por otro lado, vistas las dos funciones objetivo podemos decir que el estimador MCO es un caso particular del estimador de MCG donde $\Omega = I$ $(\sigma_i^2 = \sigma^2)$.

A continuación veremos algunas situaciones donde la varianza de la perturbación es heterocedástica. En esta situación el estimador lineal, insesgado y óptimo es el estimador de Mínimos Cuadrados Generalizados que toma el nombre especial de Mínimos Cuadrados Ponderados, dado que consiste en ponderar la suma residual de cuadrados a minimizar.

3.3.1. Heterocedasticidad causada por una variable exógena del modelo

Sea el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{Ki} + u_i \qquad i = 1, 2, \ldots, N \tag{3.3}$$

donde: $E(u_i) = 0 \quad \forall i$,
$E(u_i^2) = \sigma^2 X_{2i}$ i=1, 2, ..., N,
$E(u_i u_j) = 0, \quad \forall i, j \quad i \neq j$.

En este caso la varianza de la perturbación depende de la variable X_{2i}, por tanto es heterocedástica. La estructura de la matriz de varianzas y covarianzas de la perturbación es la siguiente:

$$E(uu') = \sigma^2 \begin{pmatrix} X_{21} & 0 & 0 & \ldots & 0 \\ 0 & X_{22} & 0 & \ldots & 0 \\ 0 & 0 & X_{23} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & X_{2N} \end{pmatrix} = \sigma^2\Omega$$

por lo que Ω es conocida. El hecho de que la varianza de la perturbación del modelo a estimar sea heterocedástica, supone que el estimador de MCO en estas circunstancias aunque es lineal, insesgado y consistente es ineficiente. El estimador lineal, insesgado y óptimo de los parámetros del modelo es el estimador de Mínimos Cuadrados Generalizados, que se define:

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

Para aplicar este estimador disponemos de dos alternativas:

- **Alternativa 1:** Aplicar el estimador de MCG directamente a los datos donde:

$$\Omega^{-1} = \begin{pmatrix} \frac{1}{X_{21}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{X_{22}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{X_{23}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{X_{2N}} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} & \dots & X_{K1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} & \dots & X_{K2} \\ \vdots & \vdots & \vdots & \dots & \vdots & \ddots & \vdots \\ 1 & X_{2N} & X_{3N} & \dots & X_{kN} & \dots & X_{KN} \end{pmatrix} \qquad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}$$

Formamos ahora las matrices $(X'\Omega^{-1}X)$ y $(X'\Omega^{-1}Y)$:

$$(X'\Omega^{-1}X) = \begin{pmatrix} \sum_1^N \frac{1}{X_{2i}} & N & \sum_1^N \frac{X_{3i}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}}{X_{2i}} \\ N & \sum_1^N X_{2i} & \sum_1^N X_{3i} & \dots & \sum_1^N X_{Ki} \\ \sum_1^N \frac{X_{3i}}{X_{2i}} & \sum_1^N X_{3i} & \sum_1^N \frac{X_{3i}^2}{X_{2i}} & \dots & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_1^N \frac{X_{Ki}}{X_{2i}} & \sum_1^N X_{Ki} & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}^2}{X_{2i}} \end{pmatrix}$$

$$(X'\Omega^{-1}Y) = \begin{pmatrix} \sum_1^N \frac{Y_i}{X_{2i}} \\ \sum_1^N Y_i \\ \sum_1^N \frac{X_{3i}Y_i}{X_{2i}} \\ \vdots \\ \sum_1^N \frac{X_{Ki}Y_i}{X_{2i}} \end{pmatrix}$$

de donde:

$$\tilde{\beta}_{MCG} = \begin{pmatrix} \sum_1^N \frac{1}{X_{2i}} & N & \sum_1^N \frac{X_{3i}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}}{X_{2i}} \\ N & \sum_1^N X_{2i} & \sum_1^N X_{3i} & \dots & \sum_1^N X_{Ki} \\ \sum_1^N \frac{X_{3i}}{X_{2i}} & \sum_1^N X_{3i} & \sum_1^N \frac{X_{3i}^2}{X_{2i}} & \dots & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_1^N \frac{X_{Ki}}{X_{2i}} & \sum_1^N X_{Ki} & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}^2}{X_{2i}} \end{pmatrix}^{-1} \begin{pmatrix} \sum_1^N \frac{Y_i}{X_{2i}} \\ \sum_1^N Y_i \\ \sum_1^N \frac{X_{3i}Y_i}{X_{2i}} \\ \vdots \\ \sum_1^N \frac{X_{Ki}Y_i}{X_{2i}} \end{pmatrix}$$

La matriz de varianzas y covarianzas del estimador de MCG sería:

$$V(\tilde{\beta}_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2 \begin{pmatrix} \sum_1^N \frac{1}{X_{2i}} & N & \sum_1^N \frac{X_{3i}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}}{X_{2i}} \\ N & \sum_1^N X_{2i} & \sum_1^N X_{3i} & \dots & \sum_1^N X_{Ki} \\ \sum_1^N \frac{X_{3i}}{X_{2i}} & \sum_1^N X_{3i} & \sum_1^N \frac{X_{3i}^2}{X_{2i}} & \dots & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_1^N \frac{X_{Ki}}{X_{2i}} & \sum_1^N X_{Ki} & \sum_1^N \frac{X_{3i}X_{Ki}}{X_{2i}} & \dots & \sum_1^N \frac{X_{Ki}^2}{X_{2i}} \end{pmatrix}^{-1}$$

Un estimador insesgado de dicha matriz de varianzas y covarianzas es:

$$\widetilde{Var}(\tilde{\beta}_{MCG}) = \tilde{\sigma}^2_{MCG}(X'\Omega^{-1}X)^{-1}$$

donde:

$$\begin{aligned}\tilde{\sigma}^2_{MCG} &= \frac{\tilde{u}'_{MCG}\Omega^{-1}\tilde{u}_{MCG}}{N-K} = \frac{(Y - X\tilde{\beta}_{MCG})'\Omega^{-1}(Y - X\tilde{\beta}_{MCG})}{N-K} \\ &= \frac{Y'\Omega^{-1}Y - \tilde{\beta}'_{MCG}X'\Omega^{-1}Y}{N-K}\end{aligned}$$

siendo $Y'\Omega^{-1}Y = \sum_1^N \frac{Y_i^2}{X_{2i}}$.

- **Alternativa 2:** Estimar por MCO el modelo transformado:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \quad \Leftrightarrow \quad Y_\star = X_\star\beta + u_\star$$

donde P es la matriz de transformación tal que $PP' = \Omega$:

$$P = \begin{pmatrix} \sqrt{X_{21}} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{X_{22}} & 0 & \dots & 0 \\ 0 & 0 & \sqrt{X_{23}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{X_{2N}} \end{pmatrix} \quad P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{X_{21}}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{X_{22}}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{X_{22}}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{X_{2N}}} \end{pmatrix}$$

Buscamos ahora las matrices transformadas $P^{-1}X$ y $P^{-1}Y$:

$$X_\star = P^{-1}X = \begin{pmatrix} \frac{1}{\sqrt{X_{21}}} & \sqrt{X_{21}} & \frac{X_{31}}{\sqrt{X_{21}}} & \dots & \frac{X_{K1}}{\sqrt{X_{21}}} \\ \frac{1}{\sqrt{X_{22}}} & \sqrt{X_{22}} & \frac{X_{32}}{\sqrt{X_{22}}} & \dots & \frac{X_{K2}}{\sqrt{X_{22}}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{X_{2N}}} & \sqrt{X_{2N}} & \frac{X_{3N}}{\sqrt{X_{2N}}} & \dots & \frac{X_{KN}}{\sqrt{X_{2N}}} \end{pmatrix} \qquad Y_\star = P^{-1}Y = \begin{pmatrix} \frac{Y_1}{\sqrt{X_{21}}} \\ \frac{Y_2}{\sqrt{X_{22}}} \\ \vdots \\ \frac{Y_N}{\sqrt{X_{2N}}} \end{pmatrix}$$

Podemos escribir el modelo transformado como:

$$\frac{Y_i}{\sqrt{X_{2i}}} = \beta_1\frac{1}{\sqrt{X_{2i}}} + \beta_2\sqrt{X_{2i}} + \beta_3\frac{X_{3i}}{\sqrt{X_{2i}}} + \ldots + \beta_k\frac{X_{Ki}}{\sqrt{X_{2i}}} + \frac{u_i}{\sqrt{X_{2i}}} \qquad i = 1, 2, \ldots, N \quad (3.4)$$

Comprobamos que la nueva perturbación, la perturbación en el modelo transformado, es homocedástica, no autocorrelada y de media cero:

$$\begin{aligned} E\left(\frac{u_i}{\sqrt{X_{2i}}}\right) &= \frac{E(u_i)}{\sqrt{X_{2i}}} = 0 \quad \forall i \\ Var\left(\frac{u_i}{\sqrt{X_{2i}}}\right) &= E\left(\frac{u_i}{\sqrt{X_{2i}}} - E\left(\frac{u_i}{\sqrt{X_{2i}}}\right)\right)^2 = E\left(\frac{u_i}{\sqrt{X_{2i}}}\right)^2 \\ &= \frac{E(u_i^2)}{X_{2i}} = \frac{\sigma^2 X_{2i}}{X_{2i}} = \sigma^2 \quad \forall i \\ Cov\left(\frac{u_i}{\sqrt{X_{2i}}}, \frac{u_j}{\sqrt{X_{2j}}}\right) &= E\left(\frac{u_i}{\sqrt{X_{2i}}}\frac{u_j}{\sqrt{X_{2j}}}\right) = \frac{E(u_iu_j)}{\sqrt{X_{2i}}\sqrt{X_{2j}}} = 0 \quad \forall i, j \quad i \neq j \end{aligned}$$

A la vista de las propiedades de la perturbación del modelo transformado el estimador de MCO es lineal, insesgado y eficiente. El estimador lo denotamos:

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

donde $X_\star = P^{-1}X$ y $Y_\star = P^{-1}Y$.
En el modelo transformado debemos notar:

a) No existe término independiente ya que el término $\left(\frac{1}{\sqrt{X_{2i}}}\right)$ es una variable cuyo valor cambia con cada observación i y no una constante.

b) La correlación entre la nueva variable endógena, $\frac{Y_i}{\sqrt{X_{2i}}}$, y el regresor $\sqrt{X_{2i}}$ es mayor que la original y tiene carácter espúreo.

3.3.2. Omisión de una variable relevante

Veamos qué sucede cuando se omite una variable relevante en el modelo. Si el modelo realmente se especifica como:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad i = 1, 2, \ldots, N \\ u_i &= Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} \\ \hat{u}_i &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} \end{aligned}$$

Pero estimamos:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + v_i \qquad i = 1, 2, \ldots, N \\ v_i &= Y_i - \beta_1 - \beta_2 X_{2i} = u_i + \beta_3 X_{3i} \\ \hat{v}_i &= Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} \\ &= \hat{u}_i - (\tilde{\beta}_1 - \hat{\beta}_1) - (\tilde{\beta}_2 - \hat{\beta}_2) X_{2i} + \hat{\beta}_3 X_{3i} \end{aligned}$$

En este contexto de omisión los estimadores MCO son sesgados en general:
$\tilde{\beta}_1$ sesgado salvo que $Cov(X_{2i}, X_{3i}) = 0$ y $\overline{X}_3 = 0$, además $(\tilde{\beta}_1 - \hat{\beta}_1) \neq 0$.
$\tilde{\beta}_2$ será sesgado salvo que $Cov(X_{2i}, X_{3i}) = 0$, además $(\tilde{\beta}_1 - \hat{\beta}_1) \neq 0$

En consecuencia, tras un análisis de los residuos $\hat{v}_i$ tanto gráfico como mediante tests, es muy probable que el investigador llegue a la conclusión de que

$$V(v_i) = a + bX_{3i}$$

o incluso, dependiendo de los sesgos, que $V(v_i) = a + bX_{2i} + cX_{3i}$.

En este caso el investigador no debe corregir la heterocedasticidad directamente, sino que debe en primer lugar especificar correctamente el modelo y sólo corregir por heterocedasticidad si el modelo está correctamente especificado y su varianza es heterocedástica.

3.3.3. Datos agregados

Sea el modelo

$$Y_j = \alpha + \beta X_j + u_j \qquad j = 1, 2, \ldots, N \tag{3.5}$$

donde $u_j \sim NID(0, \sigma^2)$, es decir, la perturbación del modelo tiene media cero, es homocedástica y no autocorrelada.
Supongamos que el número de observaciones N es tal que su manejabilidad aconseja agrupar las observaciones en m-grupos de n_i observaciones cada uno. Así, utilizaremos datos agregados. Como observación del grupo i-ésimo tomamos la media aritmética dentro del grupo.

El modelo a estimar sería:

$$\overline{Y}_i = \alpha + \beta \overline{X}_i + \overline{u}_i \qquad i = 1, 2, \ldots, m \tag{3.6}$$

donde:

$$\overline{Y}_i = \frac{\sum_{j \in i}^{n_i} Y_j}{n_i} \qquad \overline{X}_i = \frac{\sum_{j \in i}^{n_i} X_j}{n_i} \qquad \overline{u}_i = \frac{\sum_{j \in i}^{n_i} u_j}{n_i}$$

En este caso nos interesan las propiedades de la perturbación en el modelo anterior:

$$\overline{u}_i = \frac{\sum_{j \in i}^{n_i} u_j}{n_i} \qquad i = 1, 2, \ldots, m \qquad j = 1, 2, \ldots, n_i$$

que son las siguientes:

$$\begin{aligned}
E(\overline{u}_i) &= E\left(\frac{\sum_{j \in i}^{n_i} u_j}{n_i}\right) = \frac{\sum_{j \in i}^{n_i} E(u_j)}{n_i} = 0 \qquad i = 1, 2, \ldots, m \\
Var(\overline{u}_i) &= E\left(\overline{u}_i - E(\overline{u}_i)\right)^2 = E(\overline{u}_i)^2 = E\left(\frac{\sum_{j \in i}^{n_i} u_j}{n_i}\right)^2 = \\
&= \frac{\sum_{j \in i}^{n_i} E(u_j^2)}{(n_i)^2} = \frac{n_i \sigma^2}{(n_i)^2} = \frac{\sigma^2}{n_i} \qquad i = 1, 2, \ldots, m \\
Cov(\overline{u}_i, \overline{u}_\ell) &= E(\overline{u}_i \overline{u}_\ell) = E\left(\frac{\sum_{j \in i}^{n_j} u_i}{n_i} \frac{\sum_{j \in \ell}^{n_\ell} u_\ell}{n_\ell}\right) = 0 \qquad \forall i, \ell \qquad i \neq \ell
\end{aligned}$$

La varianza de la nueva perturbación $\overline{u}_i$ es heterocedástica porque depende de n_i, el número de observaciones dentro de cada grupo. La matriz de varianzas y covarianzas tiene la forma siguiente:

$$E(uu') = \sigma^2 \begin{pmatrix} \frac{1}{n_1} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{n_2} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{n_m} \end{pmatrix} = \sigma^2 \Omega$$

donde Ω es conocida siempre y cuando el número de observaciones de cada grupo sea conocido. El hecho de que la varianza de la perturbación del modelo a estimar sea heterocedástica, supone que el estimador de MCO en estas circunstancias, aunque sea lineal, insesgado y consistente, es ineficiente. El estimador lineal, insesgado y óptimo de los parámetros del modelo es el estimador de MCG. Para aplicar este estimador tenemos dos posibilidades:

- **Alternativa 1:** Aplicar el estimador de MCG directamente a los datos. El estimador de MCG se define:
$\tilde{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$
donde:

$$\Omega^{-1} = \begin{pmatrix} n_1 & 0 & 0 & \dots & 0 \\ 0 & n_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n_m \end{pmatrix}$$

y las matrices X e Y del modelo a estimar son:

$$X = \begin{pmatrix} 1 & \overline{X}_1 \\ 1 & \overline{X}_2 \\ \vdots & \vdots \\ 1 & \overline{X}_m \end{pmatrix} \qquad Y = \begin{pmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{Y}_m \end{pmatrix}$$

Formamos ahora las matrices $(X'\Omega^{-1}X)$ y $(X'\Omega^{-1}Y)$:

$$X'\Omega^{-1}X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \overline{X}_1 & \overline{X}_2 & \dots & \overline{X}_m \end{pmatrix} \begin{pmatrix} n_1 & 0 & 0 & \dots & 0 \\ 0 & n_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n_m \end{pmatrix} \begin{pmatrix} 1 & \overline{X}_1 \\ 1 & \overline{X}_2 \\ \vdots & \vdots \\ 1 & \overline{X}_m \end{pmatrix}$$

$$= \begin{pmatrix} \sum_1^m n_i & \sum_1^m n_i\overline{X}_i \\ \sum_1^m n_i\overline{X}_i & \sum_1^m n_i\overline{X}_i^2 \end{pmatrix}$$

$$X'\Omega^{-1}Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \overline{X}_1 & \overline{X}_2 & \dots & \overline{X}_m \end{pmatrix} \begin{pmatrix} n_1 & 0 & 0 & \dots & 0 \\ 0 & n_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n_m \end{pmatrix} \begin{pmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{Y}_m \end{pmatrix} =$$

$$\begin{pmatrix} \sum_1^m n_i\overline{Y}_i \\ \sum_1^m n_i\overline{X}_i\overline{Y}_i \end{pmatrix}$$

de donde:

$$\tilde{\beta}_{MCG} = \begin{pmatrix} \sum_1^m n_i & \sum_1^m n_i\overline{X}_i \\ \sum_1^m n_i\overline{X}_i & \sum_1^m n_i\overline{X}_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_1^m n_i\overline{Y}_i \\ \sum_1^m n_i\overline{X}_i\overline{Y}_i \end{pmatrix}$$

La matriz de varianzas y covarianzas del estimador de MCG sería:

$$Var(\tilde{\beta}_{MCG}) = \sigma^2(X'\Omega^{-1}X)^{-1} = \sigma^2 \begin{pmatrix} \sum_1^m n_i & \sum_1^m n_i\overline{X}_i \\ \sum_1^m n_i\overline{X}_i & \sum_1^m n_i\overline{X}_i^2 \end{pmatrix}^{-1}$$

Un estimador insesgado de dicha matriz de varianzas y covarianzas sería:

$$\widetilde{Var}(\tilde{\beta}_{MCG}) = \tilde{\sigma}^2_{MCG}(X'\Omega^{-1}X)^{-1}$$

donde:

$$
\begin{aligned}
\tilde{\sigma}^2_{MCG} &= \frac{\tilde{u}'_{MCG}\Omega^{-1}\tilde{u}_{MCG}}{N-K} = \frac{(Y - X\tilde{\beta}_{MCG})'\Omega^{-1}(Y - X\tilde{\beta}_{MCG})}{N-K} \\
&= \frac{Y'\Omega^{-1}Y - \tilde{\beta}'_{MCG}X'\Omega^{-1}Y}{N-K}
\end{aligned}
$$

siendo $Y'\Omega^{-1}Y = \sum_1^m n_i\overline{Y}_i^2$.

- **Alternativa 2:** Aplicar MCO al modelo transformado:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \quad \Leftrightarrow \quad Y_\star = X_\star\beta + u_\star$$

donde P es la matriz de transformación tal que $PP' = \Omega$ y:

$$
P = \begin{pmatrix} \frac{1}{\sqrt{n_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{n_2}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{n_m}} \end{pmatrix} \qquad
P^{-1} = \begin{pmatrix} \sqrt{n_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{n_m} \end{pmatrix}
$$

Buscamos ahora las matrices transformadas $P^{-1}X$ y $P^{-1}Y$:

$$
X_\star = P^{-1}X = \begin{pmatrix} \sqrt{n_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{n_m} \end{pmatrix} \begin{pmatrix} 1 & \overline{X}_1 \\ 1 & \overline{X}_2 \\ \vdots & \vdots \\ 1 & \overline{X}_m \end{pmatrix} = \begin{pmatrix} \sqrt{n_1} & \sqrt{n_1}\overline{X}_1 \\ \sqrt{n_2} & \sqrt{n_2}\overline{X}_2 \\ \vdots & \vdots \\ \sqrt{n_m} & \sqrt{n_m}\overline{X}_m \end{pmatrix}
$$

$$
Y_\star = P^{-1}Y = \begin{pmatrix} \sqrt{n_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{n_m} \end{pmatrix} \begin{pmatrix} \overline{Y}_1 \\ \overline{Y}_2 \\ \vdots \\ \overline{Y}_m \end{pmatrix} = \begin{pmatrix} \sqrt{n_1}\overline{Y}_1 \\ \sqrt{n_2}\overline{Y}_2 \\ \vdots \\ \sqrt{n_m}\overline{Y}_m \end{pmatrix}
$$

Podemos escribir el modelo transformado como:

$$\sqrt{n_i}\,\overline{Y}_i = \sqrt{n_i}\,\alpha + \sqrt{n_i}\,\overline{X}_i + \sqrt{n_i}\,\overline{u}_i \qquad i = 1, 2, \dots, m. \tag{3.7}$$

Este modelo puede ser estimado por MCO y los estimadores serán lineales, insesgados y eficientes si y sólo si la nueva perturbación es esférica, es decir, tiene media cero, varianza constante y covarianzas nulas. Demostración:

$$
\begin{aligned}
E(\sqrt{n_i}\overline{u}_i) &= \sqrt{n_i}E(\overline{u}_i) = 0 \quad \forall i \\
V(\sqrt{n_i}\overline{u}_i) &= E(\sqrt{n_i}\overline{u}_i - E(\sqrt{n_i}\overline{u}_i))^2 = E(\sqrt{n_i}\overline{u}_i)^2 = n_iE(\overline{u}_i^2) = n_i\frac{\sigma^2}{n_i} = \sigma^2 \quad \forall i \\
Cov(\sqrt{n_i}\overline{u}_i, \sqrt{n_\ell}\overline{u}_\ell) &= E(\sqrt{n_i}\overline{u}_i\sqrt{n_\ell}\overline{u}_\ell) = \sqrt{n_i}\sqrt{n_\ell}E(\overline{u}_i\overline{u}_\ell) = 0 \quad \forall i, \ell \quad i \neq \ell
\end{aligned}
$$

Por tanto en el modelo transformado la perturbación tiene media cero, es homocedástica y no autocorrelada, el estimador de MCO en el modelo transformado es ELIO y por tanto el estimador MCG en el modelo original también es ELIO.

- Un caso particular de este ejemplo sería el caso en que todos los grupos tuvieran igual número de observaciones, $n_i = n \quad i = 1, 2, \ldots, m$. En este caso la varianza de la perturbación del modelo sería :

$$V(\overline{u}_i) = \frac{\sigma^2}{n}$$

homodedástica con matriz de varianzas y covarianzas

$$E(uu') = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{n} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{1}{n} \end{pmatrix} = \frac{\sigma^2}{n} I_m$$

El modelo original sería estimable por MCO con propiedades adecuadas.

3.3.4. Coeficientes cambiantes

- **Variación determinista**

Supongamos el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} + u_i \qquad i = 1, 2, \ldots, N \tag{3.8}$$

donde $\beta_{3i} = \beta + \alpha Z_i$ siendo β y α dos constantes desconocidas y Z_i una variable determinista (grupo, estacionalidad, tendencia, etc). En este caso el modelo a estimar sería:

$$Y_i = \beta_1 + \beta_2 X_{2i} + (\beta + \alpha Z_i) X_{3i} + u_i$$

Pero si debido a la mala especificación de los coeficientes estimamos:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta X_{3i} + v_i$$

donde

$$v_i = Y_i - \beta_1 - \beta_2 X_{2i} - \beta X_{3i} = u_i + \alpha Z_i X_{3i}$$

Podríamos llegar a la conclusión, a semejanza del caso de omisión, de que las perturbaciones presentan heterocedasticidad. La solución en este caso no es la estimación por MCG sino especificar correctamente el modelo.

- **Variación aleatoria**

Otro caso de existencia de heterocedasticidad sería aquel en que alguno de los coeficientes del modelo es una variable aleatoria. Supongamos el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} + u_i \qquad i = 1, 2, \ldots, N \tag{3.9}$$

donde $\beta_{3i} = \beta + \epsilon_i$ siendo β una constante desconocida y ϵ un término de error. En este caso el modelo a estimar sería:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + (\beta + \epsilon_i) X_{3i} + u_i & (3.10) \\ Y_i &= \beta_1 + \beta_2 X_{2i} + \beta X_{3i} + v_i & (3.11) \end{aligned}$$

donde $v_i = u_i + \epsilon_i X_{3i}$. Si suponemos:

$$u_i \sim iid(0, \sigma_u^2) \qquad \epsilon_i \sim iid(0, \sigma_\epsilon^2) \qquad Cov(u_i, \epsilon_i) = 0 \quad \forall i$$

podremos obtener la siguiente distribución para v_i:

$$\begin{aligned}
E(v_i) &= E(u_i + \epsilon_i X_{3i}) = E(u_i) + X_{3i}E(\epsilon_i) = 0 \quad \forall i \\
\\
V(v_i) &= E(u_i + \epsilon_i X_{3i})^2 = E(u_i^2) + X_{3i}^2 E(\epsilon_i^2) + X_{3i}E(u_i\epsilon_i) = \\
&= \sigma_u^2 + X_{3i}^2\sigma_\epsilon^2 \\
Cov(v_i, v_\ell) &= E((u_i + \epsilon_i X_{3i})(u_\ell + \epsilon_\ell X_{3\ell})) = 0 \quad \forall i, \ell \quad i \neq \ell
\end{aligned}$$

La varianza de v_i tiene media cero, es heterocedástica ya que depende de X_{3i} y no autocorrelada siempre que $Cov(u_i, \epsilon_i) = 0 \quad \forall i$:

$$v_i \sim iid(0, \sigma_u^2 + X_{3i}^2\sigma_\epsilon^2)$$

En estas circunstancias el estimador de MCO es lineal, insesgado, consistente pero ineficiente.

El modelo debe ser estimado por MCG. La matriz de varianzas y covarianzas de la perturbación es:

$$E(vv') = \begin{pmatrix} \sigma_u^2 + X_{31}^2\sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 + X_{32}^2\sigma_\epsilon^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 + X_{3N}^2\sigma_\epsilon^2 \end{pmatrix} = \sum$$

Si σ_u^2 y σ_ϵ^2 fuesen conocidos, tenemos dos posibilidades para estimar el modelo. La primera aplicar el estimador de MCG a los datos directamente utilizando $\tilde{\beta}_{MCG} = (X' \sum^{-1} X)^{-1}(X' \sum^{-1} Y)$ con $V(\tilde{\beta}_{MCG}) = (X' \sum^{-1} X)^{-1}$ y la segunda estimar por MCO el siguiente modelo transformado:

$$P^{-1}Y = P^{-1}X\beta + P^{-1}u \quad \Leftrightarrow \quad Y_\star = X_\star\beta + u_\star$$

donde:

$$P = \begin{pmatrix} \sqrt{\sigma_u^2 + X_{31}^2\sigma_\epsilon^2} & 0 & \dots & 0 \\ 0 & \sqrt{\sigma_u^2 + X_{32}^2\sigma_\epsilon^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\sigma_u^2 + X_{3N}^2\sigma_\epsilon^2} \end{pmatrix}$$

$$P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\sigma_u^2 + X_{31}^2\sigma_\epsilon^2}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_u^2 + X_{32}^2\sigma_\epsilon^2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\sigma_u^2 + X_{3N}^2\sigma_\epsilon^2}} \end{pmatrix}$$

Así, podemos escribir el modelo transformado como:

$$\frac{Y_i}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}} = \beta_1\frac{1}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}} + \beta_2\frac{X_{2i}}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}} + \beta_3\frac{X_{3i}}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}} + \frac{u_i}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}}$$

y las propiedades de la perturbación en este modelo transformado serían:

$$
\begin{aligned}
E\left(\frac{u_i}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}}\right) &= \frac{E(u_i)}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}} = 0 \quad \forall i \\
Var\left(\frac{u_i}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}}\right) &= \frac{E(u_i^2)}{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2} = \frac{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2} = 1 \quad \forall i \\
Cov\left(\frac{u_i}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}}, \frac{u_\ell}{\sqrt{\sigma_u^2 + X_{3\ell}^2\sigma_\epsilon^2}}\right) &= \frac{E(u_i u_\ell)}{\sqrt{\sigma_u^2 + X_{3i}^2\sigma_\epsilon^2}\sqrt{\sigma_u^2 + X_{3\ell}^2\sigma_\epsilon^2}} = 0
\end{aligned}
$$

es decir, media cero, homocedástica y no autocorrelada. La estimación por MCO del modelo transformado proporciona estimadores ELIO.

Sin embargo, en este caso la matriz $\sum$ es **desconocida** ya que depende de σ_u^2 y de σ_ϵ^2 que son desconocidas por lo que para poder aplicar cualquiera de las dos alternativas de estimación, es necesario estimar estas varianzas previamente. Trataremos este caso en la siguiente sección.

3.4. MCGF: Mínimos Cuadrados Generalizados Factibles

En el modelo de regresión lineal general

$$Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega)$$

donde Ω es desconocida. Nos preguntamos cómo estimar los parámetros desconocidos β. Debemos responder que:

a) Si Ω es conocida, el estimador $\tilde{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$ es un estimador lineal, insesgado, eficiente y consistente de los coeficientes desconocidos, β.

b) Si Ω es desconocida, habitualmente lo es, lo tendremos que estimar para sustituirlo en la expresión del estimador de MCG y obtener así el estimador de Mínimos Cuadrados Generalizados Factible (MCGF):

$$
\begin{aligned}
&\tilde{\beta}_{MCGF} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y \qquad && \text{si} \quad E(uu') = \sigma^2\Omega \\
&\tilde{\beta}_{MCGF} = (X'\hat{\textstyle\sum}^{-1}X)^{-1}(X'\hat{\textstyle\sum}^{-1}Y) \qquad && \text{si} \quad E(uu') = \textstyle\sum
\end{aligned}
$$

3.4.1. Cómo estimar la matriz Ω (ó σ)

Si Ω es desconocida, en el modelo nos enfrentamos a la estimación de K-coeficientes y N-varianzas:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + u_i$$

$$E(uu') = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_N^2 \end{pmatrix}.$$

La estimación de $K + N$ parámetros con N observaciones no es posible si lo que deseamos es estimar estos parámetros de forma precisa. Es necesario establecer algún tipo de restricción sobre la forma funcional de los parámetros desconocidos de Ω. Habitualmente, se modela la varianza de la perturbación en función de un conjunto de parámetros θ y un conjunto de observaciones Z_i, que pueden o no formar parte del conjunto de regresores del modelo, pero en todo caso la información sobre las mismas es conocida. Este supuesto reduce el número de parámetros a estimar siempre que Z_i sea un vector de orden $(S \times 1)$, θ un vector de $(S \times 1)$ ó $((S + 1) \times 1)$ parámetros, siendo $K + S < N$. De este modo los parámetros del modelo serían estimables. Así si proponemos:

$$\sigma_i^2 = f(Z_i, \theta) \quad \forall i \quad \text{tal que} \quad \Omega = \Omega(\theta)$$

donde la función $f(\cdot)$ es la que mejor se ajusta a la información disponible (lineal, cuadrática, exponencial, etc). Entonces, una vez obtenido el estimador de θ, $\hat{\theta}$, tendremos definido el estimador de $\Omega(\theta)$, $\widehat{\Omega} = \Omega(\widehat{\theta})$ y podremos estimar el vector de coeficientes $\tilde{\beta}_{MCGF} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$

3.4.2. ¿Qué propiedades exigimos a $\hat{\Omega}$?

El estimador:

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

es una función no lineal de Y lo que dificulta la derivación analítica de sus propiedades en muestras finitas.

Por ello, nos interesaremos sólo por sus propiedades en muestras grandes o propiedades asintóticas, es decir, consistencia y normalidad asintótica, para lo que es necesario que $\hat{\Omega}$ sea un estimador consistente de Ω. Si $\hat{\Omega}$ es consistente se puede demostrar que bajo ciertas condiciones de regularidad $\tilde{\beta}_{MCGF}$ posee propiedades asintóticas deseables:

a) $\tilde{\beta}_{MCGF}$ es consistente.

b) $\tilde{\beta}_{MCGF}$ es asintóticamente normal:

$$\sqrt{N}(\tilde{\beta}_{MCGF} - \beta) \xrightarrow{d} N\left(0, \sigma^2 \lim_{N\to\infty}\left(\frac{X'\Omega^{-1}X}{N}\right)^{-1}\right)$$

donde:

$$\lim_{N\to\infty}\left(\frac{X'\Omega^{-1}X}{N}\right)^{-1} = G^{-1}$$

El estimador de MCGF es un estimador de Mínimos Cuadrados en dos etapas. En la primera etapa se busca un estimador consistente de θ para que $\hat{\Omega} = \Omega(\hat{\theta})$ sea consistente. En la segunda etapa se sustituye $\hat{\Omega}$ en la función objetivo y ésta se minimiza con respecto a β para obtener el estimador de MCGF.

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

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

Como aproximación a la estimación de Ω generalmente se utiliza la relación entre u_i y $\hat{u}_i$ y se propone:

$$\begin{aligned} \sigma_i^2 &= f(\theta, Z_i) \\ \hat{u}_i^2 &= f(\theta, Z_i) + error \\ \hat{u}_i^2 &= \theta_1 + \theta_2 Z_{2i} + \theta_3 Z_{3i} + \ldots + \theta_S Z_{Si} + \text{error} \end{aligned} \tag{3.12}$$

donde:

$$\hat{u} = Y - X\hat{\beta}_{MCO}$$

El estimador MCO del modelo auxiliar (3.12) proporciona estimadores consistentes $\{\hat{\theta}_s\}_{s=1}^S$, así $\hat{\Omega} = \Omega(\hat{\theta})$ es consistente y podemos obtener el estimador por MCGF de los parámetros desconocidos del modelo.

Observaciones:

a) Si Ω es conocida el estimador ELIO y consistente es:

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

b) Si Ω es desconocida y $\hat{\Omega}$ es consistente el estimador consistente y asintóticamente eficiente es:

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

pero nada garantiza que en muestras pequeñas la varianza sea mínima.

3.4.3. Ejercicios

• Ejercicio 1

En el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{Ki} + u_i \qquad i = 1, 2, \ldots, N \Leftrightarrow Y_i = X_i'\beta + u_i$$

donde $E(u_i) = 0 \quad \forall i$ y $V(u_i) = \sigma_i^2 = \sigma^2 E(Y_i)$

¿Cómo estimamos los parámetros desconocidos del modelo anterior?

Solución:

Proponemos como estimador el estimador de MCGF: $\tilde{\beta}_{MCGF} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}Y$ dado que Ω es desconocida.

$$V(u_i) = \sigma^2(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{Ki}) = \sigma^2(X_i'\beta)$$

donde $X_i' = (1 \;\; X_{2i} \ldots X_{Ki})$. Escribimos la matriz de varianzas y covarianzas de la perturbación.

$$E(uu') = \\ \sigma^2 \begin{pmatrix} \beta_1 + \beta_2 X_{21} + \ldots + \beta_k X_{K1} & 0 & \ldots & 0 \\ 0 & \beta_1 + \beta_2 X_{22} + \ldots + \beta_k X_{K2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \beta_1 + \beta_2 X_{2N} + \ldots + \beta_k X_{KN} \end{pmatrix}$$

$$= \sigma^2 \begin{pmatrix} X_1'\beta & 0 & \dots & 0 \\ 0 & X_2'\beta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_N'\beta \end{pmatrix} \qquad \text{de donde:} \qquad \hat{\Omega} = \begin{pmatrix} X_1'\hat{\beta} & 0 & \dots & 0 \\ 0 & X_2'\hat{\beta} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_N'\hat{\beta} \end{pmatrix}$$

donde $\hat{\beta} = \hat{\beta}_{MCO}$ se obtiene de aplicar MCO a la ecuación original.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{Ki} + u_i \qquad i = 1, 2, \ldots N,$$

Dado que $\hat{\beta}_{MCO}$ es un estimador consistente, $\hat{\Omega}$ será consistente y también lo será $\hat{\beta}_{MCGF}$.

• **Ejercicio 2**

Supongamos el modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} + u_i \qquad i = 1, 2, \ldots, N \tag{3.13}$$

donde $E(u_i) = 0 \quad \forall i$ y ; $V(u_i) = a + bX_{3i}^2$ siendo a y b constantes desconocidas. ¿Cómo se estiman los parámetros desconocidos del modelo anterior?

Solución:
En este caso debemos estimar el modelo por MCGF para lo cual podemos aplicar MCO en el siguiente modelo transformado:

$$\frac{Y_i}{\sqrt{a + bX_{3i}^2}} = \beta_1 \frac{1}{\sqrt{a + bX_{3i}^2}} + \beta_2 \frac{X_{2i}}{\sqrt{a + bX_{3i}^2}} + \beta_3 \frac{X_{3i}}{\sqrt{a + bX_{3i}^2}} + \frac{u_i}{\sqrt{a + bX_{3i}^2}} \tag{3.14}$$

Las propiedades de la perturbación en este modelo transformado serían:

$$E\left(\frac{u_i}{\sqrt{a + bX_{3i}^2}}\right) = \frac{E(u_i)}{\sqrt{a + bX_{3i}^2}} = 0 \quad \forall i$$

$$Var\left(\frac{u_i}{\sqrt{a + X_{3i}^2}}\right) = \frac{E(u_i^2)}{a + X_{3i}^2} = \frac{a + bX_{3i}^2}{a + bX_{3i}^2} = 1 \quad \forall i$$

$$Cov\left(\frac{u_i}{\sqrt{a + bX_{3i}^2}}, \frac{u_j}{\sqrt{a + bX_{3j}^2}}\right) = \frac{E(u_i u_j)}{\sqrt{a + bX_{3i}^2}\sqrt{a + bX_{3j}^2}} = 0 \quad \forall i, j \quad i \neq j$$

homocedástica y no autocorrelada. La estimación por MCO del modelo transformado proporciona estimadores ELIO.
En este caso en el modelo transformado tenemos dos constantes desconocidas a y b que deben ser previamente estimadas para poder aplicar el estimador de MCO al modelo (3.14). Para obtener estimadores consistentes de a y b podemos proceder de la forma siguiente:

a) Aplicamos MCO en el modelo (3.13) sin tener en cuenta la existencia de heterocedasticidad. Guardamos los residuos de mínimos cuadrados ordinarios.

b) Estimamos la siguiente regresión auxiliar:

$$\hat{u}^2_{MCO,i} = a + bX^2_{3i} + \epsilon_i \qquad i = 1, 2, \ldots, N$$

de esta regresión obtenemos $\hat{a}_{MCO}$ y $\hat{b}_{MCO}$ estimados consistentemente.

c) Sustituimos $\hat{a} = \hat{a}_{MCO}$ y $\hat{b} = \hat{b}_{MCO}$ en el modelo:

$$\frac{Y_i}{\sqrt{\hat{a} + \hat{b}X^2_{3i}}} = \beta_1 \frac{1}{\sqrt{\hat{a} + \hat{b}X^2_{3i}}} + \beta_2 \frac{X_{2i}}{\sqrt{\hat{a} + \hat{b}X^2_{3i}}} + \beta_3 \frac{X_{3i}}{\sqrt{\hat{a} + \hat{b}X^2_{3i}}} + \frac{u_i}{\sqrt{\hat{a} + \hat{b}X^2_{3i}}}$$

y estimamos este modelo por MCO o lo que es lo mismo el modelo original por MCGF. Los estimadores así obtenidos serán consistentes dado que los estimadores de a y b a su vez lo son.

3.5. Estimador de White de $V(\hat{\beta}_{MCO})$

Si estimamos el MRLG por MCO:

$$Y = X\beta + u \qquad E(uu') = \Sigma$$

los estimadores son lineales, insesgados, consistentes pero ineficientes. La matriz de varianzas y covarianzas del estimador MCO en presencia de heterocedasticidad es:

$$V(\hat{\beta}_{MCO}) = (X'X)^{-1}(X'\Sigma X)(X'X)^{-1}$$

para calcular estas varianzas y covarianzas es necesario conocer Σ, lo mismo que para evaluar el estimador $\tilde{\beta}_{MCG}$, ELIO, en estas circunstancias.

Dada la dificultad que entraña el conocimiento de Σ, un estimador consistente de $V(\hat{\beta}_{MCO})$ resulta útil porque de esta forma se pueden derivar estadísticos válidos, al menos asintóticamente, para contrastar hipótesis sobre el vector de coeficientes β. Sabemos que si:

$$Y = X\beta + u \qquad u \sim N(0, \Sigma)$$

entonces:

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

White (1980) proporciona un estimador consistente de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$:

$$(X'X)^{-1}(X'SX)(X'X)^{-1}$$

White demuestra que:

$$\widehat{X'\Sigma X} = \frac{1}{N}X'SX$$

donde $S = \text{diag}(\hat{u}^2_{1,MCO} \;\; \hat{u}^2_{2,MCO} \;\; \ldots \hat{u}^2_{N,MCO})$ de forma que:

$$\text{plim}\left(\frac{1}{N}X'SX\right) = \text{plim}\left(\frac{1}{N}X'\Sigma X\right) = G.$$

Por lo tanto, la distribución asintótica del estimador MCO teniendo en cuenta el estimador consistente de White de su matriz de varianzas y covarianzas viene dado por:

$$\sqrt{N}(\hat{\beta}_{MCO} - \beta) \xrightarrow{a} N\left(0, \lim_{N\to\infty}\left(\frac{X'X}{N}\right)^{-1} \lim_{N\to\infty}\left(\frac{X'SX}{N}\right) \lim_{N\to\infty}\left(\frac{X'X}{N}\right)^{-1}\right)$$

$$\sqrt{N}(\hat{\beta}_{MCO} - \beta) \xrightarrow{a} N\left(0, Q^{-1}GQ^{-1}\right)$$

Esta matriz de varianzas y covarianzas es consistente y puede ser utilizada para hacer inferencia en muestras grandes, sin tener que especificar a priori la estructura de heterocedasticidad.

3.6. Contraste de restricciones lineales con Ω desconocida

Vamos a empezar esta sección recordando cómo hacer inferencia cuando Ω es conocida.

- Sea el modelo:

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

en este caso

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

Un estimador insesgado y consistente de σ^2 es:

$$\tilde{\sigma}^2_{MCG} = \frac{(Y - X\tilde{\beta}_{MCG})'\Omega^{-1}(Y - X\tilde{\beta}_{MCG})}{N-K}.$$

Podemos contrastar restricciones lineales sobre los parámetros β de la forma $H_0 : R\beta = r$ con el estadístico:

$$\frac{(R\tilde{\beta}_{MCG} - r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r)/q}{\tilde{\sigma}^2_{MCG}} \overset{H_0}{\sim} F_{(q,N-K)}$$

con las reglas de decisión habituales.

Si queremos contrastar la significatividad individual de una de las variables exógenas del modelo, dado que $q = 1$ podemos utilizar el estadístico:

$$\frac{\tilde{\beta}_{i,MCG}}{\widetilde{desv}(\tilde{\beta}_{i,MCG})} \overset{H_0}{\sim} t_{(N-K)}$$

- Sea el modelo:

$$Y = X\beta + u \qquad u \sim N(0, \Sigma) \qquad \Sigma \qquad \text{conocida}$$

en este caso

$$\tilde{\beta}_{MCG} \sim N(\beta,\ (X'\Sigma^{-1}X)^{-1})$$

En este caso no tenemos una constante a estimar en la matriz de varianzas y covarianzas de la perturbación. Contrastamos restricciones lineales del tipo $H_0 : R\beta = r$ con el estadístico:

$$(R\tilde{\beta}_{MCG} - r)'[R(X'\Sigma^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r) \overset{H_0}{\sim} \chi_{(q)}$$

El estadístico para el contraste de significatividad individual sería:

$$\frac{\tilde{\beta}_{i,MCG}}{desv(\tilde{\beta}_{i,MCG})} \overset{H_0}{\sim} N(0,1)$$

- Ahora supongamos que:
  $Y = X\beta + u \qquad E(uu') = \sigma^2\Omega = \sum \quad$ con $\quad \sum(o \quad \Omega)$ desconocidas.
  Debemos estimar el modelo por MCGF, $\tilde{\beta}_{MCGF} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y)$, estimador consistente de β si $\hat{\Omega}$ es un estimador consistente de Ω. Su distribución asintótica es:

$$\sqrt{N}(\tilde{\beta}_{MCGF} - \beta) \xrightarrow{d} N(0, \sigma^2 G^{-1}) \qquad G = \text{plim}\left(\frac{1}{N}X'\Omega^{-1}X\right)$$

  Para contrastar hipótesis nulas del tipo $H_0 : R\beta = r$ podemos utilizar el estadístico:

$$\frac{(R\tilde{\beta}_{MCGF} - r)'[R(X'\widehat{\Omega}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCGF} - r)}{\widehat{\sigma}^2} \xrightarrow{d,H_0} \chi^2_{(q)}$$

  con distribución asintótica χ^2 con q grados de libertad.

  Si queremos contrastar la significatividad individual de una de las variables exógenas del modelo, dado que $q = 1$ podemos utilizar el estadístico:

$$\frac{\tilde{\beta}_{i,MCGF}}{\tilde{desv}(\tilde{\beta}_{i,MCGF})} \xrightarrow{d,H_0} N(0,1)$$

- Si desconocemos cómo estimar Ω o Σ podemos optar por realizar inferencia utilizando el estimador de White, el contraste de hipótesis se realiza con el estadístico:

$$(R\hat{\beta}_{MCO} - r)'(R(X'X)^{-1}(X'SX)(X'X)^{-1}R')^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_{(q)}$$

  La expresión del estadístico para el contraste de significatividad individual no varía:

$$\frac{\tilde{\beta}_{i,MCO}}{\hat{desv}(\hat{\beta}_{i,MCO})_{White}} \xrightarrow{d,H_0} N(0,1)$$

  con las reglas de aceptación y rechazo habituales.

3.7. Predicción

Recordemos como hacer la predicción por punto y por intervalo con el estimador MCO:

$$\begin{aligned}
&Y = X\beta + u \qquad u \sim N(0, \sigma^2 I_N) \\
&Y_p = X'_p\beta + u_p \qquad p \notin [1, N] \\
&E(Y_p) = X'_p\beta \qquad p \notin [1, N] \\
&X'_p = (1 \quad X_{2,p} \ldots X_{K,p}) \\
&E(u_p) = 0 \quad V(u_p) = \sigma^2
\end{aligned}$$

a) Predicción por punto del valor y el valor esperado:

$$\hat{Y}_p = X'_p\hat{\beta} \qquad\qquad E(\hat{Y}_p) = X'_p\hat{\beta}$$

b) Predicción por intervalo del valor, Y_p:
   Error de predicción: $e_p = Y_p - \hat{Y}_p$
   Distribución: $e_p \sim N(0, \sigma^2(1 + X'_p(X'X)^{-1}X_p))$

Si σ^2 es conocida:

$$\frac{e_p}{des(e_p)} \overset{H_0}{\sim} N(0,1)$$

Si σ^2 es desconocida:

$$\frac{e_p}{\widehat{des}(e_p)} \overset{H_0}{\sim} t_{(N-K)}$$

A partir de estas distribuciones se obtienen los intervalos de confianza.

c) Predicción por intervalo del valor esperado, $E(Y_{N+1})$:
Error de predicción: $\epsilon_p = E(Y_p) - \hat{Y}_p$
Distribución: $\epsilon_p \sim N(0, \sigma^2(X_p'(X'X)^{-1}X_p))$
Si σ^2 es conocida:

$$\frac{\epsilon_p}{des(\epsilon_p)} \overset{H_0}{\sim} N(0,1)$$

Si σ^2 es desconocida:

$$\frac{\epsilon_p}{\widehat{des}(\epsilon_p)} \overset{H_0}{\sim} t_{(N-K)}$$

A partir de estas distribuciones se obtienen los intervalos de confianza.

Ahora, bajo el supuesto de heterocedasticidad tenemos que

$$\begin{array}{l}
Y = X\beta + u \qquad u \sim N(0, \sigma^2\Omega) \quad V(u_i) = \sigma^2 w_{ii} \\
Y_p = X_p'\beta + u_p \qquad p \notin [1, N] \\
E(Y_p) = X_p'\beta \qquad p \notin [1, N] \\
X_p' = (1 \;\; X_{2,p} \ldots X_{K,p}) \\
E(u_p) = 0 \quad V(u_p) = \sigma^2 w_{pp}
\end{array}$$

- **Predicción por punto** La predicción por punto del valor y el valor esperado de Y_p vienen dados por:

$$\hat{Y}_p = X_p'\hat{\beta}_{MCG} \qquad\qquad E(\hat{Y}_p) = X_p'\hat{\beta}_{MCG}$$

- **Predicción por intervalo del valor, Y_p:** Error de predicción:

$$\begin{aligned}
e_p &= Y_p - \hat{Y}_p \\
&= X_p'\beta + u_p - (X_p'\tilde{\beta}_{MCG}) \\
&= -X_p'(\hat{\beta}_{MCG} - \beta) + u_p
\end{aligned}$$

Distribución:

$$\begin{aligned}
E(e_p) &= E(-X_p'(\hat{\beta}_{MCG} - \beta) + u_p) = 0 \\
\mathrm{V}(e_p) &= E[(e_p - E(e_p))^2] = E(e_p e_p') \\
&= E[(-X_p'(\tilde{\beta}_{MCG} - \beta) + u_p)(-X_p'(\tilde{\beta}_{MCG} - \beta) + u_p)'] \\
&= E(u_p^2) + X_p'E(\tilde{\beta}_{MCG} - \beta)(\tilde{\beta}_{MCG} - \beta)'X_p - 2X_p'E((\tilde{\beta}_{MCG} - \beta)u_p) \\
&= \mathrm{V}(u_p) + X_p'\mathrm{V}(\tilde{\beta}_{MCG})X_p - 0 \\
&= \mathrm{V}(u_p) + \sigma^2 X_p'(X'\Omega^{-1}X)^{-1}X_p
\end{aligned}$$

bajo normalidad de las perturbaciones, $u \sim N(0, \sigma^2\Omega)$:

$$e_p \sim N(0, V(u_p) + X_p' V(\tilde{\beta}_{MCG}) X_p)$$
$$e_p \sim N(0, V(u_p) + \sigma^2 X_p' (X'\Omega^{-1}X)^{-1} X_p)$$

de donde

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2} \sqrt{V(u_p) + X_p' V(\tilde{\beta}_{MCG}) X_p}\right]$$

Posibilidades:

a) Supongamos que $V(u) = \sigma^2\Omega$ conocida. En este caso podemos hacer inferencia con la distribución normal:

$$e_p \sim N(0, \sigma^2(w_{pp} + X_p'(X'\Omega^{-1}X)^{-1}X_p))$$

de donde

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2} \sqrt{\sigma^2(w_{pp} + \sigma^2 X_p'(X'\Omega^{-1}X)^{-1}X_p)}\right]$$

b) Supongamos que $V(u) = \sigma^2\Omega$ con Ω conocida pero σ^2 desconocido. En este otro caso tenemos:

$$\frac{e_p}{\sqrt{\tilde{\sigma}^2_{MCG}(w_{pp} + X_p'(X'\Omega^{-1}X)^{-1}X_p)}} \sim t_{(N-K)}$$

de donde

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm t_{(N-K)\alpha/2} \sqrt{\tilde{\sigma}^2_{MCG}(w_{pp} + X_p'(X'\Omega^{-1}X)^{-1}X_p)}\right]$$

c) Supongamos que $V(u) = \sigma^2\Omega$ con Ω y σ^2 desconocidos pero estimables. En este contexto tenemos que estimar el modelo por MCGF por lo que el intervalo tiene carácter asintótico:

$$\frac{e_p}{\sqrt{\tilde{\sigma}^2_{MCGF}(\hat{w}_{pp} + X_p'(X'\hat{\Omega}^{-1}X)^{-1}X_p)}} \overset{a}{\sim} N(0,1)$$

de donde

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2} \sqrt{\tilde{\sigma}^2_{MCGF}(\hat{w}_{pp} + X_p'(X'\hat{\Omega}^{-1}X)^{-1}X_p)}\right]$$

- **Predicción por intervalo del valor esperado,** $E(Y_p)$

Error de predicción:

$$\epsilon_p \quad = \quad E(Y_p) - \widehat{E(Y_p)} = X_p'\beta - X_p'\tilde{\beta}_{MCG} = -X_p'(\hat{\beta}_{MCG} - \beta)$$

Distribución:

$$\begin{aligned}
E(\epsilon_p) &= E(-X_p'(\hat{\beta}_{MCG} - \beta)) = 0 \\
V(\epsilon_p) &= E[(\epsilon_p - E(\epsilon_p))^2] = E(\epsilon_p \epsilon_p') \\
&= E[(-X_p'(\tilde{\beta}_{MCG} - \beta))(-X_p'(\tilde{\beta}_{MCG} - \beta)] \\
&= X_p' E(\tilde{\beta}_{MCG} - \beta)(\tilde{\beta}_{MCG} - \beta)' X_p X \\
&= X_p' V(\tilde{\beta}_{MCG}) X_p \\
&= \sigma^2 X_p'(X'\Omega^{-1}X)^{-1}X_p
\end{aligned}$$

bajo normalidad de las perturbaciones, $u \sim N(0, \sigma^2\Omega)$:

$$\epsilon_p \sim N(0, X_p' V(\tilde{\beta}_{MCG}) X_p) \longrightarrow \epsilon_p \sim N(0, \sigma^2 X_p'(X'\Omega^{-1}X)^{-1}X_p)$$

de donde

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2}\sqrt{X_p' V(\tilde{\beta}_{MCG}) X_p}\right]$$

Se puede observar que la única diferencia con respecto a la predicción del valor Y_p es que el término $V(u_p)$ desaparece en la varianza del error de predicción. Por tanto los intervalos de confianza correspondientes son los siguientes.

a) Cuando $V(u) = \sigma^2\Omega$ conocida.

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2}\sqrt{\sigma^2 X_p'(X'\Omega^{-1}X)^{-1}X_p}\right]$$

b) Cuando $V(u) = \sigma^2\Omega$ con Ω conocida pero σ^2 desconocido.

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm t_{(N-K)\alpha/2}\sqrt{\tilde{\sigma}^2_{MCG} X_p'(X'\Omega^{-1}X)^{-1}X_p}\right]$$

c) Cuando $V(u) = \sigma^2\Omega$ con Ω y σ^2 desconocidos pero estimables.

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm N(0,1)_{\alpha/2}\sqrt{\tilde{\sigma}^2_{MCGF} X_p'(X'\hat{\Omega}^{-1}X)^{-1}X_p}\right]$$

3.7.1. Ejercicio

Sea el modelo:

$$Y_i = \beta X_i + u_i \qquad i = 1, 2, \ldots, N$$

donde $\sigma_i^2 = \sigma^2 X_i^2$.
Dado el valor de la variable explicativa en el periodo de predicción X_p, se quiere obtener una predicción por punto y por intervalo de Y_p.

Solución:

- Predicción por punto: $\hat{Y}_p = \tilde{\beta}_{MCG} X_p$ donde $\tilde{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$ o equivalentemente, el estimador lo obtenemos aplicando MCO en el modelo:

$$\frac{Y_i}{X_i} = \beta + \frac{u_i}{X_i} \longrightarrow Y_i^\star = \beta + u_i^\star$$

siendo:

$$\begin{array}{l} E(u_i^\star) = E\left(\frac{u_i}{X_i}\right) = 0 \qquad \forall i \\ V(u_i^\star) = Var\left(\frac{u_i}{X_i}\right) = \frac{\sigma^2 X_i^2}{X_i^2} = \sigma^2 \qquad \forall i \\ Cov(u_i^\star, u_\ell^\star) = Cov\left(\frac{u_i}{X_i}, \frac{u_\ell}{X_\ell}\right) = \frac{Cov(u_i, u_\ell)}{X_i X_\ell} = 0 \qquad \forall i \neq \ell \end{array}$$

En el modelo transformado:

$$\hat{\beta}_{MCG} = \overline{Y}^\star = \frac{\sum_1^N Y_i^\star}{N} = \frac{\sum_1^N \left(\frac{Y_i}{X_i}\right)}{N} = \beta + \frac{\sum_1^N u_i^\star}{N}$$

- Predicción por intervalo de Y_p:

$$\begin{aligned}
Y_p &= \beta X_p + u_p \\
\hat{Y}_p &= \tilde{\beta}_{MCG} X_p \\
e_p &= Y_p - \hat{Y}_p = -X_p(\tilde{\beta}_{MCG} - \beta) + u_p \\
E(e_p) &= -X_p(E(\tilde{\beta}_{MCG}) - \beta) + E(u_p) = 0
\end{aligned}$$

Buscamos las propiedades de $\tilde{\beta}_{MCG}$:

$$E(\tilde{\beta}_{MCG}) = E\left(\frac{\sum_1^N \left(\frac{Y_i}{X_i}\right)}{N}\right) = \frac{E(\sum_1^N Y_i^\star)}{N} = \frac{N\beta + \sum_1^N E(u_i^\star)}{N} = \beta$$

$$\begin{aligned}
V(\tilde{\beta}_{MCG}) &= E(\tilde{\beta}_{MCG} - E(\tilde{\beta}_{MCG}))^2 = E(\tilde{\beta}_{MCG} - \beta)^2 \\
&= E\left(\frac{\sum_1^N u_i^\star}{N}\right)^2 = \frac{\sum_1^N E(u_i^{\star 2})}{N^2} = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}
\end{aligned}$$

Así, si $u_i \sim N(0, \sigma^2 X_i^2)$ entonces $\tilde{\beta}_{MCG} \sim N(\beta, \frac{\sigma^2}{N})$
Varianza del error de predicción:

$$\begin{aligned}
V(e_p) &= E(e_p)^2 = E(-X_p(\tilde{\beta}_{MCG} - \beta) + u_p)^2 \\
&= E(u_p)^2 + E(X_p^2(\tilde{\beta}_{MCG} - \beta)^2) - 2E(X_p(\tilde{\beta}_{MCG} - \beta)u_p) \\
&= \sigma^2 X_p^2 + \sigma^2 \frac{X_p^2}{N} \\
&= \sigma^2 X_p^2 (1 + \frac{1}{N})
\end{aligned}$$

Como σ^2 es desconocida debemos estimarla:

$$\tilde{\sigma}^2_{MCG} = \frac{\tilde{u}'_{MCG} \Omega^{-1} \tilde{u}_{MCG}}{N-K}$$

Por tanto el intervalo de confianza se define como:

$$IC_{1-\alpha}(Y_p) = \left[\hat{Y}_p \pm t_{\frac{\alpha}{2}(N-K)} \tilde{\sigma}_{MCG} \sqrt{X_p^2 \left(1 + \frac{1}{N}\right)}\right]$$

A lo largo del tema se ha estudiado la heterocedasticidad más simple que podemos encontrarnos. Esta clase de heterocedasticidad, aunque pueda aparecer en series temporales, es muy frecuente en datos de sección cruzada. Sin embargo, en datos de series temporales, especialmente en datos financieros, nos encontramos con otro tipo de heterocedasticidad (ARCH, GARCH,...) que deben ser analizado de forma distinta.

Tema 4

Autocorrelación

4.1. Causas y modelización

En el modelo de regresión, el término de perturbación engloba aquellos factores que determinando la variable endógena, no están recogidos en la parte sistemática del modelo. Estos factores pueden ser innovaciones, errores de medida en la variable endógena, variables omitidas, etc.

Hasta el momento uno de los supuestos básicos del modelo de regresión lineal es que la covarianza entre perturbaciones de distintos periodos es cero. Sin embargo, si estos factores están correlacionados en el tiempo o en el espacio, entonces no se satisface la hipótesis de NO autocorrelación que escribíamos como $E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$. Este fenómeno se conoce con el nombre de autocorrelación: correlación serial, en el caso de series temporales y correlación espacial en el caso de datos de sección cruzada.

En los modelos que especifican relaciones en el tiempo entre variables, la propia inercia de las variables económicas, donde el impacto de una perturbación en un periodo de tiempo puede tener efectos en subsiguientes periodos, suele generar autocorrelación en el término de perturbación. Esta dinámica, aunque no sea relevante en media, refleja un patrón sistemático que tenemos de considerar a la hora de estimar el modelo.

Concepto de autocorrelación:
Existe autocorrelación cuando el término de error de un modelo econométrico está correlacionado consigo mismo. Es decir, la covarianza entre las perturbaciones es distinta de cero para diferentes momentos del tiempo (o entre distintos individuos):

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

No es preciso que u_t esté correlacionada consigo misma en cada dos instantes distintos del tiempo, sino que basta que la correlación se extienda a algunos periodos.

La presencia de autocorrelación implica que la matriz de varianzas y covarianzas, $E(uu') = \Omega$, tiene elementos distintos de cero fuera de la diagonal principal. En este contexto el estimador MCO es ineficiente y debemos estimar el modelo por MCG si Ω es conocida para obtener estimadores lineales, insesgados y de mínima varianza. Si Ω es desconocida estimamos el modelo por MCGF siempre y cuando podamos estimarla de forma consistente.

• **Estructura de la matriz de varianzas y covarianzas de la perturbación.**

En presencia de autocorrelación la matriz de varianzas y covarianzas tiene la forma siguiente:

$$E(uu') = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2T} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \sigma_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \sigma_{T3} & \dots & \sigma_{TT} \end{pmatrix}$$

Suponiendo que $\sigma_{11} = \sigma_{22} = \ldots = \sigma_{TT} = \sigma^2$, es decir, existe autocorrelación pero hay homocedasticidad:

$$E(uu') = \begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \sigma_{23} & \dots & \sigma_{2T} \\ \sigma_{31} & \sigma_{32} & \sigma^2 & \dots & \sigma_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \sigma_{T3} & \dots & \sigma^2 \end{pmatrix}$$

donde evidentemente se cumple que $cov(u_t, u_s) = cov(u_s, u_t)\ \forall t, s$. Por tanto, tenemos una matriz de varianzas y covarianzas que tiene $T(T-1)/2$ covarianzas más una varianza, σ_u^2.

4.1.1. Causas de autocorrelación

- **Shocks aleatorios prolongados**
  Sea el modelo:
  $$R_t = \beta_1 + \beta_2 RM_t + u_t \quad t = 1, 2, \ldots, T$$
  donde R_t es la rentabilidad de un activo en el periodo t y RM_t es la rentabilidad del mercado en dicho periodo t. Si en un momento dado se produce una caída del mercado, la rentabilidad del activo se verá afectada a la baja y como consecuencia la rentabilidad obtenida será menor que la esperada. Este efecto se prolongará en el tiempo hasta que poco a poco los inversores recuperen la confianza y el mercado vuelva a estabilizarse. El shock se recogerá en el término de perturbación. Si por ejemplo, la caída se produce en (t-1), lo que estamos diciendo es que la perturbación en t dependerá de lo ocurrido en (t-1) vía u_{t-1}.

- **Existencia de ciclos y tendencias**
  Si estamos analizando un modelo econométrico cuya variable endógena presenta ciclos y/o tendencias que no se explican a través de las variables exógenas, la perturbación recoge dichas estructuras, presentando un comportamiento de autocorrelación. En este caso, los residuos presentan rachas de desviaciones por encima del promedio (en la parte alta del ciclo) y rachas de desviaciones por debajo del promedio (parte baja del ciclo).

- **Relaciones no lineales**
  Supongamos que la verdadera relación entre los tipos de interés, r_t, y el stock de Deuda Pública, D_t, es cuadrática:
  $$r_t = \beta_1 + \beta_2 D_t + \beta_3 D_t^2 + u_t \quad t = 1, 2, \ldots, T \quad \beta_2 > 0, \beta_3 < 0$$
  Este modelo implica que los tipos de interés aumentan al crecer el stock de deuda pública, aunque menos que proporcionalmente, puesto que se tiene:
  $$\frac{\partial r_t}{\partial D_t} = \beta_2 + 2\beta_3 D_t < \beta_2$$

tanto menor cuanto mayor es D_t. Pero sin embargo se especifica y se estima un modelo lineal:

$$r_t = \beta_1 + \beta_2 D_t + u_t \quad t = 1, 2, \ldots, T$$

En este caso la curvatura de la parte sistemática pasa a ser recogida por la perturbación. Los residuos presentarán una racha de residuos negativos seguida de otra racha de residuos positivos para seguir con otra negativa.

- **Variables omitidas relevantes correlacionadas**
  Si el modelo realmente se especifica como:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad t = 1, 2, \ldots, T \\ u_t &= Y_t - \beta_1 - \beta_2 X_{2t} - \beta_3 X_{3t} \\ \hat{u}_t &= Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \hat{\beta}_3 X_{3t} \end{aligned}$$

  Pero estimamos:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + v_t \qquad t = 1, 2, \ldots, T \\ v_t &= Y_t - \beta_1 - \beta_2 X_{2t} = u_t + \beta_3 X_{3t} \\ \hat{v}_t &= Y_t - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2t} \\ &= \hat{u}_t - (\tilde{\beta}_1 - \hat{\beta}_1) - (\tilde{\beta}_2 - \hat{\beta}_2) X_{2t} + \hat{\beta}_3 X_{3t} \end{aligned}$$

  En este contexto de omisión los estimadores MCO son sesgados en general:
  $\tilde{\beta}_1$ y $\tilde{\beta}_2$ están sesgados salvo que $Cov(X_{2i}, X_{3i}) = 0$ y $\overline{X}_3 = 0$, además $(\tilde{\beta}_1 - \hat{\beta}_1) \neq 0$ y $(\tilde{\beta}_1 - \hat{\beta}_1) \neq 0$

  En consecuencia, tras un análisis de los residuos $\hat{v}_i$ tanto gráfico como mediante tests, es muy probable que el investigador llegue a la conclusión de que si la variable omitida está correlacionada, presenta ciclos o tendencias:

$$Cov(v_t, v_s) \neq 0$$

  De todas formas hay que tener en cuenta que no siempre que se omite una variable relevante se causa autocorrelación, podría solamente causar heterocedastidad.

- **Existencia de variables endógenas retardas**
  El comportamiento de muchas variables económicas en un periodo t depende no sólo de otras variables sino también de cuál fue el comportamiento de esa variable en el periodo anterior(t-1). En este sentido, a la hora de especificar el modelo econométrico debemos incluirla como regresor de lo contrario esa inercia lo recoge la perturbación. Esta situación se produce principalmente con observaciones mensuales o trimestrales. El modelo a especificar sería, por ejemplo:

$$C_t = \beta_1 + \beta_2 Y_t + \beta_3 C_{t-1} + u_t \qquad t = 1, 2, \ldots, T$$

  la existencia de C_{t-1} como variable explicativa provoca autocorrelación, ya que depende de u_{t-1} y está a su vez influye en u_t:

$$u_t = C_t - \beta_1 - \beta_2 Y_t - \beta_3 C_{t-1} \qquad \text{y} \qquad C_{t-1} = \beta_1 + \beta_2 Y_{t-1} + \beta_3 C_{t-2} + u_{t-1}$$

- **Datos manipulados**

- Datos en diferencias:
En muchas ocasiones en lugar de presentar datos originales, se presentan datos suavizados: diferencias, medias, etc. Si el modelo a especificar es

$$Y_t = \alpha + \beta X_t + u_t \qquad u_t \sim (0, \sigma^2) \tag{4.1}$$

pero los datos disponibles están en diferencias, $Z_t = Y_t - Y_{t-1}$, entonces el modelo resultante es:

$$Z_t = \beta(X_t - X_{t-1}) + e_t \qquad e_t = u_t - u_{t-1}$$

cuya perturbación tiene media cero, la varianza se duplica ($var(e_t) = 2\sigma^2$) y las covarianzas entre perturbaciones que distan en un periodo no es nulo ($cov(e_t, e_{t-1}) = -\sigma^2$).

- Datos como media de observaciones pasadas.
Supongamos que el dato que se proporciona corresponde a la media de los tres últimos meses $\left(Z_t = \frac{Y_{t-1}+Y_{t-2}+Y_{t-3}}{3}\right)$, entonces el modelo resultante es:

$$Z_t = \alpha + \beta \frac{X_{t-1} + X_{t-2} + X_{t-3}}{3} + e_t \qquad e_t = \frac{u_{t-1} + u_{t-2} + u_{t-3}}{3}$$

cuya perturbación tiene media cero, la varianza es ($var(e_t) = \sigma^2/3$) y las covarianzas entre perturbaciones que distan entre un periodo y dos no son nulas:

$$cov(e_t, e_s) = \begin{cases} 2\sigma^2/9 & \text{si} \quad |t-s| = 1 \\ \sigma^2/9 & \text{si} \quad |t-s| = 2 \\ 0 & \text{si} \quad |t-s| > 2 \end{cases}$$

- Tasas de crecimiento:
Generalmente las variables económicas, que evolucionan continuamente, están medidas a intervalos de tiempo regulares (por ejemplo trimestrales) y expresadas como tasas de crecimiento sobre el periodo anterior. La utilización de datos económicos medidos en tasas de crecimiento genera autocorrelación en las perturbaciones. Por ejemplo, supongamos la variable Y_t, definimos su tasa de crecimiento como:

$$\frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

donde si $Y_t = f(X_t) + u_t$ el numerador de la expresión anterior depende de $(u_t - u_{t-1})$ no independiente de lo que ocurre en t-1, donde $(Y_{t-1} - Y_{t-2})$ dependen de $(u_{t-1} - u_{t-2})$, ambas dependen de u_{t-1}.

4.1.2. Modelización de la autocorrelación

Sea el MRLG $Y = X\beta + u$ donde:

$$E(u_t^2) = \sigma^2 \quad \forall t; \quad E(u_t u_s) \neq 0 \quad \forall t, s \quad t \neq s$$

$$E(uu') = \begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma^2 & \sigma_{23} & \dots & \sigma_{2T} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \sigma_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \sigma_{T3} & \dots & \sigma^2 \end{pmatrix}.$$

En presencia de autocorrelación tenemos que estimar σ_u^2, $\frac{T(T-1)}{2}$ covarianzas y K coeficientes con sólo T observaciones. Evidentemente, esta estimación no es factible. Para simplificar el número de parámetros a estimar tenemos que hacer supuestos sobre la estructura de autocorrelación, de forma que dependa de un conjunto de parámetros menor. Habitualmente el supuesto que se establece es el de estacionariedad débil, que entre otras cosas, implica que las covarianzas entre dos perturbaciones no dependan de los tiempos a los que pertenecen, sino del retardo de tiempo que hay entre ellas. Así la covarianza de orden "s" sería:

$$Cov(u_t, u_{t-s}) = E(u_t u_{t-s}) = \gamma_s \qquad s \pm 1, \pm 2, \pm 3 \ldots$$

donde s indica el retardo. De esta forma el número de covarianzas a estimar se reduce porque todas aquellas covarianzas que tengan el mismo retardo "s" son iguales:

$$E(u_t u_{t-1}) = E(u_{t-1} u_{t-2}) = E(u_{t-2} u_{t-3}) = \ldots = \gamma_1$$

$$E(u_t u_{t-2}) = E(u_{t-1} u_{t-3}) = E(u_{t-2} u_{t-4}) = \ldots = \gamma_2$$

Así, la matriz de varianzas y covarianzas de la perturbación se simplifica:

$$E(uu') = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-3} \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \gamma_{T-4} & \cdots & \gamma_0 \end{pmatrix}$$

donde $\gamma_0 = E(u_t u_t) = \sigma_u^2$ denota la varianza (retardo cero, $s = 0$).

Esta matriz de varianzas y covarianzas se puede expresar en función de los coeficientes de correlación, para lo cual definimos el coeficiente de correlación entre u_t y u_{t-s} como:

$$\rho_s = \frac{Cov(u_t, u_{t-s})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-s})}} = \frac{\gamma_s}{\sqrt{\sigma_u^2}\sqrt{\sigma_u^2}} = \frac{\gamma_s}{\sigma_u^2} = \frac{\gamma_s}{\gamma_0}$$

de donde podemos deducir que $\gamma_s = \rho_s \sigma_u^2 = \rho_s \gamma_0$. De esta forma la matriz de varianzas y covarianzas de la perturbación se puede escribir como:

$$E(uu') = \sigma_u^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-2} \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \rho_{T-3} \\ \rho_3 & \rho_2 & \rho_1 & 1 & \cdots & \rho_{T-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \rho_{T-4} & \cdots & 1 \end{pmatrix}$$

Bajo el supuesto de estacionariedad el número de parámetros desconocidos en la matriz de varianzas y covarianzas es T: $(\sigma_u^2, \rho_1, \rho_2, \ldots, \rho_{T-1})$. Pero todavía no es suficiente porque aún tenemos K coeficientes, σ_u^2 y $(T-1)$ covarianzas a estimar con las T observaciones muestrales disponibles. Tenemos que suponer algo adicional con objeto de reducir el número de parámetros desconocidos a estimar.

A continuación se describen algunas de las estructuras más empleadas para especificar la autocorrelación.

Proceso autorregresivo de primer orden, AR(1)

Es el proceso de autocorrelación más sencillo y uno de los que mejor se suele ajustar a datos económicos. Se especifica como:

$$u_t = \rho u_{t-1} + \epsilon_t \qquad t = 1, 2, \ldots, T$$

de forma que la perturbación en el periodo t depende de la perturbación del periodo anterior $(t-1)$ más un término aleatorio (o innovación) ϵ_t cuyas características son:

$$\begin{array}{ll} E(\epsilon_t) = 0 \quad \forall t & \\ E(\epsilon_t^2) = \sigma_\epsilon^2 \quad \forall t & \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \\ E(\epsilon_t \epsilon_s) = 0 \quad \forall t, s \quad t \neq s & \end{array}$$

y que habitualmente se le llama ruido blanco. La relación entre la perturbación y la innovación se recoge en el diagrama siguiente:

$$\begin{array}{ccccccc} u_{t-2} & \longrightarrow & u_{t-1} & \longrightarrow & u_t & \longrightarrow & u_{t+1} \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ \epsilon_{t-2} & & \epsilon_{t-1} & & \epsilon_t & & \epsilon_{t+1} \end{array}$$

por lo que cada innovación influye sobre la perturbación en el mismo periodo o períodos posteriores, pero nunca sobre los valores anteriores, es decir:

$$E(\epsilon_t u_{t-s}) = 0 \quad s > 0.$$

donde ρ mide la correlación entre u_t y u_{t-1}:

$$\rho = \frac{Cov(u_t, u_{t-1})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-1})}} = \frac{\gamma_1}{\sigma_u^2} \qquad |\rho| < 1$$

En un proceso AR(1) la condición (necesaria y suficiente) de estacionariedad es que $|\rho| < 1$. Dado que $u_t = Y_t - E(Y_t/\{X_{it}\}_{i=1}^K)$ la perturbación representa la diferencia entre el comportamiento observado y el comportamiento promedio, tenemos que:

i) Si $\rho > 0$ entonces un valor elevado de u_t genera un valor de Y_t por encima del promedio y tendrá mayor probabilidad de ir seguido por un valor elevado de u_{t+1} y así sucesivamente.

ii) Si $\rho < 0$ un valor alto de u_t irá seguido por un valor bajo de u_{t+1} y éste por uno alto de u_{t+2} y así sucesivamente.

En la Figura 4.1 se observa un proceso autorregresivo de primer orden con parámetro ρ positivo. En ella podemos una racha de residuos positivos seguidos de una racha de residuos negativos y así sucesivamente. En cambio, cuando el parámetro del proceso autorregresivo es negativo, los signos de los residuos se alternan como podemos ver en la Figura 4.2.

• Para conocer **la estructura de la matriz de varianzas y covarianzas de la perturbación bajo un AR(1)** tenemos que hallar los primeros momentos de la perturbación. Comenzamos por obtener una expresión más compacta del proceso:
Si $t = 1 \quad u_1 = \rho u_0 + \epsilon_1 = \epsilon_1 + \rho u_0$

Gráfico 4.1: Perturbaciones AR(1) positivo

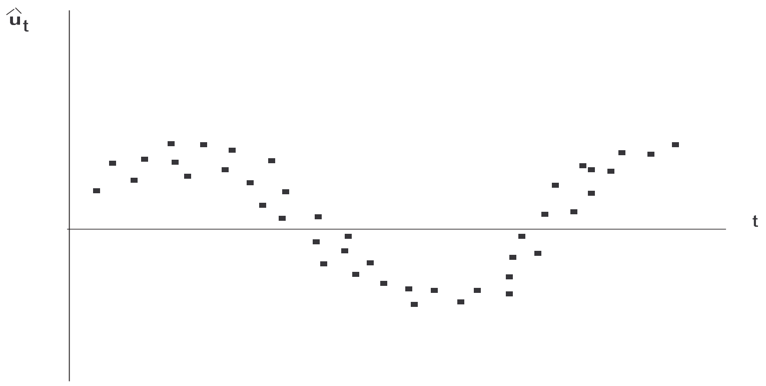
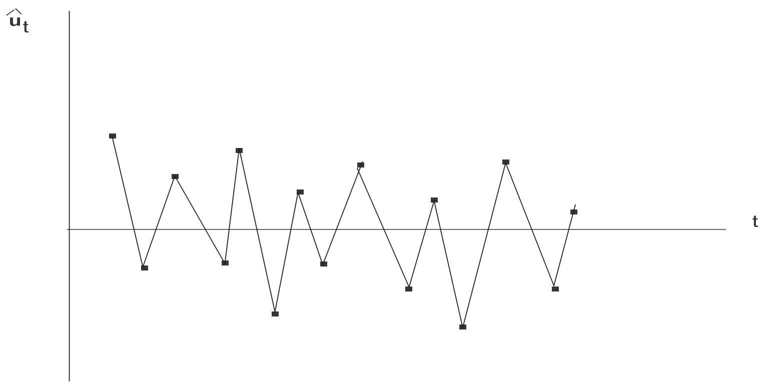


Gráfico 4.2: Perturbaciones AR(1) negativo



Si $t = 2 \quad u_2 = \rho u_1 + \epsilon_2 = \epsilon_2 + \rho(\epsilon_1 + \rho u_0) = \epsilon_2 + \rho\epsilon_1 + \rho^2 u_0$

Si $t = 3 \quad u_3 = \rho u_2 + \epsilon_3 = \epsilon_3 + \rho(\epsilon_2 + \rho\epsilon_1 + \rho^2 u_0) = \epsilon_3 + \rho\epsilon_2 + \rho^2\epsilon_1 + \rho^3 u_0$

$$\vdots$$

Si $t = T \quad u_T = \rho u_{T-1} + \epsilon_T = \epsilon_T + \rho\epsilon_{T-1} + \rho^2\epsilon_{T-2} + \ldots + \rho^{T-1}\epsilon_1 + \rho^T u_0$
entonces de forma general:

$u_t = \rho u_{t-1} + \epsilon_t = \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + \ldots + \rho^{t-1}\epsilon_1 + \rho^t u_0.$

Suponiendo que el proceso comienza en un punto remoto, $\rho^T u_0$ tomará un valor despreciable, de forma general podemos escribir:

$$u_t = \sum_{i=0}^{\infty} \rho^i \epsilon_{t-i} \longrightarrow u_t = f(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots) \tag{4.2}$$

es decir, la perturbación u_t es una combinación lineal de las innovaciones pasadas ϵ_t, ponderadas por $1, \rho, \rho^2 \ldots$ que decaen geométricamente ya que $|\rho| < 1$. Esto implica que las innovaciones ϵ_{t-i} tienen menor influencia en u_t cuanto más alejadas están en el tiempo.

A continuación, basándonos en la expresión (4.2) hallamos la media, la varianza y las covarianzas del proceso:

$$\begin{array}{rcl} E(u_t) & = & E(\sum_{i=0}^{\infty} \rho^i \epsilon_{t-i}) = \sum_{i=0}^{\infty} \rho^i E(\epsilon_{t-i}) = 0 \quad \forall t \\ Var(u_t) & = & \sigma_u^2 = \gamma_0 = E(u_t - E(u_t))^2 = E(u_t^2) = \\ & = & E(\rho u_{t-1} + \epsilon_t)^2 = \rho^2 E(u_{t-1}^2) + E(\epsilon_t^2) + 2\rho E(u_{t-1}\epsilon_t) = \\ & = & \rho^2 \sigma_u^2 + \sigma_\epsilon^2 \end{array}$$

de donde obtenemos que

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_\epsilon^2 \quad \Rightarrow \quad \sigma_u^2(1-\rho^2) = \sigma_\epsilon^2 \quad \Rightarrow \quad \sigma_u^2 = \frac{\sigma_\epsilon^2}{(1-\rho^2)} = \gamma_0$$

En cuanto a las covarianzas tenemos:

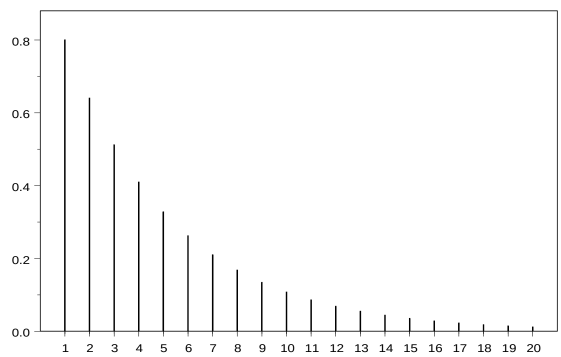
$$\begin{array}{l} Cov(u_t, u_{t-1}) = E(u_t u_{t-1}) = E((\rho u_{t-1} + \epsilon_t) u_{t-1}) = \rho E(u_{t-1}^2) + E(u_{t-1}\epsilon_t) = \rho \sigma_u^2 = \gamma_1 \\ Cov(u_t, u_{t-2}) = E(u_t u_{t-2}) = E((\rho u_{t-1} + \epsilon_t) u_{t-2}) = \rho E(u_{t-1} u_{t-2}) + E(u_{t-2}\epsilon_t) = \rho \gamma_1 = \rho^2 \sigma_u^2 = \gamma_2 \\ Cov(u_t, u_{t-3}) = E(u_t u_{t-3}) = E((\rho u_{t-1} + \epsilon_t) u_{t-3}) = \rho E(u_{t-1} u_{t-3}) + E(u_{t-3}\epsilon_t) = \rho \gamma_2 = \rho^3 \sigma_u^2 = \gamma_3 \\ \vdots \\ Cov(u_t, u_{t-s}) = E(u_t u_{t-s}) = \rho^s \sigma_u^2 = \gamma_s \end{array}$$

Si obtenemos la correlación entre dos perturbaciones que distan s retardos, tenemos

$$Cor(u_t, u_{t-s}) = \frac{Cov(u_t, u_{t-s})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-s})}} = \frac{\gamma_s}{\sigma_u^2} = \frac{\rho^s \gamma_0}{\gamma_0} = \rho^s$$

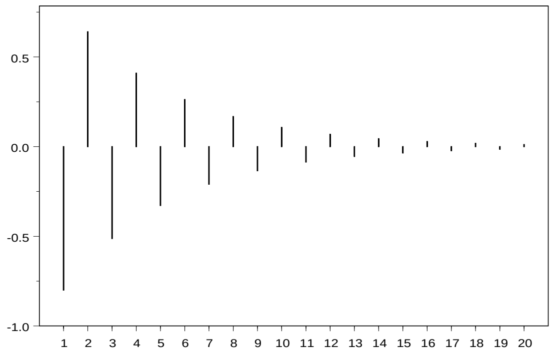
por lo que ρ mide la correlación entre u_t y u_{t-1}. Así, si dibujamos la FAC (Función de Auto-Correlación) de un AR(1) positivo, vemos que los valores decaen exponencialmente (Figura 4.3) hacia cero y si el proceso AR(1) es negativo los signos de estos valores se alternan (Figura 4.4).

Gráfico 4.3: FAC: AR(1) positivo



La matriz de varianzas y covarianzas de la perturbación cuando ésta sigue un proceso autorregresivo de orden uno es:

Gráfico 4.4: FAC: AR(1) negativo



$$E(uu') = \sigma_u^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & \dots & 1 \end{pmatrix} = \sigma_u^2 \Omega.$$

La estimación de la matriz de varianzas y covarianzas se ha simplificado de tal modo que sólo tenemos que estimar dos parámetros: σ_u^2 y ρ. En realidad, dado que $\sigma_u^2 = \frac{\sigma_\epsilon^2}{(1-\rho^2)}$, los parámetros desconocidos son ρ y σ_ϵ^2. Conocido ρ, Ω queda totalmente determinada, a excepción del factor de escala que depende de σ_ϵ^2.

Proceso autorregresivo de cuarto orden, AR(4)

Un proceso autorregresivo frecuentemente utilizado para recoger efectos estacionales en la perturbación con datos trimestrales es el llamado proceso autorregresivo de cuarto orden, AR(4). Su especificación es la siguiente:

$$u_t = \rho u_{t-4} + \epsilon_t \qquad |\rho| < 1 \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2)$$

donde se supone que u_t depende de la perturbación de cuatro periodos atrás más una innovación con propiedades esféricas. Si estamos utilizando datos de series temporales este proceso autorregresivo recoge el hecho de que cada periodo (trimestre) está relacionado con el mismo trimestre del año anterior. Podríamos haberlo escrito como:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \epsilon_t$$

con $\rho_1 = \rho_2 = \rho_3 = 0$. La expresión compacta para este caso es

$$\begin{aligned} u_t &= \rho u_{t-4} + \epsilon_t = \\ &= \rho(\rho u_{t-8} + \epsilon_{t-4}) + \epsilon_t = \\ &= \ldots = \\ &= \epsilon_t + \rho\epsilon_{t-4} + \rho^2\epsilon_{t-8} + \ldots = \\ &= \sum_{i=0}^{\infty} \rho^j \epsilon_{t-4j} \end{aligned}$$

$$\begin{array}{ll} E(u_t) & = E(\sum_{i=0}^{\infty} \rho^j \epsilon_{t-4j}) = \sum_{i=0}^{\infty} \rho^j E(\epsilon_{t-4j}) = 0 \quad \forall t \\ Var(u_t) & = \sigma_u^2 = \gamma_0 = \\ & = E(u_t - E(u_t))^2 = E(u_t^2) = E(\rho u_{t-4} + \epsilon_t)^2 = \\ & = \rho^2 E(u_{t-4}^2) + E(\epsilon_t^2) + 2\rho E(u_{t-4}\epsilon_t) = \\ & = \rho^2 \sigma_u^2 + \sigma_\epsilon^2 \end{array}$$

de donde

$$\sigma_u^2 = \rho^2 \sigma_u^2 + \sigma_\epsilon^2 \quad \Rightarrow \quad \sigma_u^2(1-\rho^2) = \sigma_\epsilon^2 \quad \Rightarrow \quad \sigma_u^2 = \frac{\sigma_\epsilon^2}{(1-\rho^2)} = \gamma_0.$$

En cuanto a las covarianzas:

$$\begin{array}{l} Cov(u_t, u_{t-1}) = E(u_t u_{t-1}) = E((\rho u_{t-4} + \epsilon_t) u_{t-1}) = 0 \\ Cov(u_t, u_{t-2}) = E(u_t u_{t-2}) = E((\rho u_{t-4} + \epsilon_t) u_{t-2}) = 0 \\ Cov(u_t, u_{t-3}) = E(u_t u_{t-3}) = E((\rho u_{t-4} + \epsilon_t) u_{t-3}) = 0 \\ Cov(u_t, u_{t-4}) = E(u_t u_{t-4}) = E((\rho u_{t-4} + \epsilon_t) u_{t-4}) = \rho E(u_{t-4}^2) + E(u_{t-4}\epsilon_t) = \rho \sigma_u^2 = \gamma_4 \\ Cov(u_t, u_{t-8}) = E(u_t u_{t-8}) = E((\rho u_{t-4} + \epsilon_t) u_{t-8}) = \rho E(u_{t-4} u_{t-8}) + E(u_{t-8}\epsilon_t) = \rho \gamma_4 = \rho^2 \sigma_u^2 = \gamma_8 \end{array}$$

Por tanto la matriz de varianzas y covarianzas de la perturbación cuando ésta sigue un proceso autorregresivo de cuarto orden es:

$$E(uu') = \sigma_u^2 \begin{pmatrix} 1 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & \rho^2 & 0 & \dots & \dots & \dots \\ 0 & 1 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & \rho^2 & \dots & \dots & \dots \\ 0 & 0 & 1 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & \rho^2 & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 \end{pmatrix} = \sigma_u^2 \Omega$$

en la que los únicos parámetros a estimar son σ_u^2 y ρ, en realidad σ_ϵ^2 y ρ.

Proceso de medias móviles de orden uno, MA(1)

El proceso de medias móviles más sencillo es el MA(1) que se define como:

$$u_t = \epsilon_t + \theta \epsilon_{t-1} \qquad t = 1, 2, \dots T$$

donde $|\theta| < 1$ para que el proceso sea estacionario. En este caso, la perturbación u_t es una combinación lineal de sólo dos innovaciones ϵ_t y ϵ_{t-1}, por lo que se dice que es un proceso de memoria corta. Buscamos las propiedades de u_t:

$$\begin{array}{lcl} E(u_t) & = & E(\epsilon_t + \theta\epsilon_{t-1}) = E(\epsilon_t) + \theta E(\epsilon_{t-1}) = 0 \\ Var(u_t) & = & E(u_t^2) = E(\epsilon_t + \theta\epsilon_{t-1})^2 = E(\epsilon_t^2) + \theta^2 E(\epsilon_{t-1}^2) + 2\theta E(\epsilon_t \epsilon_{t-1}) = \\ & = & \sigma_\epsilon^2 + \theta^2 \sigma_\epsilon^2 = \sigma_\epsilon^2 (1+\theta^2) \\ Cov(u_t, u_{t-1}) & = & E(u_t u_{t-1}) = E((\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})) = \\ & = & E(\epsilon_t\epsilon_{t-1}) + \theta E(\epsilon_t\epsilon_{t-2}) + \theta E(\epsilon_{t-1}^2) + \theta^2 E(\epsilon_{t-1}\epsilon_{t-2}) = 0 + 0 + \theta\sigma_\epsilon^2 + \theta^2 0 = \theta\sigma_\epsilon^2 \\ Cov(u_t, u_{t-2}) & = & E(u_t u_{t-2}) = E((\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-2} + \theta\epsilon_{t-3})) = \\ & = & E(\epsilon_t\epsilon_{t-2}) + \theta E(\epsilon_t\epsilon_{t-3}) + \theta E(\epsilon_{t-1}\epsilon_{t-2}) + \theta^2 E(\epsilon_{t-1}\epsilon_{t-3}) = 0 \\ Cov(u_t, u_{t-s}) & = & 0 \quad \forall s \geq 2 \quad \forall t \end{array}$$

de donde:

$$E(uu') = \sigma^2_\epsilon \begin{pmatrix} (1+\theta^2) & \theta & 0 & 0 & \dots & 0 \\ \theta & (1+\theta^2) & \theta & 0 & \dots & 0 \\ 0 & \theta & (1+\theta^2) & \theta & \dots & 0 \\ \vdots & 0 & \theta & (1+\theta^2) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \theta \\ 0 & 0 & 0 & \dots & \theta & (1+\theta^2) \end{pmatrix} = \sigma^2_\epsilon \Omega$$

la matriz de varianzas y covarianzas de la perturbación u queda totalmente determinada una vez conocida θ, salvo por el factor de escala σ^2_ϵ. En el caso de un proceso MA(1) el número de correlaciones se reduce a uno ya que:

$$\begin{aligned} Cor(u_t, u_{t-1}) &= \frac{Cov(u_t, u_{t-1})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-1})}} = \frac{\theta\sigma^2_\epsilon}{\sigma^2_u} = \theta \\ Cor(u_t, u_{t-2}) &= \frac{Cov(u_t, u_{t-2})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-2})}} = \frac{0}{\sigma^2_u} = 0 \\ &\vdots \\ Cor(u_t, u_{t-s}) &= \frac{Cov(u_t, u_{t-s})}{\sqrt{Var(u_t)}\sqrt{Var(u_{t-s})}} = \frac{0}{\sigma^2_u} = 0 \quad s > 1 \end{aligned}$$

por lo que θ mide la correlación entre u_t y u_{t-1}. Así, la FAC correspondiente a un MA(1) tiene un único valor distinto de cero: el parámetro θ. Como se puede observar en las Figuras 4.6 y 4.5, si el proceso MA(1) es positivo la correlación será positiva y viceversa.

En algunas ocasiones se define el proceso MA(1) como $u_t = \epsilon_t - \phi\epsilon_{t-1}$ donde reparametrizamos $\theta = -\phi$ y estamos en los mismos resultados que los obtenidos anteriormente:

$$E(uu') = \sigma^2_\epsilon \begin{pmatrix} (1+\phi^2) & -\phi & 0 & 0 & \dots & 0 \\ -\phi & (1+\phi^2) & -\phi & 0 & \dots & 0 \\ 0 & -\phi & (1+\phi^2) & -\phi & \dots & 0 \\ \vdots & 0 & -\phi & (1+\phi^2) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & -\phi \\ 0 & 0 & 0 & \dots & -\phi & (1+\phi^2) \end{pmatrix} = \sigma^2_\epsilon \Omega$$

Otros procesos más generales serían por ejemplo:

$$\begin{array}{ll} AR(2) & u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t \\ AR(p) & u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \epsilon_t \\ MA(2) & u_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \\ MA(q) & u_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q} \end{array}$$

El modelo más general es un ARMA(p,q) donde u_t depende de sus valores pasados, de la innovación ϵ_t y de los valores pasados de ésta. El modelo se especifica:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_p u_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}$$

en el que se combina un proceso AR(p) con un proceso MA(q).

Gráfico 4.5: FAC: MA(1) positivo

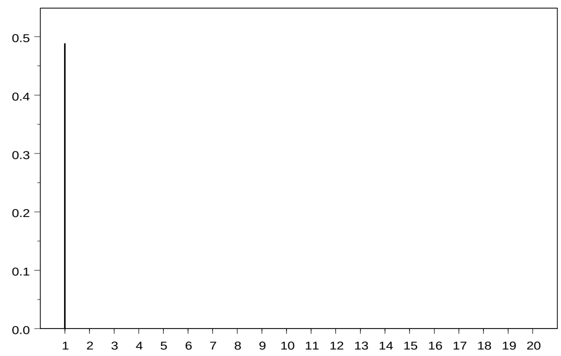
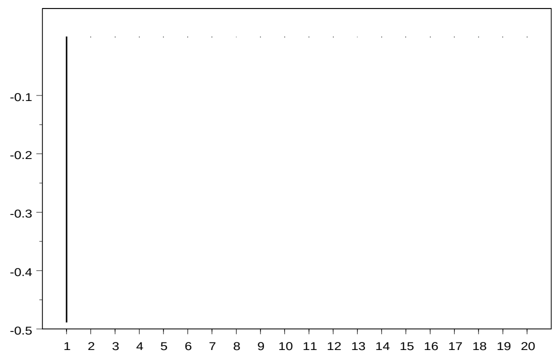


Gráfico 4.6: FAC: MA(1) negativo



4.2. Contrastes de autocorrelación

En la práctica no se conoce a priori si existe autocorrelación ni cuál es el proceso más adecuado para modelarla. Para determinar su existencia es necesario contrastar dicha hipótesis mediante un estadístico de contraste.

Si embargo, ningún contraste de autocorrelación debe excluir un examen riguroso de los residuos generados en la estimación del modelo. El gráfico de los mismos puede indicarnos la existencia de autocorrelación. Dado que los residuos son una aproximación a la perturbación, la existencia de patrones o comportamientos sistemáticos indicaría la posible existencia de autocorrelación.

Por ejemplo, si en el gráfico de la evolución temporal de $\hat{u}_t$ contra la de $\hat{u}_{t-s}$ para $s = 1$ encontramos que la mayoría de los puntos en dicho gráfico se hallan en el primer o tercer cuadrante, Figura 4.7, ello es un indicio de autocorrelación positiva. Si se hallan en el segundo y cuarto cuadrante indicará autocorrelación negativa.

Gráfico 4.7: Perturbaciones AR(1) positivo

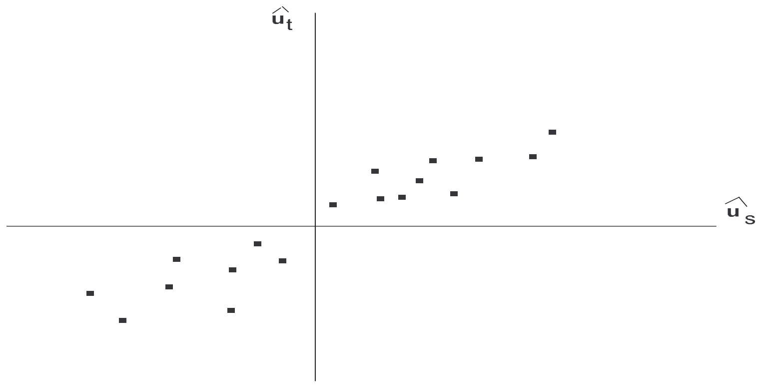
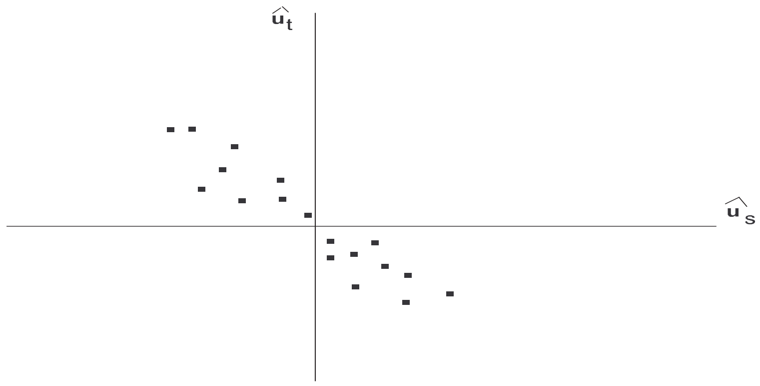


Gráfico 4.8: Perturbaciones AR(1) negativo

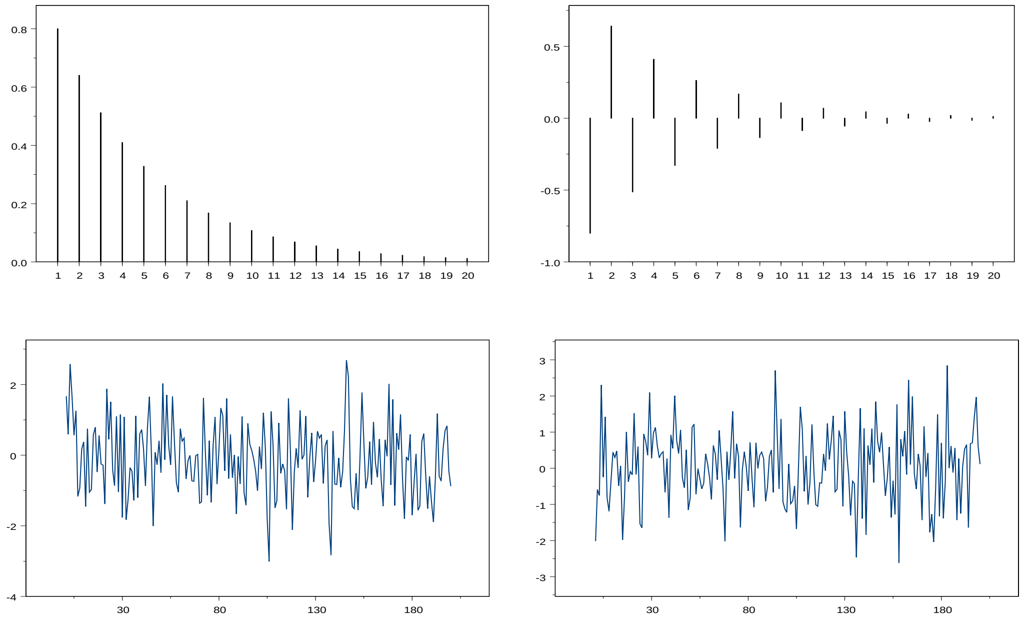


Cuando los residuos no se comportan aleatoriamente, si no que recogen una estructura, ello puede indicar la existencia de autocorrelación. Tras el análisis gráfico, debemos contrastarla. Si el resultado del contraste es que existe autocorrelación y ésta no es debida a una mala especificación del modelo el método se estima por MCG o MCGF.

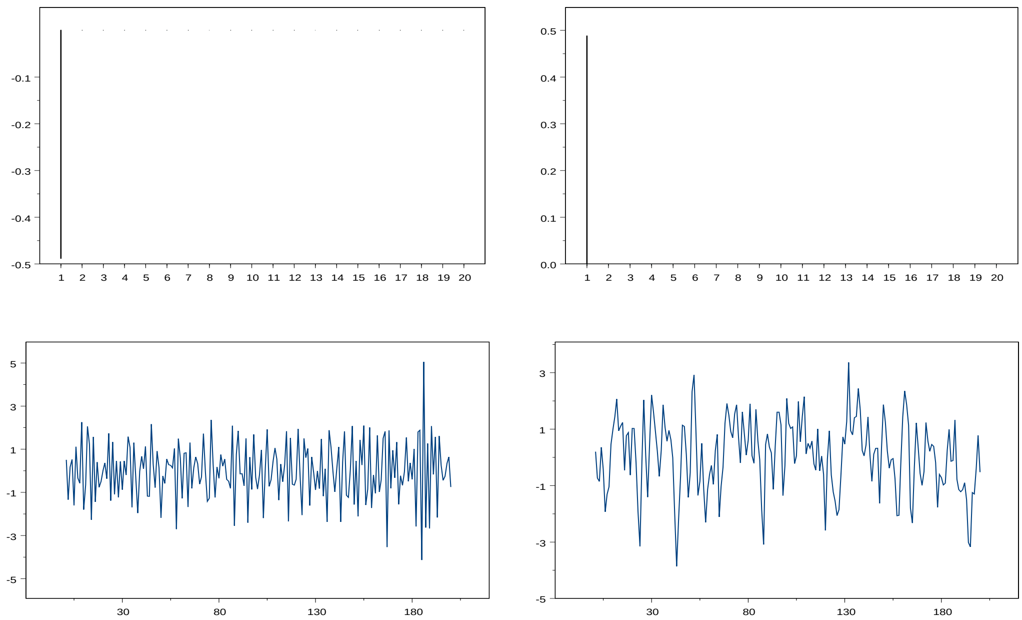
Si la autocorrelación es originada por una mala especificación del modelo primero se ha de corregir esta especificación y una vez el modelo esté correctamente especificado analizar las propiedades de la perturbación.

No obstante, a menudo y especialmente ante tamaños de muestra considerables, el análisis gráfico puede no sugerir nada en especial. En las siguientes figuras se muestra un proceso AR(1) positivo y negativo junto con su Función de Autocorrelación (FAC) y también un proceso MA(1) positivo y negativo. En todas ellas resulta imposible divisar un comportamiento sistemático, por lo que se hace necesario emplear un contraste.

Gráfico 4.9: Funciones de autocorrelación y realizaciones de modelos



$Y_t = 0{,}8Y_{t-1} + u_t$ $\qquad$ $Y_t = -0{,}8Y_{t-1} + u_t$



$Y_t = \epsilon_t - 0{,}8\epsilon_{t-1}$ $\qquad$ $Y_t = \epsilon_t + 0{,}8\epsilon_{t-1}$

4.2.1. Contraste de Durbin Watson

Durbin y Watson propusieron en 1951 un estadístico para contrastar la existencia de un proceso AR(1) en el término de perturbación. La hipótesis nula es la no existencia de autocorrelación:

$$H_0 : \rho = 0$$

frente a la alternativa

$$H_a : \rho \neq 0 \quad \text{en} \quad u_t = \rho u_{t-1} + \epsilon_t \quad \epsilon_t \sim (0, \sigma^2_\epsilon)$$

y se contrasta mediante el estadístico:

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^T \hat{u}_t^2} = \frac{\sum_{t=2}^T \hat{u}_t^2 + \sum_{t=2}^T \hat{u}_{t-1}^2 - 2\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2}$$

donde $\hat{u}_t$ son los residuos mínimo-cuadráticos ordinarios de estimar el modelo original sin tener en cuenta la existencia de autocorrelación en las perturbaciones.

Si el tamaño muestral es suficientemente grande podemos emplear las aproximaciones

$$\sum_{t=2}^T \hat{u}_t^2 \simeq \sum_{t=2}^T \hat{u}_{t-1}^2 \simeq \sum_{t=1}^T \hat{u}_t^2$$

con lo que

$$DW \simeq \frac{2\sum_{t=2}^T \hat{u}_t^2 - 2\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^T \hat{u}_t^2} \simeq 2 - 2\frac{\sum_{t=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^T \hat{u}_{t-1}^2} \simeq 2(1-\hat{\rho})$$

donde $\hat{\rho}$ es el estimador de ρ por MCO en el modelo $u_t = \rho u_{t-1} + \epsilon_t$, empleando como proxy de u_t el residuo MCO.

- Interpretación del estadístico DW:
  Si existe autocorrelación positiva de primer orden, valores positivos del término de error u_t tiendan a ir seguidos de valores positivos y asimismo, valores negativos tiendan a ir seguidos de valores negativos. Dado que la aproximación a la perturbación es el residuo, los patrones en la perturbación serán detectados en el residuo. Así, observaremos rachas de residuos positivos seguidas de rachas de residuos negativos. En estas circunstancias, generalmente $|\hat{u}_t - \hat{u}_{t-1}| < |\hat{u}_t| \Rightarrow (\hat{u}_t - \hat{u}_{t-1})^2 < \hat{u}_t^2$ y el numerador del estadístico "será pequeño" en relación al denominador, con lo que el estadístico "será pequeño". En consecuencia cuanto más cercano esté el parámetro ρ a la unidad más próximo a cero estará el DW. En el extremo positivo tenemos que $\rho \longrightarrow 1 \quad \Rightarrow DW \longrightarrow 0$.

  Si existe autocorrelación negativa de primer orden, valores positivos de $\hat{u}_t$ tienden a ir seguidos de valores negativos, en este caso $|\hat{u}_t - \hat{u}_{t-1}| > |\hat{u}_t| \Rightarrow (\hat{u}_t - \hat{u}_{t-1})^2 > \hat{u}_t^2$ con lo que el estadístico DW tenderá a tomar valores grandes. En el extremo negativo tenemos que $\rho \longrightarrow -1 \quad \Rightarrow DW \longrightarrow 4$.

  A partir de la relación $DW \simeq 2(1-\hat{\rho})$ se puede establecer el rango de valores que puede tomar el estadístico DW.

$$\begin{array}{ll} 0 < \hat{\rho} < 1 & DW \in (0,2) \\ \hat{\rho} = 0 & DW \simeq 2 \\ -1 < \hat{\rho} < 0 & DW \in (2,4) \end{array}$$

  la distribución del estadístico DW bajo H_0 depende de la matriz de regresores X ya que $\hat{u} = Mu = (I - X(X'X)^{-1}X')u$ por lo que los valores críticos del contraste también serán diferentes para cada posible X. Durbin y Watson tabularon los valores máximo (d_U) y mínimo (d_L) que puede tomar el estadístico independientemente de cuál sea X, y tal que $d_L < DW < d_U$. La distribución de d_L y d_U depende del tamaño de la muestra, T, y de K' que denota el número de variables explicativas del modelo exceptuando el término independiente.

  Contraste de existencia de autocorrelación positiva:
  $H_0 : \rho = 0$

$H_a : \rho > 0 \quad \text{en} \quad u_t = \rho u_{t-1} + \epsilon_t \quad |\rho| < 1 \quad \epsilon_t \sim iid(0, \sigma^2_\epsilon)$

a) Si $DW < d_L$ se rechaza la H_0 para un nivel de significatividad α dado, por tanto existe autocorrelación positiva.

b) Si $DW > d_U$ no se rechaza la H_0 para un nivel de significatividad α dado, por tanto no existe autocorrelación positiva.

c) Si $d_L < DW < d_U$ estamos en una zona de incertidumbre y no podemos concluir si existe o no autocorrelación positiva de primer orden.

Contraste de existencia de autocorrelación negativa:
$H_0 : \rho = 0$
$H_a : \rho < 0 \quad \text{en} \quad u_t = \rho u_{t-1} + \epsilon_t \quad |\rho| < 1 \quad \epsilon_t \sim iid(0, \sigma^2_\epsilon)$

a) Si $DW < 4 - d_U$ no se rechaza la H_0 para un nivel de significatividad α dado, por tanto no existe autocorrelación negativa.

b) Si $DW > 4 - d_L$ se rechaza la H_0 para un nivel de significatividad α dado, por tanto existe autocorrelación negativa.

c) Si $4 - d_U < DW < 4 - d_L$ estamos en una zona de incertidumbre y como en el caso anterior, no podemos concluir si existe o no autocorrelación negativa de primer orden.

Gráficamente:

$$H_o : \rho = 0$$

Autocorrelación positiva — Autocorrelación negativa

$\longleftarrow - - - - H_a : \rho > 0 - - - - - \longrightarrow \quad \longleftarrow - - - - - - - - H_a : \rho < 0 - - - - - \longrightarrow$

Rechazar $\rho = 0$	Duda	Aceptar $\rho = 0$	Duda	Rechazar $\rho = 0$
0 — d_L	d_L — d_U	d_U — 2 — $4 - d_U$	$4 - d_U$ — $4 - d_L$	$4 - d_L$ — 4

El contraste de Durbin Watson también se puede considerar un contraste de mala especificación del modelo. La omisión de variables relevantes correlacionadas, una forma funcional inadecuada, cambios estructurales no incluidos en el modelo, etc., pueden originar un estadístico DW significativo. Esto nos puede llevar a errores si consideramos que hay evidencia de autocorrelación y se modela un proceso AR(1). Por otro lado, si u_t sigue un proceso distinto de un AR(1), supongamos un AR(2), es probable que el estadístico DW lo detecte. Por lo tanto, el estadístico de Durbin Watson es útil porque nos indica la existencia de problemas en el modelo, pero a veces no nos ayuda a establecer cuál es la estructura real. En caso de no rechazar la H_o, podemos afirmar que no tenemos un AR(1), pero no sabemos si tenemos alguna otra estructura alternativa.

Por otro lado el estadístico DW sólo debe aplicarse cuando los regresores son **fijos**, en presencia de regresores aleatorios como la variable endógena retardada no tiene validez.

Cuando el estadístico DW cae en zona de duda, y si no podemos llevar a cabo un contraste alternativo, no debemos concluir que no existe autocorrelación. El procedimiento conservador aconseja rechazar la hipótesis nula y estimar por MCGF ya que las consecuencias de ignorar su existencia cuando sí la hay son más graves que las correspondientes al caso contrario.

4.2.2. Contraste de Wallis

El estadístico de Durbin-Watson es adecuado para contrastar la existencia de autocorrelación de primer orden. Wallis modificó dicho estadístico con el objetivo de recoger esquemas de autocorrelación estacional muy presentes en datos trimestrales en los cuales la correlación se produce con cuatro periodos de desfase. Se supone que en este caso las perturbaciones siguen el esquema:

$$u_t = \rho u_{t-4} + \epsilon_t \quad t = 1, 2, \ldots, T \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \quad |\rho| < 1$$

donde se contrasta: $H_0 : \rho = 0$ con el estadístico:

$$DW = \frac{\sum_{t=5}^{T} (\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

Wallis computó los valores críticos precisos para llevar a cabo el contraste bajo el supuesto de X fijas y dependiendo de si el modelo incluye término independiente o no. Así hay dos tablas distintas, la primera para un modelo sin variables ficticias estacionales pero con término independiente y la segunda cuando se incluyen variables ficticias estacionales. La regla de decisión es similar al anterior, lo único que varía son las tablas de referencia.

4.2.3. Contraste h de Durbin

El contraste de Durbin Watson se deriva bajo el supuesto de que los regresores son fijos, sin embargo, éstos pueden ser estocásticos. En muchos modelos econométricos se incluyen retardos de la variable endógena como regresores:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + \gamma_1 Y_{t-1} + \ldots + \gamma_s Y_{t-s} + u_t. \tag{4.3}$$

Recordar que en este caso, el estadístico de Durbin-Watson no tiene validez ya que la condición de que los regresores sean fijos no se cumple.

En estos casos, cuando los únicos regresores estocásticos del modelo son los retardos de la variable endógena aplicaremos el estadístico h de Durbin:

$$\begin{array}{lll} H_0: & \rho = 0 & \text{o bien} \quad H_0 : \rho = 0 \\ H_a: & \rho > 0 & \qquad\quad\;\; H_a : \rho < 0 \end{array}$$

$$h = \hat{\rho}\sqrt{\frac{T}{1 - T\widehat{Var}(\hat{\gamma}_1)}} \xrightarrow{a,H_0} N(0,1)$$

donde T es el tamaño muestral, $\hat{\rho}$ la estimación por MCO del coeficiente de autocorrelación de primer orden en el proceso AR(1), y $\widehat{Var}(\hat{\gamma}_1)$ la varianza estimada de $\hat{\gamma}_1$, el coeficiente estimado correspondiente al primer retardo de Y_t (independientemente de que otros retardos de Y_t aparezcan como regresores.) Dado que la hipótesis alternativa es la existencia de autocorrelación positiva o negativa el contraste se realiza a una cola comparando el valor calculado para h con el valor crítico $N(0,1)_\alpha$. La regla de decisión es:

a) Si $h > N(0,1)_\alpha$ rechazo la H_0 y existe autocorrelación positiva.

b) Si $h < -N(0,1)_\alpha$ rechazo la H_0 y existe autocorrelación negativa.

Observaciones:

a) Si la muestra no es suficientemente grande no se pueden garantizar los resultados del contraste.

b) El único coeficiente que influye en el estadístico es el del primer retardo de Y_t, independientemente del número de retardos incluidos en el modelo.

c) Si $T\widehat{Var}(\hat{\gamma}_1) \geq 1$ el estadístico h de Durbin no es calculable y en este caso Durbin propone calcular la regresión de $\hat{u}_t$ sobre $\hat{u}_{t-1}$ más los regresores del modelo original:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \ldots + \alpha_K X_{Kt} + \delta_1 Y_{t-1} + \ldots + \delta_s Y_{t-s} + \lambda \hat{u}_{t-1} + e_t$$

y ver si el coeficiente asociado a $\hat{u}_{t-1}$ (λ) es significativamente distinto de cero, en cuyo caso no se rechaza la existencia de autocorrelación.

En caso de que sospechemos de un AR de mayor orden (AR(p)), se incluirían dichos retardos:

$$\hat{u}_t = \alpha_1 + \alpha_2 X_{2t} + \ldots + \alpha_K X_{Kt} + \delta_1 Y_{t-1} + \ldots + \delta_s Y_{t-s} + \lambda_1 \hat{u}_{t-1} + \ldots + \lambda_p u_{t-p} + e_t$$

y se analizaría la significatividad conjunta de las variables.

4.2.4. Contraste de Breusch y Godfrey

Este contraste de autocorrelación, además de no imponer la condición de que los regresores sean fijos, permite que la hipótesis alternativa incluya especificaciones más generales que las del AR(1):

$$\begin{array}{ll} H_0: & \text{no autocorrelación de orden } \boldsymbol{p} \\ H_a: & \left\{ \begin{array}{ll} AR(\boldsymbol{p}): & u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_{\boldsymbol{p}} u_{t-\boldsymbol{p}} + \epsilon_t \\ MA(\boldsymbol{p}): & u_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_{\boldsymbol{p}} \epsilon_{t-\boldsymbol{p}} \end{array} \right. \end{array}$$

El contraste de Breusch y Godfrey sugiere:

a) Estimar el modelo original por MCO ignorando la existencia de posible autocorrelación y obtener la serie de residuos mínimo cuadráticos, $\hat{u}_{MCO,t}$.

b) Estimar una regresión auxiliar de $\hat{u}_{MCO,t}$ sobre los p retardos (el orden de autocorrelación de que sospechamos) $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p}$ y las variables explicativas del modelo original:

$$\hat{u}_{MCO,t} = \delta_0 + \delta_1 \hat{u}_{MCO,t-1} + \ldots + \delta_p \hat{u}_{MCO,t-p} + \gamma_2 X_{2t} + \ldots + \gamma_K X_{Kt} + v_t \qquad t = p+1, p+2, \ldots, T$$

y obtener el R^2 de esta regresión.

c) Contrastamos la hipótesis nula de no autocorrelación:

$$H_0 : \text{no autocorrelación}$$

$$H_0 : \delta_1 = \ldots = \delta_p = \gamma_2 = \ldots = \gamma_K = 0$$

con el siguiente estadístico asintótico:

$$TR^2 \xrightarrow{d,H_0} \chi^2_p$$

donde T es el tamaño muestral a pesar de que la regresión incluye retardos y p el número de residuos retardados incluidos en la regresión auxiliar. Rechazamos la H_0 de no autocorrelación si el estadístico calculado es superior al valor de la distribución en las tablas para un nivel de significación dado.

- Interpretación del contraste. Si realmente no existe autocorrelación, entonces en la regresión auxiliar, hay dos tipos de variables explicativas:

  1. Las variables explicativas de la regresión y los residuos son ortogonales, es decir $X'\hat{u} = 0$ (propiedad de la recta de regresión).
  2. Los retardos de los residuos MCO no tienen capacidad explicativa para los residuos.

  Esto implica que el R^2 de la regresión auxiliar es muy pequeño, no rechazaríamos H_0. En caso contrario los retardos de los residuos tienen poder explicativo sobre los residuos, el R^2 es alto y rechazamos la H_0 para un α dado. Existiría autocorrelación del orden contrastado.

  Desventajas:

  a) El orden de la autocorrelación (p) tiene que estar determinado para realizar el contraste, de lo contrario puede ser necesario hacer pruebas sobre el valor de p.

  b) Si se rechaza H_0, no sabemos cuál de las dos especificaciones (AR(p) o MA(p)) es la correcta, por lo que tenemos que comparar los resultados de emplear una u otra.

4.3. MCG: Modelo transformado para AR(1)

Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T$$

donde las perturbaciones siguen un proceso AR(1) tal que $u_t = \rho u_{t-1} + \epsilon_t$, $\epsilon_t \sim iid(0, \sigma^2_\epsilon)$ con ρ conocido y $|\rho| < 1$. La matriz de varianzas y covarianzas viene dado por:

$$E(uu') = \frac{\sigma^2_\epsilon}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \ldots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \rho & \ldots & \rho^{T-3} \\ \rho^3 & \rho^2 & \rho & 1 & \ldots & \rho^{T-4} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & \ldots & 1 \end{pmatrix} = \sigma^2 \Omega$$

como hemos supuesto que ρ es conocido la matriz de varianzas y covarianzas de la perturbación es conocida salvo por σ^2_ϵ. En este contexto el estimador MCO es ineficiente y debemos estimar el modelo por MCG. Tenemos dos procedimientos alternativos para estimar el modelo en presencia de autocorrelación:

1) Aplicar el estimador MCG directamente a la muestra, $\tilde{\beta}_{MCG} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$, obteniendo estimadores lineales, insesgados, de mínima varianza y consistentes. En este caso, y para el ejemplo elegido tenemos:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_T \end{pmatrix} \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_T \end{pmatrix}$$

En cuanto a la matriz de varianzas y covarianzas, $E(uu') = \sigma^2\Omega$, se puede demostrar que:

$$\Omega^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 & 0 \\ -\rho & (1+\rho^2) & -\rho & \dots & 0 & 0 & 0 \\ 0 & -\rho & (1+\rho^2) & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & (1+\rho^2) & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}$$

y el término σ_ϵ^2 se puede estimar de forma insesgada y consistente como:

$$\hat{\sigma}_\epsilon^2 = \frac{\tilde{u}'_{MCG}\Omega^{-1}\tilde{u}_{MCG}}{T-K}$$

2) Estimar el modelo transformado por MCO. La matriz de transformación P^{-1} en el caso de un AR(1) es tal que

$$P^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & -\rho & 1 \end{pmatrix}$$

siendo $\Omega^{-1} = \frac{1}{(1-\rho^2)}(P^{-1})'P^{-1}$. Así, podemos escribir el modelo transformado como:

$$\begin{array}{ll} \sqrt{1-\rho^2}Y_1 = \beta_1\sqrt{1-\rho^2} + \beta_2\sqrt{1-\rho^2}X_1 + \sqrt{1-\rho^2}u_1 & t=1 \\ Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \epsilon_t & t = 2,3,\dots,T \end{array}$$

dado que del proceso AR(1) obtenemos que $\epsilon_t = u_t - \rho u_{t-1}$. En el modelo transformado observamos dos cosas, primero que está formado por dos ecuaciones ya que a la primera observación le corresponde una transformación distinta a las demás. Y segundo, que el término independiente ahora ha cambiado.

De forma alternativa, ante un modelo con perturbaciones que siguen un proceso AR(1) y por tanto no tiene propiedades esféricas, el propio proceso AR(1) nos puede indicar una solución a seguir, ya que ϵ_t es un ruido blanco, es decir, una variable aleatoria con perturbaciones esféricas. En consecuencia, si obtenemos un modelo transformado tal que su única variable aleatoria sea ϵ_t, éste puede ser estimado por MCO obteniendo las propiedades habituales.

Dado $u_t = \rho u_{t-1} + \epsilon_t$, y por tanto $\epsilon_t = u_t - \rho u_{t-1}$ debemos tomar el modelo en el momento t y (t-1), premultiplicar este último por ρ y tomar diferencias para obtener perturbaciones esféricas. Si lo hacemos así tendremos la siguiente ecuación:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \epsilon_t \qquad t = 2,3,\dots,T$$

donde $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ por lo que aplicar MCO en este modelo transformado es correcto.

Lo que hemos hecho en el modelo transformado es descontar la información que se tenía con anterioridad ya que si $u_t = \rho u_{t-1} + \epsilon_t$ lo ocurrido en t se explica en parte por lo ocurrido en (t-1), al tener $(Y_t - \rho Y_{t-1})$ nos queda únicamente la información nueva. En este caso las variables transformadas se definen como:

$$Y_t^\star = (Y_t - \rho Y_{t-1}) \qquad X_t^\star = (X_t - \rho X_{t-1}).$$

Hay que notar que en este modelo hemos perdido la primera observación pero si la muestra es suficientemente grande es una pérdida de eficiencia irrelevante.

Si la muestra es pequeña, es mejor estimar el modelo transformado utilizando las dos ecuaciones (es decir incluyendo la ecuación para t=1) ya que podríamos tener cambios significativos en las estimaciones:

$$\begin{array}{ll} \sqrt{1-\rho^2}Y_1 = \beta_1\sqrt{1-\rho^2} + \beta_2\sqrt{1-\rho^2}X_1 + \sqrt{1-\rho^2}u_1 & t=1 \\ Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \epsilon_t & t = 2,3,\dots,T \end{array}$$

Las matrices correspondientes serían:

$$Y^\star = P^{-1}Y = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}\ Y_1 \\ Y_2 - \rho Y_1 \\ Y_3 - \rho Y_2 \\ \vdots \\ Y_T - \rho Y_{T-1} \end{pmatrix}$$

$$X^\star = P^{-1}X = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_T \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\ X_1 \\ (1-\rho) & X_2 - \rho X_1 \\ (1-\rho) & X_3 - \rho X_2 \\ \vdots & \vdots \\ (1-\rho) & X_T - \rho X_{t-1} \end{pmatrix}$$

$$u^\star = P^{-1}u = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & -\rho & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_t \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}\ u_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

Cuando el modelo transformado está formado solamente por la segunda ecuación, las matrices correspondientes se obtienen de suprimir la primera fila de cada una de ellas.

4.4. MCGF: Aplicación para un AR(1)

Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1,2,\dots,T$$

donde las perturbaciones siguen un proceso AR(1) tal que $u_t = \rho u_{t-1} + \epsilon_t$ y $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ con ρ desconocido y $|\rho| < 1$. En este caso la matriz de varianzas y covarianzas de la perturbación sigue teniendo la misma forma que en la sección anterior, $E(uu') = \frac{\sigma_\epsilon^2}{1-\rho^2}\Omega$, pero al ser ρ desconocido, Ω es desconocida. En este contexto el método de estimación apropiado es el de MCGF.

Para emplear el estimador por MCGF tenemos que estimar el parámetro ρ consistentemente porque así logramos que $\hat{\Omega}$ sea consistente. El parámetro ρ puede ser estimado a priori o bien conjuntamente con los coeficientes del modelo.

A continuación describimos algunas formas alternativas de estimar ρ:

- **Metodo de Durbin en dos etapas:**
  Se basa en la estimación del modelo transformado en el que las perturbaciones son esféricas:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t + \rho\beta_2 X_{t-1} + \epsilon_t.$$

  En este modelo la variable a explicar es desconocida dado que depende de ρ por lo que se traslada este término a la parte sistemática:

$$Y_t = \underbrace{\beta_1(1-\rho)}_{\alpha_0} + \beta_2 X_t + \underbrace{\rho\beta_2}_{\alpha_1} X_{t-1} + \rho Y_{t-1} + \epsilon_t$$

  que estimado por MCO proporciona estimadores consistentes. Una vez obtenida la estimación de ρ, se procede a estimar el modelo original por cualquiera de las dos alternativas.

- **Proceso iterativo de Cochrane-Orcutt:**

  El proceso iterativo consta de los siguientes pasos:

  1. Estimar por MCO el modelo original ignorando la existencia de un proceso AR(1) en la perturbación:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T \tag{4.4}$$

     guardar los residuos MCO, $\hat{u}_t$.
  2. Utilizar los residuos MCO de la regresión anterior para estimar el parámetro ρ en la regresión $\hat{u}_{t,MCO} = \rho\hat{u}_{t-1,MCO} + v_t \quad t = 2, 3, \ldots, T$:

$$\tilde{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2}$$

  3. Se utiliza $\tilde{\rho}$ para transformar el modelo:

$$Y_t - \tilde{\rho} Y_{t-1} = \beta_1(1-\tilde{\rho}) + \beta_2(X_t - \tilde{\rho} X_{t-1}) + \epsilon_t \qquad t = 2, 3, \ldots, T \tag{4.5}$$

     que se estima por MCO, obteniendo las estimaciones MCGF de los coeficientes, $\tilde{\beta}_1$ y $\tilde{\beta}_2$.

  Este proceso se itera hasta que dos estimaciones consecutivas de $'$ alcancen un grado de convergencia prefijado de antemano: $|\tilde{\rho}^{(i+1)} - \tilde{\rho}^{(i)}| < \xi$, donde ξ es una cantidad prefijada de antemano. Para iterar basta con generar una nueva serie de residuos con las estimaciones de $\tilde{\beta}_1$ y $\tilde{\beta}_2$ Dada la nueva serie de residuos, se repite el proceso desde el segundo paso.

  Notar que:

  1) El término independiente en el modelo transformado es $\beta_1(1-\tilde{\rho}^2)$ por lo que será necesario recuperar β_1.
  2) Este procedimiento de estimación puede presentar dos problemas: convergencia y existencia de mínimos locales en vez de mínimos absolutos.

3) Es preciso tener en cuenta que los dos métodos considerados minimizan la suma de cuadrados sin tener en cuenta la primera observación, por lo que sólo son aproximaciones al estimador de MCGF. Asintóticamente ambos son equivalentes al estimador de MCGF pero para muestras pequeñas puede haber diferencias, a veces, importantes.
4) Si desde el tercer paso se tienen en cuenta las dos ecuaciones a la hora de estimar el modelo transformado, mejoran los resultados en muestras pequeñas. Este proceso iterativo con información completa fue propuesto por Prais-Winsten (1957).

- **Método de red de búsqueda de Hildreth-Lu:**

  El procedimiento de estimación de ρ de Hildreth-Lu (1960) es similar al procedimiento anterior. La diferencia radica en que en lugar de hallar una primera aproximación de ρ mediante MCO, se propone recorrer el dominio de ρ, es decir, el intervalo (-1, 1):

  a) Se particiona el intervalo $(-1, 1)$ en n-puntos arbitrarios equidistantes entre sí: $\rho_i \quad i = 1, 2, \ldots, n$.
  b) Una vez prefijados estos puntos, para cada uno de los valores de ρ_i fijados se transforma el modelo de la forma usual:

  $$Y_t - \rho_i Y_{t-1} = \beta_1(1 - \rho_i) + \beta_2(X_t - \rho_i X_{t-1}) + \epsilon_t \qquad t = 2, 3, \ldots, T$$

  se estima por MCO y se obtienen las correspondientes sumas de cuadrados de los residuos.
  c) Como estimación definitiva, tanto de ρ como de los parámetros desconocidos del modelo (en el ejemplo β_1 y β_2), se toma aquella que proporciona la menor SCR_i.

  Este método presenta la ventaja de que, utilizando una red suficientemente fina -n elevado- queda prácticamente garantizada la aproximación al mínimo global. Por ello, con frecuencia se utiliza este procedimiento inicialmente para detectar la región donde está el mínimo absoluto y proseguir, a continuación, con el método de Cochrane-Orcutt.

4.5. MCO: Estimador de Newey-West de $V(\hat{\beta}_{MCO})$

Al igual que en el caso de heterocedasticidad el estimador de White proporciona un estimador consistente de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$ cuando se desconoce totalmente la estructura de heterocedasticidad, para autocorrelación tenemos un estimador en la misma línea.

Este estimador, de Newey-West (1987), propone utilizar como estimador de $G = \lim_{T\to\infty} \frac{\sigma^2 X'\Omega X}{T}$ la expresión:

$$S_T = \frac{1}{T}\sum_{j=1}^{T} \hat{u}_t^2 X_t X_t' + \frac{1}{T}\sum_{\ell=1}^{L}\sum_{t=\ell+1}^{T} w_\ell \hat{u}_t \hat{u}_{t-\ell}[X_t X_{t-\ell}' + X_{t-\ell} X_s]$$

donde $w_\ell = 1 - \frac{\ell}{L+1}$ siendo L el orden máximo de la autocorrelación del término de error, que no siempre será fácil de determinar.

Newey-West demostraron que:

$$\text{plim}\,(S_T) = \text{plim}\left(\frac{\sigma^2}{T} X'\Omega X\right) = G.$$

Por lo tanto, la distribución asintótica del estimador MCO bajo autocorrelación viene dado por:

$$\sqrt{N}(\hat{\beta}_{MCO} - \beta) \xrightarrow{a} N\left(0, \lim_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1} \operatorname{plim}_{T\to\infty}(S_T) \lim_{T\to\infty}\left(\frac{X'X}{T}\right)^{-1}\right)$$

$$\sqrt{N}(\hat{\beta}_{MCO} - \beta) \xrightarrow{a} N\left(0, M^{-1}GM^{-1}\right)$$

Esta matriz de varianzas y covarianzas es consistente y puede ser utilizada para hacer inferencia en muestras grandes, sin tener que especificar a priori la estructura de autocorrelación.

4.6. El estimador de la varianza de la perturbación

El parámetro desconocido σ^2_ϵ la estimamos mediante la expresión habitual

$$\tilde{\sigma}^2_\epsilon = \frac{\tilde{u}'_{MCGF}\tilde{\Omega}^{-1}\tilde{u}_{MCGF}}{T-K}$$

para lo cual habremos tenido que estimar el parámetro ρ de antemano y obtener las estimaciones MCGF de los coeficientes del modelo. Bajo las mismas condiciones que vimos en el primer tema, este estimador será consistente.

Si optamos por aplicar MCO en el transformado podemos estimar σ^2_ϵ como: $\sigma^2_\epsilon = \frac{\tilde{u}'_{MCO}\tilde{u}_{MCO}}{T-k}$ en el caso de que $'$ sea desconocido debemos estimar σ^2_ϵ y $'$ y estaríamos aplicando el estimador de MCGF. En este caso hay varias formas correctas de estimar σ^2_ϵ por MCO:

- $u_t = \rho u_{t-1} + \epsilon_t$ esta regresión permite obtener estimadores para $'$ y σ^2_ϵ
- también podemos obtenerlo en: $(Y_t - \beta_1 - \beta_2 X_t) = \rho(Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) + \epsilon_t$

en cualquier caso tras obtener $\hat{\sigma}^2_\epsilon$ tendremos

$$\tilde{\sigma}^2_u = \frac{\tilde{\sigma}^2_\epsilon}{(1-\tilde{\rho}^2)}$$

y así el conocimiento de la matriz de varianzas y covarianzas es total.

4.7. Contraste de restricciones lineales con Ω desconocida

Los contrastes de restricciones lineales se llevan a cabo de la misma forma que vimos en el tema anterior.

- Si $var(u) = \sigma^2\Omega$ conocida
  Contrastamos restricciones lineales del tipo $H_0 : R\beta = r$ con el estadístico:

$$\frac{(R\tilde{\beta}_{MCG} - r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCG} - r)}{\sigma^2} \overset{H_0}{\sim} \chi_{(q)}$$

- Si $var(u) = \sigma^2\Omega$ con σ^2 desconocido
  Contrastamos restricciones lineales del tipo $H_0 : R\beta = r$ con el estadístico:

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

- Si $var(u) = \sigma^2\Omega$ desconocido pero estimable
Para contrastar hipótesis nulas del tipo $H_0 : R\beta = r$ tendremos que aplicar el estadístico:

$$\frac{(R\tilde{\beta}_{MCGF} - r)'[R(X'\hat{\Omega}^{-1}X)^{-1}R']^{-1}(R\tilde{\beta}_{MCGF} - r)}{\hat{\sigma}^2} \xrightarrow{d,H_0} \chi^2_{(q)}$$

- Si $var(u) = \sigma^2\Omega$ totalmente desconocido
Aplicamos el estimador de Newey-West, el contraste de hipótesis se realiza con el estadístico:

$$T(R\hat{\beta}_{MCO} - r)'(RM^{-1}GM^{-1}R')^{-1}(R\hat{\beta}_{MCO} - r) \xrightarrow{d,H_0} \chi^2_{(q)}$$

donde lo que varía es cómo obtener G que ahora correspondería a la expresión proporcionada por Newey-West.

4.8. Predicción bajo autocorrelación de primer orden

Bajo el supuesto de autocorrelación tenemos que

$$\begin{array}{ll} Y = X\beta + u & u \sim N(0, \sigma^2\Omega) \\ u_t = \rho u_{t-1} + \epsilon & \epsilon_t \sim (0, \sigma^2_\epsilon) \\ \rho \quad \Omega \quad \text{conocidos} & \end{array}$$

por lo que en el modelo transformado está dado por:

$$\begin{array}{l} Y^* = X^*\beta + u^* \qquad u^* = \epsilon \quad \epsilon \sim N(0, \sigma^2_\epsilon I) \\ Y_t - Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + \ldots + \epsilon_t \quad \epsilon \sim (0, \sigma^2_\epsilon) \\ Y^*_{T+1} = X^{*\prime}_{T+1}\beta + \epsilon_{T+1} \end{array}$$

Si queremos predecir el valor Y^*_{T+1} tenemos que:

$$\begin{array}{l} Y^*_{T+1} = X^{*\prime}_{T+1}\beta + \epsilon_{T+1} \\ Y_{T+1} - \rho Y_T = X'_{T+1}\beta - \rho X'_T\beta + \epsilon_{T+1} \\ Y_{T+1} = X'_{T+1}\beta + \rho(Y_T - X'_T\beta) + \epsilon_{T+1} \end{array}$$

Predicción por punto

La predicción por punto en presencia de un proceso AR(1) en las perturbaciones se obtiene como:

$$\begin{array}{l} \hat{Y}_{T+1} = X'_{T+1}\tilde{\beta}_{MCG} + \rho(Y_T - X'_T\tilde{\beta}_{MCG}) \\ \hat{Y}_{T+1} = X'_{T+1}\tilde{\beta}_{MCG} + \rho\hat{u}_T \end{array}$$

donde podemos observar que la diferencia es que se incorpora un término de corrección debido a que parte del error nuevo u_{T+1} puede ser predicho por $\rho\hat{u}_T$.

Predicción por intervalo del valor, Y_p:

Error de predicción:

$$\begin{aligned}
e_{T+1} &= Y_{T+1} - \hat{Y}_{T+1} = \\
&= X'_{T+1}\beta + \rho(Y_T - X'_T\beta) + \epsilon_{T+1} - X'_{T+1}\tilde{\beta}_{MCG} - \rho(Y_T - X'_T\tilde{\beta}_{MCG}) = \\
&= -(X'_{T+1} - \rho X'_T)(\tilde{\beta}_{MCG} - \beta) + \epsilon_{T+1} = \\
&= -X^{*'}_{T+1}(\tilde{\beta}_{MCG} - \beta) + \epsilon_{T+1}
\end{aligned}$$

Distribución:

$E(e_{T+1}) = E(-X^{*'}_{T+1}(\tilde{\beta}_{MCG} - \beta) + \epsilon_{T+1}) = 0$

$$\begin{aligned}
\text{Var}(e_{T+1}) &= E[(e_{T+1} - E(e_{T+1}))^2] = E(e_{T+1}e'_{T+1}) \\
&= E[(-X^{*'}_{T+1}(\tilde{\beta}_{MCG} - \beta) + \epsilon_{T+1})(-X^{*'}_{T+1}(\tilde{\beta}_{MCG} - \beta) + \epsilon_{T+1})'] \\
&= E(\epsilon^2_{T+1}) + X^{*'}_{T+1}E(\tilde{\beta}_{MCG} - \beta)(\tilde{\beta}_{MCG} - \beta)'X^*_{T+1} - 2X^{*'}_{T+1}E((\tilde{\beta}_{MCG} - \beta)\epsilon_{T+1}) \\
&= \sigma^2_\epsilon + \sigma^2_\epsilon X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X^*_{T+1} - 2X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(u\ \epsilon_{T+1}) \\
&= \sigma^2_\epsilon + \sigma^2 X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X^*_{T+1}
\end{aligned}$$

ya que la esperanza $E(u\ \epsilon_{T+1})$ es cero:

$$\begin{aligned}
E(u\epsilon_{T+1}) &= E(u(u_{T+1} - \rho u_T)) = \\
&= \begin{pmatrix} E(u_1\epsilon_{T+1}) \\ E(u_2\epsilon_{T+1}) \\ E(u_3\epsilon_{T+1}) \\ \vdots \\ E(u_T\epsilon_{T+1}) \end{pmatrix} = \begin{pmatrix} E(u_1(u_{T+1} - \rho u_T)) \\ E(u_2(u_{T+1} - \rho u_T)) \\ E(u_3(u_{T+1} - \rho u_T)) \\ \vdots \\ E(u_T(u_{T+1} - \rho u_T)) \end{pmatrix} = \begin{pmatrix} \gamma_T - \rho\gamma_{T-1} \\ \gamma_{T-1} - \rho\gamma_{T-2} \\ \gamma_{T-2} - \rho\gamma_{T-3} \\ \vdots \\ \gamma_1 - \rho\gamma_0 \end{pmatrix} = \\
&= \begin{pmatrix} \sigma^2\rho^T - \sigma^2\rho\rho^{T-1} \\ \sigma^2\rho^{T-1} - \sigma^2\rho\rho^{T-2} \\ \sigma^2\rho^{T-2} - \sigma^2\rho\rho^{T-3} \\ \vdots \\ \sigma^2\rho - \sigma^2\rho \end{pmatrix} = \begin{pmatrix} \sigma^2\rho^T - \sigma^2\rho^T \\ \sigma^2\rho^{T-1} - \sigma^2\rho^{T-1} \\ \sigma^2\rho^{T-2} - \sigma^2\rho^{T-2} \\ \vdots \\ \sigma^2\rho - \sigma^2\rho \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\end{aligned}$$

Bajo normalidad de las perturbaciones, $u \sim N(0, \sigma^2\Omega)$:

$$e_{T+1} \sim N(0, \text{Var}(\epsilon_{T+1}) + X^{*'}_{T+1}\text{Var}(\tilde{\beta}_{MCG})X^*_{T+1})$$

$$e_{T+1} \sim N\left(0, \sigma^2_\epsilon + \sigma^2 X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X^*_{T+1}\right)$$

$$e_{T+1} \sim N\left(0, \sigma^2_\epsilon + \frac{\sigma^2_\epsilon}{1-\rho}X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X^*_{T+1}\right)$$

de donde

$$IC_{1-\alpha}(Y_{T+1}) = \left[\hat{Y}_{T+1} \pm N(0,1)_{\alpha/2}\frac{\sigma_\epsilon}{\sqrt{1-\rho}}\sqrt{1 + X^{*'}_{T+1}(X'\Omega^{-1}X)^{-1}X^*_{T+1}}\right]$$

$$IC_{1-\alpha}(Y_{T+1}) = \left[(X'_{T+1}\hat{\beta}_{MCG} + \rho\hat{u}_T) \pm N(0,1)_{\alpha/2}\sigma\sqrt{1 + (X'_{T+1} - \rho X'_T)(X'\Omega^{-1}X)^{-1}(X'_{T+1} - \rho X'_T)}\right]$$

A partir de aquí, los ajustes necesarios no distan de los realizados para heterocedasticidad.

Tema 5

Regresores Estocásticos

5.1. Introducción

Durante buena parte de estos contenidos hemos mantenido el supuesto básico sobre los regresores, de que X era una matriz de variables explicativas fijas (no estocásticas). Este supuesto es apropiado para experimentos de laboratorio o para variables como la tendencia o variables ficticias, sin embargo, no se ajusta a la realidad. En este tema y siguientes relajaremos dicho supuesto para adecuarnos a la realidad económico-social. Por ejemplo, si analizamos la relación entre consumo y renta no podemos suponer que la variable explicativa renta sea fija ya que tanto el consumo como la renta vienen determinados por el mismo sistema económico-social y son aleatorios.

En este tema analizaremos si los métodos de estimación e inferencia vistos hasta ahora son válidos cuando X es estocástica. En caso de que no sea así analizaremos qué métodos alternativos están disponibles.

Sea $\quad Y = X\beta + u \quad$ donde X es estocástica. En este entorno el estimador MCO de β se define como $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u \quad$ es decir, es una función estocástica no lineal de X y u y por tanto sus propiedades dependen de la distribución conjunta de estas. Para analizar si el estimador es insesgado buscamos su valor esperado:

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

para poder obtener

$$E[(X'X)^{-1}X'u]$$

debemos suponer o conocer la distribución conjunta de las variables aleatorias X y u . Bajo el supuesto de regresores fijos, el problema se soluciona fácilmente:

$$E[(X'X)^{-1}X'u] = (X'X)^{-1}X'E(u) = 0 \quad \text{ya que} \quad E(u) = 0$$

pero bajo regresores estocásticos la igualdad no se cumple y es preciso contar con la distribución conjunta de X y u para poder derivar propiedades de los estimadores $\hat{\beta}_{MCO}$ así como las distribuciones de los estadísticos de contraste habituales.

Una forma de enfocar el problema es utilizar la distribución de Y condicionada a X. Podemos escribir la distribución conjunta $f(Y, X; \beta, \sigma_u^2)$ como:

$$f(Y, X; \beta, \sigma_u^2) = f(Y/X; \beta, \sigma_u^2) \bullet f(X; \gamma)$$

$$\text{distribución conjunta} = \text{distribución condicionada} \bullet \text{distribución marginal}$$

En términos de la perturbación u :

$$f(u, X; \beta, \sigma_u^2) = f(u/X; \beta, \sigma_u^2) \bullet f(X; \gamma)$$

$$f(u/X) = \frac{f(u, X)}{f(X)}$$

Si nuestro interés se centra en los parámetros de la distribución condicionada (β, σ_u^2) y estos no están relacionados con los parámetros de la distribución marginal γ, podemos olvidarnos de ella y considerar sólo la distribución de Y **condicionada a unos valores fijos** de las variables aleatorias X.

El modelo lineal general condicionado a X se puede escribir como:

$$Y = X\beta + u$$

donde:

$$\begin{array}{l} E(u/X) = 0 \\ E(uu'/X) = \sigma_u^2 I_T \\ rg(X) = k > T \\ u/X \sim N(0, \sigma_u^2 I_T) \end{array}$$

y podemos derivar los siguientes resultados condicionados:

$$\begin{array}{ll} E(\hat{\beta}/X) & = \beta + E[(X'X)^{-1}X'u/X] = \beta + (X'X)^{-1}X'E(u/X) = \beta \\ \sum_{(\hat{\beta}/X)} & = E[(X'X)^{-1}X'uu'X(X'X)^{-1}/X] = \\ & = (X'X)^{-1}X'E(uu'/X)X(X'X)^{-1} = \\ & = (X'X)^{-1}X'\sigma_u^2 I_T X(X'X)^{-1} = \\ & = \sigma_u^2(X'X)^{-1} \\ E(\hat{\sigma}_u^2/X) & = \sigma_u^2 \\ \widehat{\sum}_{(\hat{\beta}/X)} & = \hat{\sigma}_u^2(X'X)^{-1} \end{array}$$

El estimador de $\hat{\beta}_{MCO}$ no es un estimador lineal, sino una función estocástica no lineal de X e Y, por lo que estrictamente hablando no podemos aplicar el Teorema de Gauss Markov y decir que tiene menor varianza entre todos los lineales e insesgados. Sin embargo, si consideramos la varianza del estimador como condicionada a valores fijos de X, entonces, el estimador es eficiente. La distribución condicionada a los valores fijos de X es:

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

y los estadísticos t y F condicionados a X siguen una distribución t-student y F-snedecor respectivamente. De esta forma, aunque en principio, las variables X son v.a. si condicionamos a unos valores fijos de éstas, los resultados habituales se mantienen. Esta conclusión no es sorprendente, porque trabajar condicionado a X es tratar a X como si fuera una matriz de variables fijas y los resultados dependen de los valores concretos que toman estas variables en la muestra.

El problema se plantea cuando nos encontramos con situaciones en las que los regresores son estocásticos y no tiene sentido realizar un análisis condicionado a unos valores fijos de X. Para ilustrar en qué situaciones no podemos hacer este supuesto vamos a considerar tres ejemplos:

Ejemplo 1

Suponemos el siguiente modelo de regresión:

$$Y_t = \alpha + \beta Y_{t-1} + u_t \qquad t = 2, 3, \ldots, T \qquad u_t \sim N(0, \sigma_u^2)$$

en este modelo aparece como regresor la variable dependiente retardada un periodo. Dado que Y_t es una v.a. el regresor Y_{t-1} también lo es. En esta situación la matriz de regresores X es estocástica y se define como:

$$X = [\vec{1}; Y_{t-1}]$$

Por otro lado, no podemos realizar el análisis condicionado a unos valores fijos de Y_{t-1} ya que no tendría sentido porque es el propio modelo estocástico el que indica cómo se generan.

Ejemplo 2
Dado el siguiente modelo de regresión:

$$Y_t = \alpha + \beta X_t + v_t \qquad v_t \sim N(0, \sigma_v^2) \qquad t = 1, 2, \ldots, T$$

siendo X_t la habilidad de un trabajador, variable no observable ya que es difícil de medir. En su lugar observamos la variable $X_t^\star$ años de experiencia del trabajador en el puesto de trabajo, tal que:

$$X_t^\star = X_t + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2) \qquad Cov(\epsilon_t, v_t) = 0 \quad \forall t \qquad t = 1, 2, \ldots, T$$

donde ϵ_t es una v.a. que recoge el error de medida en t. En esta situación, $X_t^\star$ es una v.a. aunque consideremos a X_t como fija. El modelo estimable sería:

$$Y_t = \alpha + \beta X_t^\star + (v_t - \beta \epsilon_t)$$

llamamos $u_t = (v_t - \beta\epsilon_t)$, en este caso u_t es una función de ϵ_t y X por ello no tendría sentido realizar un análisis condicionado a unos valores fijos de X.

Ejemplo 3
Supongamos que se quiere estimar los parámetros de la siguiente ecuación de demanda de un bien:

$$Q_t = \alpha + \beta P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde Q es la cantidad demandada y P es el precio. Dado que en el momento t observamos la cantidad y precio de equilibrio, ambas variables se determinan simultáneamente en el mercado. Luego tanto Q como P son variables endógenas. Si en t se produce un shock en la demanda del bien debido por ejemplo, a un cambio en los gustos de los consumidores. Al ser recogido por u_t se genera un cambio tanto en la cantidad demandada como en el precio. En este contexto dado que las variables se determinan simultáneamente ambas son aleatorias. Este es otro ejemplo donde la matriz de regresores es estocástica y no tiene sentido realizar el análisis condicionado a valores fijos de $P_t \quad t = 1, 2, \ldots, T$, dado que P_t se determina simultáneamente a Q_t.

5.2. Propiedades de los MCO

Estudiaremos las propiedades del estimador MCO en las siguientes situaciones:

2.1 Independencia entre regresor y error.

2.2 Incorrelación contemporánea entre regresor y error.

2.3 Correlación entre regresor y error.

5.2.1. Independencia entre regresor y error

Vamos a buscar las propiedades del estimador MCO cuando X y u son independientes. Sea $Y = X\beta + u$ donde:

1. X es una matriz estocástica, (alguno de sus regresores es una v.a) con una determinada función de densidad $f(X)$, es decir X toma diferentes valores con diferentes probabilidades.
2. $u \sim N(0, \sigma_u^2 I_T)$
3. X y u se distribuyen independientemente, es decir $E(X'u) = E(X')E(u) = 0$.
4. $plim \frac{1}{T} X'X = Q$ simétrica, semidefinida (+) y no singular.
5. $plim \frac{1}{T} X'u = 0$ ya que se cumple el teorema de Mann y Wald.

La hipótesis dos garantiza que la función de densidad marginal de $(X_{1t}, \quad X_{2t}, \quad \ldots \quad, X_{kt})$ no depende de los parámetros $(\beta, \sigma_u^2) \quad \forall t$ y por tanto:

$$f(u/X) = f(u) \quad \text{ya que :}$$

$$f(u/X) = \frac{f(u, X)}{f(X)} = \frac{f(u)f(X)}{f(X)}$$

lo que en términos de valores esperados significa:

$$\begin{array}{l} E(u/X) = E(u) = 0 \\ E(uu'/X) = E(uu') = \sigma_u^2 I_T \\ E(Y/X) = E(X\beta + u/X) = E(X\beta) + E(u/X) = X\beta + E(u) = X\beta \\ Var(Y/X) = Var(u/X) = Var(u) = \sigma_u^2 I_T \end{array}$$

Esto significa que podríamos estimar el modelo de forma condicionada, lo que implica tratar a X como fija y por tanto tendríamos los resultados conocidos. Pero si las X son estocásticas lo lógico es hacer inferencia no condicionada.

- Resultados en muestras finitas para valores de X no condicionados:

  a) Linealidad:
  $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'u$ **función no lineal** de X y u.
  El estimador $\hat{\beta}_{MCO}$ ya no es una combinación lineal de las perturbaciones, sino que es una función estocástica no lineal de X y u, por lo que sus propiedades dependen de la distribución conjunta de éstas.

  b) Insesgadez:
  Dado que X y u son independientes y $E(u) = 0$ por hipótesis básica:

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

  por tanto $\hat{\beta}_{MCO}$ es insesgado si X y u son independientes y $E[(X'X)^{-1}X']$ existe y es finito.
  Para demostrar esta propiedad hemos utilizado el siguiente resultado estadístico:

  $$E(a) = E_b[E_{a/b}] \qquad \text{y} \qquad E_{WX} = E_X[E_{W/X}]$$

  ya que :

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

c) Matriz de varianzas y covarianzas:

$$\begin{array}{ll}
\sum_{\hat{\beta}_{MCO}} = E(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)' & = E[(X'X)^{-1}X'uu'X(X'X)^{-1}] = \\
\text{Aplicando} \quad E_{WX} = E_X[E_{W/X}] & \\
 & = E_X\{E_{u/X}[(X'X)^{-1}X'uu'X(X'X)^{-1}]\} = \\
 & = E_X\{(X'X)^{-1}X'E_{u/X}(uu')X(X'X)^{-1}\} = \\
\text{dado que} \quad E(uu') = \sigma^2_u I_T & \\
 & = E_X\{(X'X)^{-1}X'\sigma^2_u I_T X(X'X)^{-1}\} = \\
 & = \sigma^2_u E_X\{(X'X)^{-1}\}
\end{array}$$

$\hat{\beta}_{MCO}$ sigue siendo el estimador insesgado de mínima varianza.
siendo:

E_X el valor esperado de la distribución marginal de X.
$E_{u/X}$ el valor esperado de la distribución condicional de u dado X.
$E_X(X'X)^{-1}$ es la matriz de covarianzas poblacional de los regresores calculada en la distribución marginal de X.

Un estimador insesgado de $\sum_{\hat{\beta}_{MCO}}$ donde σ^2_u y $E_X(X'X)^{-1}$ son los dos elementos desconocidos es:

$$\widehat{\sum}(\hat{\beta}_{MCO}) = \hat{\sigma}^2_u\{E_X(X'X)^{-1}\}$$

$$E\left[\sum_{\hat{\beta}_{MCO}}\right] = E_X\{E_{S^2/X}[\hat{\sigma}^2_u(X'X)^{-1}]\} = E_X\{\sigma^2_u(X'X)^{-1}\} = \sigma^2_u E_X(X'X)^{-1}$$

- Distribución e inferencia:
El estimador $\hat{\beta}_{MCO}$ es una combinación no lineal de X y u y por tanto no tiene porqué tener una distribución normal incluso aunque X e u la tengan. Como consecuencia no tenemos garantizado que $\hat{\beta}_{MCO} \sim N(\quad,\quad)$ y por tanto los estadísticos t y F no tienen una distribución exacta conocida, por ello la inferencia no es válida.

Conclusión: La eliminación del supuesto de que X es fija sustituyéndolo por X estocástica pero independiente de u no altera las propiedades deseables ni la variabilidad de la estimación mínimo cuadrática.

(Nota: Green tiene una referencia a Hamilton en que éste demuestra que los estadísticos t y F son válidos para este caso de independencia entre X y u.)

- Propiedades en muestras grandes:
Bajo los supuestos habituales, y si además se satisface que $plim\left(\frac{1}{T}X'X\right) = Q$ finita y no singular es posible derivar las propiedades asintóticas para los estimadores MCO utilizando el Teorema de Mann y Wald y el Teorema de Cramer.

a) Consistencia:

$$plim\hat{\beta}_{MCO} = plim\beta + plim\left(\frac{1}{T}X'X\right)^{-1} plim\left(\frac{1}{T}X'u\right)$$

I) $plim\left(\frac{1}{T}X'X\right)^{-1} = Q^{-1}$

II) $plim\left(\frac{1}{T}X'u\right) = 0$ ya que se cumple el Teorema de Mann y Wald. Veamos sus condiciones:

i) $u_1, u_2, \ldots, u_T$ v.a tal que $u_t \sim iid(0, \sigma^2_u)$
ii) $E(X'u) = E(X')E(u) = 0$ por independencia entre X y u.
ii) $plim\left(\frac{1}{T}X'X\right) = Q$ finita, simétrica y no singular

i) + ii) + iii) implican los resultados siguientes:

1. $plim\left(\frac{1}{T}X'u\right) = 0$
2. $\frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma_u^2 Q)$

Así,

$$plim(\hat{\beta}_{MCO}) = \beta + Q^{-1} \cdot 0 = \beta$$

por tanto $\hat{\beta}_{MCO}$ es un estimador consistente del parámetro β.

b) Distribución asintótica e inferencia asintótica:
Aplicando el Teorema de Cramer:
Si $Y_T = A_T \cdot Z_T$ y :

$$\begin{array}{ll} plim A_T = A & \\ Z_T \overset{a}{\sim} N(\mu, \sum) & \Rightarrow A_T Z_T \overset{a}{\sim} N(A_\mu, A \sum A') \end{array}$$

de donde podemos escribir:

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

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

Dado que como acabamos de ver se cumple el Teorema de Mann y Wald, con los dos resultados anteriores obtenemos:

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'u\right) \xrightarrow{d} N(0, Q^{-1} \cdot \sigma_u^2 \cdot Q \cdot (Q^{-1})')$$

de donde

$$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \sigma_u^2 Q^{-1})$$

Dado que se cumple el Teorema de Mann y Wald las propiedades asintóticas se mantienen y tiene sentido la inferencia asintótica.

Bajo $H_0 : R\beta = r$, los estadísticos t y F se distribuyen asintóticamente como $N(0,1)$ y $\chi^2_{(q)}$ respectivamente, si el tamaño de la muestra es suficientemente grande. Por lo tanto, podemos utilizar estas distribuciones asintóticas para aproximar la distribución exacta de los estadísticos. Así, para contrastar q restricciones lineales de la forma: $Ho : R\beta = r$
$Ha : R\beta \neq r$
utilizamos el siguiente estadístico:

$$\frac{(R\hat{\beta}_{MCO} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r)}{\hat{\sigma}_u^2} \xrightarrow{d} \chi^2_{(q)}$$

Si q=1 podemos utilizar el estadístico:

$$\frac{R\hat{\beta}_{MCO} - r}{\hat{\sigma}_u\sqrt{(R(X'X)^{-1}R')}} \xrightarrow{d} N(0,1)$$

Si por ejemplo queremos contrastar la significatividad de una de las variables explicativas: $Ho : \beta_i = 0$ versus $Ha : \beta_i \neq 0$ en este caso q=1 y podemos escribir el estadístico a utilizar como:

$$\frac{\hat{\beta}_{i,MCO}}{\widehat{des}_{(\hat{\beta}_{i,MCO})}} \xrightarrow{d} N(0,1)$$

Observación: El supuesto de independencia entre regresores y perturbación es muy restrictivo en ciertas ocasiones. Por ejemplo, en los casos ilustrados anteriormente no se cumple.

Ejercicio:
En el modelo lineal simple:

$$Y_t = \alpha + \beta X_t + u_t \qquad u_t \sim iid(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T$$

donde X_t es una variable estocástica pero independiente de la perturbación buscar el estimador de MCO y sus propiedades en muestras finitas.

El estimador de MCO se define.

$$\hat{\beta}_{MCO} = \frac{\sum y_t x_t}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2}$$

Buscamos ahora sus propiedades en muestras finitas:

$$\begin{aligned}
E(\hat{\beta}_{MCO}) &= E\left[\beta + \frac{\sum x_t u_t}{\sum x_t^2}\right] = \\
&= \beta + E\left[\frac{x_1 u_1}{\sum x_t^2} + \frac{x_2 u_2}{\sum x_t^2} + \ldots + \frac{x_T u_T}{\sum x_t^2}\right] = \\
&= \beta + E\left[\frac{x_1}{\sum x_t^2}\right] E(u_1) + E\left[\frac{x_2}{\sum x_t^2}\right] E(u_2) + \ldots + E\left[\frac{x_T}{\sum x_t^2}\right] E(u_T) = \beta
\end{aligned}$$

Y por lo tanto, dado que $E(u_t) = 0 \quad \forall t$, $\hat{\beta}_{MCO}$ es insesgado si $E\left[\frac{x_t}{\sum x_t^2}\right]$ existe y es finito $\forall t$.

$$\begin{aligned}
Var(\hat{\beta}_{MCO}) &= E(\hat{\beta}_{MCO} - \beta)^2 = \\
&= E\left[\frac{\sum x_t u_t}{\sum x_t^2}\right]^2 = \\
&= E\left[\frac{x_1 u_1}{\sum x_t^2} + \frac{x_2 u_2}{\sum x_t^2} + \ldots + \frac{x_T u_T}{\sum x_t^2}\right]^2 = \\
&= E\left[\frac{x_1^2 u_1^2}{(\sum x_t^2)^2} + \frac{x_2^2 u_2^2}{(\sum x_t^2)^2} + \ldots + \frac{x_T^2 u_T^2}{(\sum x_t^2)^2} + \frac{2\sum\sum x_t x_s u_t u_s}{(\sum x_t^2)^2}\right] = \\
&= \sum \left[E\left[\frac{x_t^2}{(\sum x_t^2)^2}\right] E(u_t^2)\right] = \\
&= \sigma_u^2 \left[\sum_{t=1}^T E\left[\frac{x_t^2}{(\sum x_t^2)^2}\right]\right]
\end{aligned}$$

y la esperanza debe existir.

5.2.2. Incorrelación contemporánea entre regresores y error

X, u no son independientes pero son incorreladas contemporáneamente $\quad Cov(X_{it}, u_t) = 0$, es decir, mantenemos que $E(X_{it} u_t) = 0 \quad i = 1, 2, \ldots, k$, pero no la independencia entre X_{it} y u_t.

En este caso no podemos derivar analíticamente las propiedades en muestras finitas del estimador:

a) Valor medio:

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

y en general será sesgado ya que $E[(X'X)^{-1} X'u]$ puede ser distinto de cero.

b) Matriz de varianzas y covarianzas:
Su cálculo es complicado dada la no linealidad del estimador en X y u.

c) No conocemos su distribución exacta, no siguen una distribución normal aún en el caso de que $X_{it} \quad \forall i \quad \forall t$ y u la sigan. Como consecuencia los estadísticos t y F no tienen distribución exacta conocida.

d) Las propiedades asintóticas de consistencia y distribución asintótica se mantienen ya que podemos aplicar los Teoremas de Mann y Wald, Cramer y Slutsky.

Ejercicio 1:
En el modelo:

$$Y_t = \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$u_t \sim N(0, \sigma_u^2)$
X_t es v.a tal que $E(X_t u_t) = 0$

pero X_t e u_t no son independientes.

Definimos el estimador MCO como :

$$\hat{\beta}_{MCO} = \frac{\sum X_t Y_t}{\sum X_t^2} = \beta + \frac{\sum X_t u_t}{\sum X_t^2}$$

Buscamos las propiedades en muestras finitas:

$$E(\hat{\beta}_{MCO}) = E\left[\beta + \frac{\sum X_t u_t}{\sum X_t^2}\right] = \beta + E\left[\frac{\sum X_t u_t}{\sum X_t^2}\right] \neq \beta$$

$E\left[\frac{\sum X_t u_t}{\sum X_t^2}\right]$ puede ser distinto de cero, por tanto el estimador MCO será sesgado en general. Dado que X_t y u_t no son independientes:

$$E\left[\frac{\sum X_t u_t}{\sum X_t^2}\right] \neq E\left[\frac{\sum X_t}{\sum X_t^2}\right] E(u_t)$$

además no podemos hacer

$$E\left[\frac{\sum X_t u_t}{\sum X_t^2}\right] = \frac{E[\sum X_t u_t]}{E[\sum X_t^2]}$$

con lo que no podemos derivar sus propiedades estadísticas en muestras pequeñas analíticamente. Sin embargo, si las propiedades habituales se cumplen podemos aplicar el Teorema de Mann y Wald y el estimador MCO será consistente, con distribución asintótica conocida y podremos hacer inferencia en el límite.

$$plim \hat{\beta}_{MCO} = \beta + plim\left(\frac{1}{T}\sum X_t^2\right)^{-1} plim\left(\frac{1}{T}\sum X_t u_t\right)$$

Por el Teorema de Mann y Wald, aplicado sus condiciones al ejercicio:

i) $u_1, u_2, \ldots, u_T$ v.a tal que $u_t \sim iid(0, \sigma_u^2)$

ii) $E(X_t u_t) = 0$ ya que X_t e u_t son incorreladas.

iii) $plim\left(\frac{1}{T}\sum X_t^2\right) = q_{XX}$

de i)+ii)+iii) obtenemos que:

$$\begin{array}{l} 1.plim\left(\frac{1}{T}\sum X_t u_t\right) = 0 \\ 2.\frac{1}{\sqrt{T}}\sum X_t u_t \xrightarrow{d} N(0, \sigma_u^2 q_{XX}) \end{array}$$

de i) $\Rightarrow plim\hat{\beta}_{MCO} = \beta$ por lo tanto es un estimador consistente de β.

de ii) $\Rightarrow$ $\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N(0, \frac{\sigma_u^2}{q_{XX}})$ y haremos inferencia del tipo $H_0 : R\beta = r$ con el siguiente estadístico asintótico:

$$\frac{(R\hat{\beta}_{MCO} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}_{MCO} - r)}{\hat{\sigma}_u^2} \xrightarrow{d} \chi^2_{(q)}$$

si sólo hay una hipótesis de contraste, es decir q=1 podemos realizar el contraste con el siguiente estadístico:

$$\frac{R\hat{\beta}_{MCO} - r}{\hat{\sigma}_u\sqrt{(R(X'X)^{-1}R')}} \xrightarrow{d} N(0,1)$$

Ejercicio 2:
Sea el modelo:

$$Y_t = \alpha + \beta Y_{t-1} + u_t \qquad t = 2, \ldots, T \qquad |\beta| < 1$$

donde:

$$E(u_t) = 0 \quad \forall t$$
$$E(u_t^2) = \sigma_u^2 \quad \forall t$$
$$E(u_t u_s) = 0 \quad \forall t, s \quad t \neq s$$

dado que Y_t es v.a. Y_{t-1} también lo es, luego la matriz de regresores es estocástica. Vamos a tratar de determinar las propiedades del estimador MCO del parámetro β. Para ello tenemos que analizar las relaciones entre Y_{t-1} y u_t.
Retardando el modelo obtenemos:

$$\begin{aligned} Y_t &= \alpha + \beta Y_{t-1} + u_t = \\ &= \alpha + \beta(\alpha + \beta Y_{t-2} + u_{t-1}) + u_t = \\ &= \alpha + \alpha\beta + \beta^2 Y_{t-2} + \beta u_{t-1} + u_t = \\ &= \alpha(1+\beta) + \beta^2(\alpha + \beta Y_{t-3} + u_{t-2}) + \beta u_{t-1} + u_t = \\ &= \alpha(1+\beta+\beta^2) + \beta^3 Y_{t-3} + \beta^2 u_{t-2} + \beta u_{t-1} + u_t = \\ &= \ldots\ldots\ldots\ldots = \\ &= \alpha(1+\beta+\beta^2+\ldots+\beta^{t-1}) + \beta^t Y_0 + u_t + \beta u_{t-1} + \beta^2 u_{t-2} + \ldots + \beta^{t-1} u_1 = \\ &= \alpha\textstyle\sum_{t=1}^{\infty}\beta^{t-1} + \beta^t Y_0 + \sum_{s=0}^{\infty}\beta^s u_{t-s} \end{aligned}$$

de donde:
Y_t depende de $Y_0, u_1, u_2, u_3, \ldots, u_{t-1}, u_t$ pero no depende de u_t por lo tanto podemos mantener que Y_{t-1} y u_t están incorreladas contemporáneamente $Cov(Y_{t-1}, u_t) = 0$ ya que $Y_{t-1} = X_t \Rightarrow Cov(X_t, u_t) = 0$ incorrelación contemporánea. Ahora vamos a probar que no son independientes:

$$y_{t-1} = Y_{t-1} - \overline{Y} = Y_{t-1} - \frac{1}{T}(Y_1 + Y_2 + \ldots + Y_t + Y_{t+1} + \ldots)$$

Y_t depende de u_t, como Y_{t-1} incorpora a Y_t y esta a su vez a u_t, no son independientes.
En este caso:

1) $Y_1, Y_2, \ldots, Y_T$ no serían v.a. independientes de $u_1, u_2, \ldots, u_T$.

2) Como $E(u_t u_s) = 0 \quad \forall t \neq s$ entonces $E(Y_{t-1}u_t) = 0 \quad \forall t$, el regresor y la perturbación están contemporáneamente incorrelados.

- Propiedades de $\hat{\beta}_{MCO}$ en muestras finitas:

$$\hat{\beta}_{MCO} = \beta + \frac{\sum_{t=2}^{T} y_{t-1}u_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

$$E(\hat{\beta}_{MCO}) = \beta + E\left(\frac{\sum_{t=2}^{T} y_{t-1}u_t}{\sum_{t=2}^{T} y_{t-1}^2}\right) \neq \beta$$

Notar que:

$$E\left(\frac{\sum_{t=2}^{T} y_{t-1}u_t}{\sum_{t=2}^{T} y_{t-1}^2}\right) \neq \frac{E\left[\sum_{t=2}^{T} y_{t-1}u_t\right]}{E\left[\sum_{t=2}^{T} y_{t-1}^2\right]}$$

ya que numerador y denominador no son independientes.

Sabemos que $E\left(\sum_{t=2}^{T} y_{t-1}u_t\right) = 0$ pero esto no implica que

$$E\left(\frac{\sum_{t=2}^{T} y_{t-1}u_t}{\sum_{t=2}^{T} y_{t-1}^2}\right) = 0$$

por tanto no podemos mantener la insesgadez y tampoco podemos hablar de eficiencia. De igual manera no conocemos su distribución exacta en muestras finitas.

- Propiedades asintóticas. Consistencia: Podemos demostrar que el estimador es consistente ya que se cumple el Teorema de Mann y Wald. Aplicamos las condiciones del teorema al ejercicio:

1. $u_t \sim iid(0, \sigma_u^2)$
2. $E(X_t u_t) = E(Y_{t-1}u_t) = 0$
3. Suponemos que $plim\left(\frac{1}{T}X'X\right) = Q^{-1}$ tal que $plim\left(\frac{1}{T}\sum y_{t-1}^2\right) = \kappa, \quad \kappa$ es un escalar

de [1]+[2]+[3] obtenemos que $plim\left(\frac{1}{T}X'u\right) = 0$ que aplicado al ejercicio, dado que $X = [\vec{1}; Y_{t-1}]$ implica:

1. $plim\left(\frac{1}{T}\sum u_t\right) = 0$
2. $plim\left(\frac{1}{T}\sum Y_{t-1}u_t\right) = 0 \Rightarrow plim\left(\frac{1}{T}\sum y_{t-1}u_t\right) = 0$

Por tanto el límite en probabilidad del estimador MCO de la pendiente del modelo es:

$$\begin{aligned} plim\hat{\beta}_{MCO} &= plim\beta + plim\left[\frac{\sum_{t=2}^{T} y_{t-1}u_t}{\sum_{t=2}^{T} y_{t-1}^2}\right] = \\ &= \beta + plim\left[\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2\right]^{-1} plim\left[\frac{1}{T}\sum_{t=2}^{T} y_{t-1}u_t\right] = \\ &= \beta + \kappa \cdot 0 \end{aligned}$$

- Distribución asintótica:
  El segundo resultado del Teorema de Mann y Wald es que $\frac{1}{\sqrt{T}} \sum y_{t-1}u_t \xrightarrow{d} N(0, \sigma_u^2 \kappa)$ por tanto la distribución asintótica del estimador MCO de la pendiente del modelo es:
  $$\sqrt{T}(\hat{\beta}_{MCO} - \beta) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\kappa}\right)$$
  y podemos hacer inferencia asintótica.

Conclusión: Cuando u_t cumple las hipótesis básicas está justificado el uso de los MCO en un modelo donde entre los regresores aparece la variable endógena retardada, independientemente del número de retardos, pero los resultados de MCO se cumplen sólo asintóticamente.

5.2.3. Correlación entre regresores y error

En muchas ocasiones la hipótesis $Cov(X_{it}, u_t) = 0$ no es válida. En este caso los estimadores MCO no son ni siquiera consistentes. Hay cuatro puntos a solucionar en relación a este tema:

a) ¿Cómo aparecen las correlaciones entre X y u ?

   I) En modelos con variable endógena retardada como regresor y perturbación autocorrelada.
   II) Cuando la variable exógena está medida con error.
   III) En modelos de ecuaciones simultáneas, por ejemplo en el modelo de oferta y demanda donde P y Q se determinan simultáneamente.
   IV) Si el modelo tiene un problema de omisión de variable relevante y la variable omitida está correlada con los regresores. Esta correlación aparecerá vía la perturbación ya que la perturbación recoge las variables omitidas.

b) ¿Qué importancia puede tener este problema? En realidad depende del caso concreto, pero de forma general podemos decir que se pierden las propiedades en muestras pequeñas y grandes.

c) ¿Cómo podemos detectar que existe el problema? Usando test de contraste que sean capaces de detectar la correlación entre X y u .

d) ¿Cómo podemos solucionar el problema? Si la existencia de correlación entre regresores y perturbación se debe a un problema de omisión de variable relevante debemos especificar correctamente el modelo. En el resto de casos tendremos que buscar un método de estimación alternativo a MCO que nos produzca buenas propiedades, aunque éstas se logren sólo en muestras grandes.

Algunos ejemplos.

Ejemplo 1: Omisión de variable relevante.
Sea el modelo correctamente especificado:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad t = 1, 2, \ldots, T$$

donde

$$E(X_{2t}v_t) = 0$$
$$E(X_{3t}v_t) = 0$$
$$v_t \sim iid(0, \sigma_v^2)$$

pero el modelo que se estima es el siguiente:

$$Y_t = \beta_1 + \beta_2 X_{2t} + u_t \qquad t = 1, 2, \ldots, T$$

tal que $u_t = \beta_3 X_{3t} + v_t$

En el modelo que finalmente se estima tendremos:

$$E(X_{2t}u_t) = E[X_{2t}(\beta_3 X_{3t} + v_t)] = \beta_3 E(X_{2t}X_{3t}) + E(X_{2t}v_t) = \beta_3 E(X_{2t}X_{3t})$$

$$E(u_t) = \beta_3 E(X_{3t}) + E(v_t) = \beta_3 E(X_{3t})$$

luego:

$$E(X_{2t}u_t) \neq 0 \quad \text{si} \quad \beta_3 \neq 0 \quad \text{y} \quad E(X_{2t}X_{3t}) \neq 0$$

$$E(u_t) \neq 0 \quad \text{si} \quad \beta_3 \neq 0 \quad \text{y} \quad E(X_{3t}) \neq 0$$

Por tanto X y u son independientes y $E(u/X) = E(u) = 0$ si X_{2t} y X_{3t} son variables aleatorias independientes y $E(X_{3t}) = 0 \quad \forall t$.

Ejemplo 2: Simultaneidad.
Sea el modelo formado por las siguientes dos ecuaciones:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad t = 1, 2, \ldots, T \tag{5.1}$$

$$X_t = \gamma_1 + \gamma_2 Y_t + v_t \qquad t = 1, 2, \ldots, T \tag{5.2}$$

tal que:

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$$

Es decir u_t y v_t son v.a normales e independientemente distribuidas en el tiempo. Estamos interesados en estimar β_1 y β_2 en (5.1), para ello queremos saber si X y u son independientes y/o incorreladas.

$$\begin{array}{l} E(u_t) = 0 \quad \forall t \\ E(X_t u_t) = E[(\gamma_1 + \gamma_2(\beta_1 + \beta_2 X_t + u_t) + v_t)u_t] = \\ \gamma_1 E(u_t) + \gamma_2\beta_1 E(u_t) + \gamma_2\beta_2 E(X_t u_t) + \gamma_2 E(u_t^2) + E(v_t u_t) = \gamma_2\beta_2 E(X_t u_t) + \gamma_2\sigma_u^2 + \sigma_{uv} \end{array}$$

resolviendo:

$$E(X_t u_t) = \frac{1}{1 - \gamma_2\beta_2}(\sigma_{uv} + \gamma_2\sigma_u^2)$$

con lo que $E(X_t u_t) \neq 0$ si $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$

5.3. Método de Variables Instrumentales

En modelos donde existe correlación entre regresores y error como por ejemplo:

$$\begin{array}{ll} Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t & t = 1, 2, \ldots, T \\ u_t \sim iid(0, \sigma_u^2) & \\ X_{2t} \quad \text{variable fija} & \\ X_{3t} \quad \text{variable aleatoria tal que} \quad E(X_{3t}u_t) \neq 0 & \end{array}$$

O también:

$$
\begin{aligned}
Y_t &= \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_t + u_t \quad t = 1, 2, \ldots, T \\
u_t &= \rho u_{t-1} + \epsilon_t \\
\epsilon_t &\sim iid(0, \sigma^2_\epsilon) \\
|\rho| &< 1
\end{aligned}
$$

el estimador MCO es inconsistente y sesgado ya que $E(X'u) \neq 0$. El procedimiento para obtener estimadores consistentes en un modelo de este tipo es el Método de Variables Instrumentales. Una variable instrumental es una variable Z_t que satisface tres condiciones:

1. No está incluida en el modelo como variable explicativa.
2. Está incorrelacionada con el término de error, $E(Z_t u_t) = 0$.
3. Está correlacionada con la variable para la cual hace de instrumento.

Cualquiera de los dos modelos anteriores podemos escribirlos de la forma matricial habitual $Y = X\beta + u$ y sus ecuaciones normales son:

$$X'Y = X'X\hat{\beta}$$

donde lo que hacemos es premultiplicar el modelo por X'. Si $plim\left(\frac{1}{T}X'u\right) = 0$ los estimadores así obtenidos son consistentes. Por tanto parece que sugiere que cuando $plim\left(\frac{1}{T}X'u\right) \neq 0$ en vez de premultiplicar por X' lo hagamos por Z' tal que $plim\left(\frac{1}{T}Z'u\right) = 0$ y llamamos a Z matriz de instrumentos. Al estimador así obtenido se le conoce por el estimador de Variables Instrumentales y se define:

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

Demostración:

$$
\begin{aligned}
&Y = X\beta + u \\
&Z'Y = Z'X\beta + Z'u
\end{aligned}
$$

Denotamos por $u_\star = Z'u$

$$
\begin{aligned}
&u_\star = Z'u = Z'(Y - X\beta) \\
&\hat{u}_\star = Z'\hat{u} = Z'(Y - X\hat{\beta}) \\
&\hat{u}'_\star \hat{u}_\star = \left[Z'(Y - X\hat{\beta})\right]' \left[Z'(Y - X\hat{\beta})\right] = (Y - X\hat{\beta})'ZZ'(Y - X\hat{\beta})
\end{aligned}
$$

La función objetivo podemos escribirla como:

$$Min_{\hat{\beta}} \hat{u}'_\star \hat{u}_\star = Min_{\hat{\beta}} \left[Y'ZZ'Y - 2\hat{\beta}'X'ZZ'Y + \hat{\beta}'X'ZZ'X\hat{\beta}\right]$$

por las condiciones de primer orden obtenemos:

$$
\begin{aligned}
\frac{\partial(\hat{u}'_\star \hat{u}_\star)}{\partial\hat{\beta}} &= -2X'ZZ'Y + 2X'ZZ'X\hat{\beta} = 0 \\
&= X'ZZ'(Y - X\hat{\beta}) = 0
\end{aligned}
$$

De donde las ecuaciones normales del modelo serían: $X'ZZ'X\hat{\beta} = X'ZZ'Y$ tal que:

$$
\begin{aligned}
\hat{\beta}_{VI} &= (X'ZZ'X)^{-1}(X'ZZ'Y) \\
\hat{\beta}_{VI} &= (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'Y) \\
\hat{\beta}_{VI} &= (Z'X)^{-1}(Z'Y)
\end{aligned}
$$

5.3.1. Propiedades del estimador de Variables Instrumentales

1. Linealidad:
   $\hat{\beta}_{VI} = \beta + (Z'X)^{-1}(Z'u)$ **no lineal**

1. Insesgadez:
   $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$
   $E(\hat{\beta}_{VI}) = \beta + E[(Z'X)^{-1}(Z'u)] \neq 0$
   en general el estimador será **sesgado** ya que $E[(Z'X)^{-1}(Z'u)] \neq 0$. El requisito impuesto para obtener el estimador de Variables Instrumentales $\hat{\beta}_{VI}$ es $plim\left(\frac{1}{T}Z'u\right) = 0$ y no la independencia entre Z y u necesaria para que el estimador sea insesgado.

2. Consistencia:
   Suponiendo que $plim\left(\frac{1}{T}Z'X\right) = Q_{ZX}$ y dado que buscamos los instrumentos tal que $plim\left(\frac{1}{T}Z'u\right) = 0$, $(E(Z_t u_t) = 0)$, el estimador de Variables Instrumentales será consistente. En este contexto hay que tener cuidado con la aplicación del Teorema de Mann y Wald. El estimador de Variables Instrumentales es siempre consistente y su consistencia proviene de la ausencia de correlación entre los instrumentos y el término de error, con independencia de que éste tenga o no autocorrelación. Así el límite en probabilidad del estimador es:

$$plim\hat{\beta}_{VI} = \beta + plim\left(\frac{1}{T}Z'X\right)^{-1} plim\left(\frac{1}{T}Z'u\right) = \beta + Q_{zx}^{-1} \cdot 0 = \beta$$

3. Distribución asintótica:
   Con respecto a la distribución asintótica de $\hat{\beta}_{VI}$, sólo la podremos caracterizar si el término de perturbación no está autocorrelado.

   En ausencia de autocorrelación en las perturbaciones y suponiendo:

   i) $E(Z'u) = 0$
   ii) $plim\left(\frac{1}{T}Z'Z\right) = Q_{ZZ}$ finita, simétrica y definida positiva.
   iii) $plim\left(\frac{1}{T}Z'X\right) = Q_{ZX}$ finita, no singular

   entonces el estimador $\hat{\beta}_{VI}$ es consistente y se tiene que por el Teorema de Mann y Wald:

   a) $plim\left(\frac{1}{T}Z'u\right) = 0$
   b) $\frac{1}{\sqrt{T}}Z'u \xrightarrow{d} N(0, \sigma_u^2 Q_{ZZ})$

   de donde aplicando el Teorema de Cramer tendremos:

$$\sqrt{T}(\hat{\beta}_{VI} - \beta) = \left(\frac{1}{T}Z'X\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'u\right) \xrightarrow{d} N(0, \sigma_u^2 \cdot Q_{ZX}^{-1} \cdot Q_{ZZ} \cdot (Q_{ZX}^{-1})')$$

El resultado anterior justifica que en muestras grandes se utilice como matriz de covarianzas asintótica del estimador de variables instrumentales a:

$$\sum_{(\hat{\beta}_{VI})} = \frac{\sigma_u^2}{T}(Q_{ZX}^{-1} Q_{ZZ} (Q_{ZX}^{-1})')$$

y si se utilizan las matrices de momentos muestrales $\frac{Z'X}{T}$ y $\frac{Z'Z}{T}$ para aproximar sus límites respectivos Q_{ZX} y Q_{ZZ} entonces se tiene como matriz de covarianzas asintótica a:

$$\begin{aligned}\sum\nolimits_{(\hat{\beta}_{VI})} &= \tfrac{\sigma_u^2}{T}\left(\tfrac{Z'X}{T}\right)^{-1}\tfrac{Z'Z}{T}\left[\left(\tfrac{X'Z}{T}\right)^{-1}\right]' = \\ &= \sigma_u^2(Z'X)^{-1}Z'Z((Z'X)^{-1})'\end{aligned}$$

Un estimador consistente de la misma sería:

$$\sum\nolimits_{(\hat{\beta}_{VI})} = \hat{\sigma}^2_{VI}(Z'X)^{-1}Z'Z((Z'X)^{-1})'$$

donde:

$$\hat{\sigma}^2_{VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-k}$$

es un estimador consistente de σ_u^2. El resultado anterior no puede generalizarse fácilmente al caso en el que la perturbación tiene autocorrelación. Modelos de este tipo son estimados por máximaverosimilitud.

5.3.2. Cómo buscar los instrumentos

Con respecto a los instrumentos sabemos que éstos tienen que cumplir tres condiciones:

1. No estar incluidos como regresores en el modelo de interés.
2. Estar incorrelacionados con la perturbación del modelo de interés.
3. Estar correlacionados con la variable para la cuál hacen de instrumento.

En la práctica existen dos situaciones diferentes en la búsqueda de instrumentos:

a) Que el número de instrumentos disponibles coincida con el número de variables que necesiten instrumento.

b) Que el número de instrumentos disponibles sea mayor que el número de variables que necesiten instrumento.

Número de instrumentos igual al número de variables explicativas que lo necesitan

Supongamos que el número de instrumentos de que se dispone es igual al número de variables explicativas que necesitan instrumento. En este caso cada instrumento constituye una variable instrumental para sustituir a su correspondiente variable exógena correlada con la perturbación. En general en la matriz de regresores X sólo habrá unas variables que no satisfagan la condición $E(X_{it}u_t) = 0$ y son estas variables las que necesitan de variables instrumentales. Es decir, las matrices Z y X tendrán en común aquellas columnas correspondientes a las variables incorrelacionadas con el término de error. Notar que necesariamente Z y X deben tener el mismo número de columnas ya que en otro caso $(Z'X)$ no sería cuadrada.

Los instrumentos deben cumplir tres requisitos, mencionados anteriormente, no deben estar incluidos como variables explicativas en el modelo original, no deben estar correlados con la perturbación y deben de estar correlados con la variable exógena para la que actúan de instrumento. En cuanto a esta correlación, debe existir pero no puede ser muy importante pues en este caso, si lo fuera, $E(Z_t u_t) \neq 0$ y Z_t no serviría de instrumento. A continuación buscaremos algunos ejemplos.

Ejemplo 1:
Sea el modelo:

$$Y_t = \alpha + \beta X_t + u_t \quad u_t \sim iid(0, \sigma_u^2) \quad t = 1, 2, \ldots, T$$

$$\begin{array}{ll} \text{tal que} & X_t = 0,7X_{t-1} + v_t \quad v_t \sim iid(0, \sigma_v^2) \\ & E(u_t v_s) = 5 \quad \text{si} \quad t = s \\ & E(u_t v_s) = 0 \quad \text{si} \quad t \neq s \end{array}$$

En este caso el regresor X_t es un regresor estocástico correlado con la perturbación:

$$E(X_t u_t) = E((0,7X_{t-1} + v_t)u_t) = 0,7E(X_{t-1}u_t) + E(v_t u_t) = 5$$

El estimador MCO de los parámetros del modelo es inconsistente y debemos estimar por el Método de Variables Instrumentales. Para ello buscamos un instrumento para X_t, podemos pensar en $Z_t = X_{t-1}$ ya que X_{t-1} no es un regresor del modelo, está incorrelado con la perturbación, $E(X_{t-1}u_t) = 0$ y correlado con X_t ya que ésta se genera por un proceso autorregresivo, $E(X_t X_{t-j}) \neq 0 \quad \forall j > 0$.

Aplicamos el estimador de Variables Instrumentales para $Z_t = X_{t-1}$:

$$Y = \begin{pmatrix} Y_2 \\ Y_3 \\ \cdots \\ Y_T \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_2 \\ 1 & X_3 \\ \cdots & \cdots \\ 1 & X_T \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & X_1 \\ 1 & X_3 \\ \cdots & \cdots \\ 1 & X_{T-1} \end{pmatrix}$$

Y el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ sería el siguiente:

$$\begin{pmatrix} \hat{\alpha}_{VI} \\ \hat{\beta}_{VI} \end{pmatrix} = \begin{pmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_2^T Y_t \\ \sum_2^T X_{t-1} Y_t \end{pmatrix}$$

Su matriz de varianzas y covarianzas estimada se define:

$$\sum\nolimits_{(\hat{\beta}_{VI})} = \hat{\sigma}^2_{u,VI}(Z'X)^{-1}(Z'Z)((Z'X)^{-1})'$$

$$\sum_{(\hat{\beta}_{VI})} = \hat{\sigma}^2_{u,VI} \begin{pmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} (T-1) & \sum_2^T X_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}^2 \end{pmatrix} \begin{pmatrix} (T-1) & \sum_2^T X_t \\ \sum_2^T X_{t-1} & \sum_2^T X_t X_{t-1} \end{pmatrix}^{-1}$$

siendo:

$$\hat{\sigma}^2_{u,VI} = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-k}$$

Ejemplo 2:
Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t \qquad t = 1, 2, \ldots, T$$

$$u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \quad |\rho| < 1$$

$$X_t \quad \text{no estocástica}$$

en este caso Y_{t-1} depende de u_{t-1} y éste se relaciona con u_t vía el proceso AR(1) por tanto $E(Y_{t-1}u_t) \neq 0$ y el estimador MCO de los parámetros es inconsistente. En este caso si nuestro objetivo es encontrar estimadores consistentes podemos estimar por Variables Instrumentales.

Sin embargo cómo la perturbación está autocorrelada no podremos encontrar su distribución asintótica para hacer inferencia asintótica con este estimador. Más adelante propondremos un estimador alternativo al de Variables Instrumentales basado en la estimación del modelo transformado. Sin embargo ahora vamos a implementar el estimador de VI cómo una práctica más. Necesitamos instrumentalizar Y_{t-1}, un instrumento válido es X_{t-1} ya que:

a) No aparece como regresor en el modelo.

b) Al ser una variable fija está incorrelada con la perturbación, $E(X_{t-1}u_t) = 0$

c) X_{t-1} influye en Y_{t-1} a través del propio modelo por tanto $E(X_{t-1}X_t) \neq 0$ ya que

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \beta_3 Y_{t-2} + u_{t-1}$$

En este caso las matrices Y, X y Z serían:

$$Y = \begin{pmatrix} Y_2 \\ Y_3 \\ \dots \\ Y_T \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_2 & Y_1 \\ 1 & X_3 & Y_2 \\ \dots & \dots & \dots \\ 1 & X_T & Y_{T-1} \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & X_2 & X_1 \\ 1 & X_3 & X_2 \\ \dots & \dots & \dots \\ 1 & X_T & X_{T-1} \end{pmatrix}$$

Así el estimador de variables instrumentales sería:

$$\begin{pmatrix} \hat\beta_{1,VI} \\ \hat\beta_{2,VI} \\ \hat\beta_{3,VI} \end{pmatrix} = \begin{pmatrix} (T-1) & \sum_2^T X_t & \sum_2^T Y_{t-1} \\ \sum_2^T X_t & \sum_2^T X_t^2 & \sum_2^T X_t Y_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}X_t & \sum_2^T X_{t-1}Y_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_2^T Y_t \\ \sum_2^T X_t Y_t \\ \sum_2^T X_{t-1} Y_t \end{pmatrix}$$

Y su matriz de varianzas y covarianzas estimada sería:

$$\sum(\hat\beta_{VI}) = \hat\sigma^2_{u,VI} \begin{pmatrix} (T-1) & \sum_2^T X_t & \sum_2^T Y_{t-1} \\ \sum_2^T X_t & \sum_2^T X_t^2 & \sum_2^T X_t Y_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}X_t & \sum_2^T X_{t-1}Y_{t-1} \end{pmatrix}^{-1} \bullet$$

$$\begin{pmatrix} (T-1) & \sum_2^T X_t & \sum_2^T X_{t-1} \\ \sum_2^T X_t & \sum_2^T X_t^2 & \sum_2^T X_t X_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}X_t & \sum_2^T X_{t-1}^2 \end{pmatrix} \begin{pmatrix} (T-1) & \sum_2^T X_t & \sum_2^T Y_{t-1} \\ \sum_2^T X_t & \sum_2^T X_t^2 & \sum_2^T X_t Y_{t-1} \\ \sum_2^T X_{t-1} & \sum_2^T X_{t-1}X_t & \sum_2^T X_{t-1}Y_{t-1} \end{pmatrix}^{-1}$$

siendo:

$$\hat\sigma^2_{u,VI} = \frac{(Y - X\hat\beta_{VI})'(Y - X\hat\beta_{VI})}{T-k}$$

Ejemplo 3:
Sea el modelo:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t \qquad t = 1, 2, \dots, T$$

$$u_t = \epsilon_t + \theta\epsilon_{t-1}$$

En este caso u_t se relaciona con u_{t-1} pero no con $u_{t-2}, u_{t-3}, \dots$ etc. ya que u_t sigue un proceso MA(1). El instrumento adecuado para Y_{t-1} sería su propio retardo ya que Y_{t-1} se relaciona con u_{t-1} y éste a su vez con u_t, de donde Y_{t-2} se relaciona con u_{t-2} y u_{t-3} pero no con u_{t-1}, por tanto tampoco con u_t, es decir $E(Y_{t-1}u_t) \neq 0$ pero $E(Y_{t-2}u_t) = 0$. Para este modelo las matrices de datos e instrumentos serían:

$$Y = \begin{pmatrix} Y_3 \\ Y_4 \\ \ldots \\ Y_T \end{pmatrix} \qquad X = \begin{pmatrix} 1 & Y_2 \\ 1 & Y_3 \\ \ldots & \ldots \\ 1 & Y_{T-1} \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & Y_1 \\ 1 & Y_2 \\ \ldots & \ldots \\ 1 & Y_{T-2} \end{pmatrix}$$

Y el estimador $\hat{\beta}_{VI} = (Z'X)^{-1}(Z'Y)$ sería el siguiente:

$$\begin{pmatrix} \hat{\beta}_{1,VI} \\ \hat{\beta}_{2,VI} \end{pmatrix} = \begin{pmatrix} (T-2) & \sum_3^T Y_{t-1} \\ \sum_3^T Y_{t-2} & \sum_3^T Y_{t-2}Y_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_3^T Y_t \\ \sum_3^T Y_{t-2}Y_t \end{pmatrix}$$

En este caso el estimador de VI no tiene una distribución asintótica conocida y deberíamos estimar el modelo por otro método que nos permita hacer inferencia, método que más adelante veremos.

Método de Mínimos Cuadrados en dos etapas

Hasta ahora hemos desarrollado el estimador de variables instrumentales para el caso en que el número de instrumentos sea igual al número de variables que lo necesitan. Generalmente se dispondrá de un número mayor de instrumentos que de variables explicativas a sustituir, en este caso habría muchas formas de construir las variables instrumentales que precisamos para obtener consistencia. Pero dado que la matriz de covarianzas del estimador de variables instrumentales depende de los valores de éstas, el modo en que se combinan los instrumentos para generar variables instrumentales influye sobre la eficiencia del estimador de variables instrumentales respecto a otro estimador de variables instrumentales de su misma clase. De ahí que en ocasiones se hable de la **eficiencia relativa** de los estimadores de variables instrumentales.

Ejemplo 1:
En el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$$\begin{array}{ll} u_t \sim iid(0, \sigma_u^2) & \\ X_{3t}, X_{4t} \quad \text{variables fijas} & \\ X_{2t} = 0,5X_{2,t-1} + v_t & v_t \sim iid(0, \sigma_v^2) \\ E(u_t v_t) = 5 \quad \forall t = s & E(u_t v_s) = 0 \quad \forall t \neq s \end{array}$$

En este caso:

$$E(X_{2t}u_t) = E[(0,5X_{2,t-1} + v_t)u_t] = 0,5E(X_{2,t-1}u_t) + E(v_t u_t) = 0,5 \cdot 0 + 5 = 5$$

En este ejercicio la variable explicativa X_{2t} y la perturbación están contemporáneamente correladas, por tanto $E(X'u) \neq 0$ donde $X = [\vec{1}; X_{2t}; X_{3t}; X_{4t}]$, por lo tanto el estimador MCO será no lineal y sesgado. En muestras grandes además será inconsistente. Deberíamos estimar por el Método de Variables Instrumentales, buscando un instrumento para X_{2t}.

En este caso hay varios instrumentos que cumplen los requisitos, por ejemplo:

$$E(X_{2,t-1}u_t) = 0$$

$$E(X_{3,t-1}u_t) = 0$$

$$E(X_{4,t-1}u_t) = 0$$

Las variables $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ además de estar correladas con X_{2t} no son regresores del modelo por tanto $X_{2,t-1}$, $X_{3,t-1}$ y $X_{4,t-1}$ serían buenos instrumentos. También lo serían combinaciones lineales de los mismos. Dado que en esta situación el número de instrumentos supera al de variables que lo precisan se trata de buscar cuál de todos los posibles instrumentos minimiza la varianza del estimador resultante. Puesto que la matriz de varianzas y covarianzas de $\hat{\beta}_{VI}$ depende de la matriz de instrumentos, Z, el modo en que los instrumentos se combinan para generar variables instrumentales influye sobre la eficiencia del estimador de VI respecto a otro estimador de su misma clase. De ahí que se hable en ocasiones de la eficiencia relativa del estimador de VI. Una posibilidad consiste en generar la variable instrumental con mayor correlación con Y. Para ello se estima por MCO una regresión auxiliar de esta variable sobre todos los posibles instrumentos. Para el ejemplo que nos ocupa, si suponemos que los únicos instrumentos de que disponemos son los primeros retardos de los regresores originales la regresión auxiliar sería:

$$X_{2t} = \gamma_1 + \gamma_2 X_{2,t-1} + \gamma_3 X_{3,t-1} + \gamma_4 X_{4,t-1} + \eta_t \quad t = 2, 3, \ldots, T$$

así $\widehat{X_{2t}}$ es una combinación lineal de todos los posibles instrumentos ponderado cada uno por su correlación con X_{2t}, la variable a instrumentalizar. A continuación se reestima el modelo por VI con:

$$Z = [\vec{1}; \widehat{X_{2t}}; X_{3t}; X_{4t}]$$

$$X = [\vec{1}; X_{2t}; X_{3t}; X_{4t}]$$

Al estimador así obtenido se le llama estimador de Mínimos Cuadrados en dos Etapas, MC2E. (Nota: En este ejemplo perdemos una observación al construir $\widehat{X_{2t}} \quad t = 2, 3, \ldots, T$). Así:

$$Z = \begin{pmatrix} 1 & \widehat{X_{22}} & X_{32} & X_{42} \\ 1 & \widehat{X_{23}} & X_{33} & X_{43} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \widehat{X_{2T}} & X_{3T} & X_{4T} \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_{22} & X_{32} & X_{42} \\ 1 & X_{23} & X_{33} & X_{43} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2T} & X_{3T} & X_{4T} \end{pmatrix}$$

Ejemplo 2:
En el modelo:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 X_{1t} + \beta_4 X_{2t} + \beta_5 X_{3t} + u_t \qquad t = 2, 3, \ldots, T$$

donde:

$$u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) \qquad |\rho| < 1$$

$$X_{1t}, X_{2t}, X_{3t} \quad \text{variables determinísticas}$$

Las variables X_{1t}, X_{2t} y X_{3t} al ser consideradas determinísticas están incorrelacionadas con el término de error, de igual manera lo están sus retardos y éstos son instrumentos válidos para Y_{t-1}. Es un caso en que el número de instrumentos es superior al de variables explicativas a sustituir. Además cualquier combinación lineal de estos retardos sería asímismo un instrumento válido. Se trataría, por tanto, de buscar cuál de todas las posibles variables instrumentales minimiza la

Un posibilidad consiste en generar la variable instrumental con mayor correlación con Y_{t-1}. Para ello, se estima una regresión auxiliar de ésta variable sobre los instrumentos de que disponemos. Si suponemos que los únicos instrumentos disponibles son $X_{1,t-1}$, $X_{2,t-1}$ y $X_{3,t-1}$, la regresión auxiliar para instrumentalizar a la variable Y_{t-1} sería:

$$Y_{t-1} = \gamma_0 + \gamma_1 X_{1,t-1} + \gamma_2 X_{2,t-1} + \gamma_3 X_{3,t-1} + \eta_t \qquad t = 2, 3, \ldots T$$

así $\widehat{Y_{t-1}}$ será una combinación lineal de $X_{1,t-1}$, $X_{2,t-1}$ y $X_{3,t-1}$ y como tal una variable instrumental válida. A continuación se utiliza $\widehat{Y_{t-1}}$ como instrumento para Y_{t-1} y se reestima el modelo con:

$$Z = [\vec{1}; \widehat{Y_{t-1}}; X_{1t}; X_{2t}; X_{3t}]$$

$$X = [\vec{1}; Y_{t-1}; X_{1t}; X_{2t}; X_{3t}]$$

en $\hat{\beta}_{VI} = (Z'X)^{-1}Z'Y$ generando el estimador de MC2E.

Ejemplo 3:
Sea el modelo:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_{1t} + u_t \qquad t = 1, 2, \ldots, T$$

$$X_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 Z_{2t} + \gamma_4 Z_{3t} + v_t$$

$$u_t \sim NID(0, \sigma_u^2)$$

$$v_t \sim NID(0, \sigma_v^2)$$

$$Cov(u_t, v_t) = \sigma_{uv}$$

dado que $\gamma_2 \neq 0$ y/o $\sigma_{uv} \neq 0$ existe correlación entre X_t y u_t, $E(X_t u_t) \neq 0$. El estimador adecuado es el estimador de Variables Instrumentales. Necesitamos buscar un instrumento para X_t y tenemos dos disponibles Z_{2t} y Z_{3t}. Para combinarlos de forma óptima podemos realizar la siguiente regresión:

$$X_t = \alpha_1 + \alpha_1 Z_{2t} + \alpha_3 Z_{3t} + \eta_t \qquad t = 1, 2, \ldots, T$$

Así el instrumento adecuado para X_t es $\hat{X}_t$ obtenido de la estimación por MCO de la regresión anterior. Utilizamos el instrumento para fijar la matriz de instrumentos Z. Así:

$$Z = \begin{pmatrix} 1 & \hat{X}_1 & Z_{11} \\ 1 & \hat{X}_2 & Z_{12} \\ \ldots & \ldots & \ldots \\ 1 & \hat{X}_T & Z_{1T} \end{pmatrix}$$

Y aplicamos el estimador de Variables Instrumentales para el cuál:

$$X = \begin{pmatrix} 1 & X_1 & Z_{11} \\ 1 & X_2 & Z_{12} \\ \ldots & \ldots & \ldots \\ 1 & X_T & Z_{1T} \end{pmatrix} \qquad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_T \end{pmatrix}$$

En general la regresión de **todas** las variables explicativas sobre **todos** los posibles instrumentos, recogidos en la matriz W, produce los coeficientes $(W'W)^{-1}W'X$ y genera el vector de variables explicadas $\hat{X} = W(W'W)^{-1}W'X$ que utilizadas como variables instrumentales, $Z = \hat{X}$, conducen finalmente al estimador de mínimos cuadrados en dos etapas:

Sea:

$$Y = X\beta + u$$

siendo:
X la matriz de variables explicativas.

Y la matriz de variables a explicar.
W la matriz de posibles instrumentos.
Regresamos X/W:

$$\begin{array}{l} X = W\delta + u \\ \hat{\delta} = (W'W)^{-1}W'X \\ \hat{X} = W\hat{\delta} = W(W'W)^{-1}W'X \end{array}$$

si utilizamos $\hat{X}$ como matriz de variables instrumentales, es decir $Z = \hat{X}$ tendremos que:

$$\begin{array}{ll} \hat{\beta}_{VI} & = (Z'X)^{-1}Z'Y = (\hat{X}'X)^{-1}\hat{X}'Y = \\ & = [X'W(W'W)^{-1}W'X]^{-1}[X'W(W'W)^{-1}W'Y] = \hat{\beta}_{MC2E} \end{array}$$

que sería el estimador de Mínimos Cuadrados en dos Etapas. Su matriz de varianzas y covarianzas sería:

$$\begin{array}{ll} \sum_{\hat{\beta}_{MC2E}} & = \sigma_u^2(Z'X)^{-1}Z'Z((X'Z)^{-1})' = \\ & = \sigma_u^2(\hat{X}'X)^{-1}\hat{X}'\hat{X}((X'\hat{X})^{-1})' = \sigma_u^2(\hat{X}'\hat{X})^{-1} \\ & = \sigma_u^2[X'W(W'W)^{-1}W'X]^{-1} \end{array}$$

y

$$\hat{\sigma}_u^2 = \frac{SRC_{VI}}{T-k}$$

Puede probarse que el estimador de MC2E es el estimador lineal de variables instrumentales eficiente, en el sentido de tener mínima matriz de covarianzas entre los estimadores que utilizan como variables instrumentales combinaciones lineales de los instrumentos disponibles. Sobre este estimador podemos hacer dos observaciones:

- Hay que notar que en el estimador de MC2E se regresan **todas** las variables explicativas sobre los posibles instrumentos, es decir, se parte de la idea de que si $E(X_t u_t) = 0$ el mejor instrumento para X_t es ella misma.

- La inclusión de retardos de las variables exógenas como instrumentos aumentaría el conjunto de información utilizado en la construcción del estimador de MC2E. Así, hay un estimador de MC2E para cada conjunto de instrumentos que se considere. Al utilizar más información, el estimador MC2E resultante sería más eficiente que otro estimador similar que utilizase menos información; sin embargo, el uso de retardos obliga a prescindir de algunas observaciones muestrales, lo que disminuye algo la eficiencia.

5.3.3. Contraste de hipótesis con el estimador de MC2E

Para hacer contraste de restricciones lineales con el estimador de MC2E podemos utilizar el siguiente estadístico:
$H_0 : R\beta = r$
$H_a : R\beta \neq r$

$$\frac{(R\hat{\beta}_{MC2E} - r)'[R(\hat{X}'\hat{X}R']^{-1}(R\hat{\beta}_{MC2E} - r)}{\hat{\sigma}_u^2} \xrightarrow{d} \chi^2_{(q)}$$

donde q es el número de restricciones que se contrastan y $\hat{\sigma}_u^2$ es el estimador obtenido desde $\hat{u}_{MC2E} = Y - X\hat{\beta}_{MC2E}$

- Estadístico de diferencias en las sumas residuales de cuadrados:
  Por paralelismo con el estimador de MCO el cual minimiza la expresión siguiente: $S_{MCO} = (Y - X\beta)'(Y - X\beta)$, el estimador de MC2E es aquel que minimiza la expresión

  $$S_{MC2E} = (Y - X\beta)'W(W'W)^{-1}W'(Y - X\beta)$$

  cuyas condiciones de primer orden son:

  $$X'W(W'W)^{-1}W'(Y - X\beta) = 0_k$$

  que coinciden con las ecuaciones normales del estimador $\hat{\beta}_{MC2E}$. Si minimizamos ahora S_{MC2E} sujeto a restricciones de la forma $R\beta = r$ obtenemos el siguiente estimador restringido:

  $$\hat{\beta}^r_{MC2E} = \hat{\beta}_{MC2E} - (\hat{X}'\hat{X})^{-1}R'(R(\hat{X}'\hat{X})^{-1}R')^{-1}(R\hat{\beta}_{MC2E} - r)$$

  de donde:

  $$\hat{u}^r_{MC2E} = Y - X\hat{\beta}^r_{MC2E} = \hat{u}_{MC2E} + X(\hat{X}'\hat{X})^{-1}R'[R(\hat{X}'\hat{X})^{-1}R']^{-1}(R\hat{\beta}_{MC2E} - r)$$

  de donde el estadístico de sumas de cuadrados sería:

  $$\frac{SRC(\hat{\beta}^r_{MC2E}) - SRC(\hat{\beta}_{MC2E})}{\hat{\sigma}^2_u} \xrightarrow{d} \chi^2_{(q)}$$

y las sumas de cuadrados restringida y sin restringir se obtendrían utilizando los residuos de las estimaciones restringida y sin restringir. Así:

$$SRC(\hat{\beta}^r_{MC2E}) - SRC(\hat{\beta}_{MC2E}) = \hat{u}^r_{MC2E}W(W'W)^{-1}W\hat{u}^r_{MC2E} - \hat{u}_{MC2E}W(W'W)^{-1}W\hat{u}_{MC2E}$$

5.3.4. Contraste de Sargan de validez de instrumentos

Dado que los instrumentos los elige en cierta manera el investigador y la elección es en cierto modo subjetiva resulta de utilidad disponer de un contraste para la validez de estos instrumentos. Sargan mostró que el estadístico:

$$\frac{S_{MC2E}}{\hat{\sigma}^2_u} \xrightarrow{d} \chi^2_{(p-k)}$$

sirve para contrastar la validez de los instrumentos utilizados siendo:

i) $S_{MC2E} = \hat{u}_{VI}W(W'W)^{-1}W'\hat{u}_{VI}$ valor que puede calcularse como la suma explicada en una regresión de los residuos de las variables instrumentales $\hat{u}_{VI}$ sobre el vector de variables W, calculándose $\hat{u}_{VI}$ con las variables del modelo original.

ii) p es el número total de instrumentos utilizados.

iii) k es el número de variables explicativas del modelo original.

Si el valor del estadístico calculado es mayor que el de la distribución $\chi^2_{(p-k)}$ para un α dado se acepta que el modelo está mal especificado o bien que no todos los instrumentos utilizados son válidos, es decir alguno-s están correlacionados con el término de perturbación.

5.3.5. Perturbación heterocedástica

¿Qué ocurre si la varianza de la perturbación es heterocedástica, por ejemplo $E(u_t^2) = \sigma_t^2$, es decir $E(uu') = \sum$?

Nuestro problema ahora es tal que:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$$\begin{array}{lll} E(u_t) = 0 \quad \forall t & E(u_t^2) = \sigma_t^2 & E(u_t u_s) = 0 \quad \forall t \neq s \\ X_{2t}, X_{3t} \quad \text{variables fijas} & & \\ E(X_{4t}, u_t) \neq 0 & & \end{array}$$

en este caso el estimador MCO es inconsistente y deberíamos estimar el modelo por variables instrumentales tal que encontremos instrumentos que cumplan $plim \left(\frac{1}{T} Z'u\right) = 0$.

En este caso no podemos aplicar el Teorema de Mann y Wald tal y como lo hemos enunciado, pero el segundo Teorema Central del Límite nos garantiza que vamos a encontrar una distribución asintótica para el estimador. Si recordamos del tema de heterocedasticidad vimos como White (1980) probó la existencia de un estimador para la matriz de covarianzas del estimador MCO bajo heterocedasticidad cuando la forma específica de ésta es desconocida. Entonces teníamos:

$$Y = X\beta + u$$

$$X \quad \text{fija} \quad \text{y} \quad u \sim N(0, \Omega)$$

de donde:

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

el estimador consistente de ésta matriz de varianzas y covarianzas era aquel que utilizaba como estimador de Ω a una matriz diagonal de los residuos MCO al cuadrado en t. es decir,

$$S = \begin{pmatrix} \hat u_1^2 & 0 & \ldots & 0 \\ 0 & \hat u_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & \hat u_T^2 \end{pmatrix}$$

Siguiendo el enunciado del ejemplo el estimador de MCO es inconsistente y debemos estimar por VI. Dado que $E(u_t^2) = \sigma_t^2$ podemos incluir la corrección de White en el estimador de Variables Instrumentales. White (1982) demuestra que cuando se desconoce la forma funcional de heterocedasticidad y se utiliza como estimador a aquel que usa como variables instrumentales a la matriz Z tal que $\hat\beta_{VI} = (Z'X)^{-1}Z'Y$ la matriz de covarianzas del estimador, aproximada en muestras finitas es:

$$\sum\nolimits_{(\hat\beta_{VI})} = (Z'X)^{-1}(Z'SZ)((Z'X)^{-1})'$$

donde S es la matriz diagonal de residuos mínimo cuadráticos ordinarios a cuadrado definida anteriormente.

Así para el estimador de MC2E la matriz de covarianzas aproximada en muestras finitas es:

$$\sum\nolimits_{(\hat\beta_{MC2E})} = (\hat X'X)^{-1}(\hat X'S\hat X)((\hat X'X)^{-1})'$$

donde $\hat X = W(W'W)^{-1}W'X$ y S es la matriz diagonal definida anteriormente.

Nota: En términos de sumatorios podemos escribir:

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

$$Var(\hat{\beta}_{VI}) = \left(\frac{1}{T}\sum_1^T z_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_1^T \hat{u}_t^2 z_t z_t'\right)\left[\left(\frac{1}{T}\sum_1^T z_t x_t'\right)^{-1}\right]'$$

por lo que para el caso del estimador de MC2E:

$$Var(\hat{\beta}_{MC2E}) = (\hat{X}'X)^{-1}\left(\frac{1}{T}\sum_1^T \hat{u}_t^2 \hat{x}_t \hat{x}_t'\right)\left[\left(\hat{X}'X\right)^{-1}\right]'$$

donde $\quad \hat{X} = W(W'W)^{-1}W'X$

en esta situación, de carácter general, se tiene que la diferencia $Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MC2E})$ es semidefinida positiva por lo que el estimador de MC2E sigue siendo relativamente más eficiente que otro estimador de Variables Instrumentales.
Ahora bien, **introduciendo más generalidad**, si permitimos que las variables instrumentales disponibles no sean necesariamente independientes del término de error del modelo, White probó en el trabajo citado que existe un estimador aún más eficiente que el estimador de MC2E, este estimador se denomina estimador de **Variables Instrumentales en dos Etapas(VI2E)**. En una primera etapa se estima el modelo por un procedimiento de variables instrumentales, por ejemplo MC2E, se guardan los residuos del estimador de Variables Instrumentales utilizado, por ejemplo $\hat{u}_{MC2E}$, y se obtiene en la segunda etapa el estimador:

$$\hat{\beta}_{VI2E} = \left[X'Z\left(\frac{1}{T}\sum_1^T \hat{u}_t^2 z_t z_t'\right)^{-1} Z'X\right]^{-1}\left[X'Z\left(\frac{1}{T}\sum_1^T \hat{u}_t^2 Z_t Z_t'\right)^{-1} Z'Y\right]$$

Cuya matriz de covarianzas puede aproximarse como:

$$Var(\hat{\beta}_{VI2E}) = \left[X'Z\left(\frac{1}{T}\sum_1^T \hat{u}_t^2 z_t z_t'\right)^{-1} Z'X\right]^{-1}$$

5.3.6. ¿Qué ocurre si existe autocorrelación en la perturbación?

Supongamos ahora el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 Y_{t-1} + u_t \qquad t = 1, 2, \ldots, T$$

donde:

$$\begin{array}{ll} u_t = \rho u_{t-1} + \epsilon_t \qquad \epsilon_t \sim iid(0, \sigma_\epsilon^2) & |\rho| < 1 \\ X_{2t}, X_{3t} \qquad \text{variables determinísticas} & \end{array}$$

en este caso $E(X_{2t}u_t) = E(X_{3t}u_t) = 0$ pero $E(Y_{t-1}u_t) \neq 0$ en concreto, ya que:

$$E(u_t^2) = \sigma_u^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

$$E(Y_{t-1}u_t) = E\left[(\sum_{i=0}^{\infty} \beta_1^i u_{t-1-i})u_t\right] = \frac{\rho\sigma_u^2}{1-\beta\rho} = \frac{\rho}{1-\beta\rho}\cdot\frac{\sigma_\epsilon^2}{1-\rho^2} \neq 0$$

la estimación por MCO proporciona estimadores inconsistentes. En cualquier caso, suponiendo que $E(Y_{t-1}u_t) = 0$ y dado que $u_t \sim AR(1)$ el estimador de MCO sería ineficiente y por tanto con ρ conocido estaríamos aplicando MCG o MCGF si ρ es desconocido. La pregunta es, ¿cómo estimamos ahora? El problema principal es que $E(Y_{t-1}u_t) \neq 0$ y **no** la existencia de autocorrelación, por tanto, según eso nosotros deberíamos estimar por Variables Instrumentales. Nuestro problema es que dada la existencia de autocorrelación en la perturbación no vamos a poder aplicar el Teorema de Mann y Wald y no podremos buscar directamente la distribución asintótica del estimador. No obstante dado que buscamos Z, matriz de instrumentos, tal que $plim\left(\frac{1}{T}Z'u\right) = 0$ nuestro estimador de Variables Instrumentales será consistente y si no necesitamos hacer inferencia el estimador de VI está justificado.

Si nosotros queremos hacer inferencia, lo mejor es que estimemos por máxima verosimilitud, pero también podemos optar por mirar al modelo transformado.

- **Caso 1:** ρ conocido.
  En este caso el estimador de MCG es consistente y asintóticamente normal ya que podemos obtenerlo estimando por MCO el correspondiente modelo transformado:

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2,t-1}) + \beta_3(X_{3t} - \rho X_{3,t-1}) + \beta_4(Y_{t-1} - \rho Y_{t-2}) + \epsilon_t$$

$$\text{y} \quad Cov((Y_{t-1} - \rho Y_{t-2})\epsilon_t) = 0$$
$$E((X_{2t} - \rho X_{2,t-1})\epsilon_t) = 0$$
$$E((X_{3t} - \rho X_{3,t-1})\epsilon_t) = 0$$

  con lo que la inferencia basada en la estimación Mínimo Cuadrática Generalizada es asintóticamente válida.

- **Caso 2:** ρ desconocido.
  En este caso podemos optar por implementar un proceso de estimación del tipo Cochrane-Orcutt tal que:

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2,t-1}) + \beta_3(X_{3t} - \rho X_{3,t-1}) + \beta_4(Y_{t-1} - \rho Y_{t-2}) + \epsilon_t$$

  donde $\epsilon_t = (u_t - \rho u_{t-1})$.
  En este caso $E((X_{2t} - \rho X_{2,t-1}), \epsilon_t) = E((X_{3t} - \rho X_{3,t-1})\epsilon_t) = 0$
  y además:

$$E((Y_{t-1} - \rho Y_{t-2})\epsilon_t) = 0$$

  por lo que podemos aplicar MCO a la ecuación anterior previa estimación del parámetro desconocido ρ. Sin embargo debemos notar que el estimador de ρ no debe conseguirse vía el estimador de MCO en el modelo original ya que ahora este estimador es inconsistente. Debemos estimar ρ por Variables Instrumentales de la forma siguiente:

  1. Estimamos el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 Y_{t-1} + u_t \qquad t = 2, 3, \ldots, T$$

     por VI obteniendo $\hat{\beta}_{1,VI}$, $\hat{\beta}_{2,VI}$, $\hat{\beta}_{3,VI}$ y $\hat{\beta}_{4,VI}$ consistentes. Guardamos los residuos del modelo $\hat{u}_{VI,t}$.
     Notar que necesitamos encontrar un instrumento para Y_{t-1} y que tanto el retardo de X_{2t} como el retardo de X_{3t} son válidos así, que para este ejemplo en concreto, deberíamos utilizar técnicas de MC2E.

2. Regresamos por MCO:

$$\hat{u}_{VI,t} = \rho \hat{u}_{VI,t-1} + \epsilon_t$$

y el estimador de ρ así obtenido será consistente e igual a :

$$\hat{\rho}_{VI} = \frac{\sum_2^T \hat{u}_{VI,t}\hat{u}_{VI,t-1}}{\sum_2^T \hat{u}^2_{VI,t-1}}$$

3. Sustituimos $\hat{\rho}_{VI}$ en el modelo transformado y estimamos éste por MCO:

$$\begin{array}{l}(Y_t - \hat{\rho}_{VI} Y_{t-1}) = \\ \beta_1(1-\hat{\rho}_{VI}) + \beta_2(X_{2t} - \hat{\rho}_{VI}X_{2,t-1}) + \beta_3(X_{3t} - \hat{\rho}_{VI}X_{3,t-1}) + \beta_4(Y_{t-1} - \hat{\rho}_{VI}Y_{t-2}) + \epsilon_t\end{array}$$

El estimador así conseguido es consistente y con distribución asintótica conocida. Los contrastes basados en las técnicas de MCGF son asintóticamente válidos.

Sin embargo el estimador obtenido no es totalmente eficiente. Hatanaka (1974) mostró que una ligera modificación en el modelo transformado permite obtener la eficiencia. La corrección de Hatanaka se basa en estimar por MCO el siguiente modelo transformado:

$$\begin{array}{l}(Y_t - \hat{\rho}_{VI} Y_{t-1}) = \beta_1(1-\hat{\rho}_{VI}) \\ +\beta_2(X_{2t} - \hat{\rho}_{VI}X_{2,t-1}) + \beta_3(X_{3t} - \hat{\rho}_{VI}X_{3,t-1}) + \beta_4(Y_{t-1} - \hat{\rho}_{VI}Y_{t-2}) + \rho_1\hat{u}_{VI,t-1}\epsilon_t\end{array}$$

donde el estimador $\hat{\rho}_{VI}$ se obtiene del modelo anterior y $\hat{u}_{VI,t-1} = \hat{\rho}_{VI}\hat{u}_{VI,t-2}$

En este caso podemos seguir haciendo inferencia con las técnicas de MCGF descritas anteriormente y la matriz de covarianzas del estimador se aproxima de la misma manera que en el caso anterior.

5.4. Contraste de Hausman

Necesitamos conocer un test de contraste que sea capaz de juzgar la incorrelación entre X y u. Supongamos que disponemos de dos estimadores:
$\hat{\beta}_0$ que bajo H_0 es consistente y eficiente pero inconsistente bajo H_1.
$\hat{\beta}_1$ consistente bajo H_0 y H_1 pero ineficiente bajo H_0.

y siendo:

H_0: X y u incorreladas
H_1: X y u no incorreladas

$$\begin{array}{ll}\text{Bajo} \quad H_0: & \hat{q} = \hat{\beta}_1 - \hat{\beta}_0 \xrightarrow{p} 0 \\ \text{Bajo} \quad H_1: & \hat{q} = \hat{\beta}_1 - \hat{\beta}_0 \not\xrightarrow{p} 0\end{array}$$

$$Var(\hat{q}) = Var(\hat{\beta}_1) - Var(\hat{\beta}_0)$$

siendo:
$Var(\hat{\beta}_1)$ bajo H_0

$Var(\hat{\beta}_0)$ bajo H_0
$\widehat{Var}(\hat{q})$ es un estimador consistente de $Var(\hat{q})$ el test de contraste será:

$$T\hat{q}'\widehat{Var}(\hat{q})^{-1}\hat{q} \sim \chi^2_{(p)}$$

donde:
$\hat{\beta}_0$ es el estimador de MCO
$\hat{\beta}_1$ es el estimador de VI
p es el número de restricciones que se contrastan.

El estadístico también podemos escribirlo como:

$$(\hat{\beta}_{VI} - \hat{\beta}_{MCO})'[\widehat{\sum}_{(\hat{\beta}_{IV}} - \widehat{\sum}_{\hat{\beta}_{MCO}}]^{-1}(\hat{\beta}_{VI} - \hat{\beta}_{MCO}) \xrightarrow{d} \chi^2_{(p)}$$

Cuando hay una única restricción de contraste, p=1, podemos escribir el estadístico de contraste como:

$$\frac{(\hat{\beta}_{VI} - \hat{\beta}_{MCO})^2}{\widehat{Var}(\hat{\beta}_{IV}) - \widehat{Var}(\hat{\beta}_{MCO})} \xrightarrow{d} \chi^2_{(1)}$$

Demostración: h.q.d:

$$Var(\hat{q}) = Var(\hat{\beta}_1) - Var(\hat{\beta}_0)$$

es decir:

$$Cov(\hat{\beta}_0, \hat{q}) = 0$$

i) Bajo $H_0 : \hat{\beta}_0 \quad$ y $\quad \hat{\beta}_1$ son consistentes para β y por tanto:

$$plim\hat{q} = plim\hat{\beta}_1 + plim\hat{\beta}_0 = \beta - \beta = 0$$

ii) Consideremos un nuevo estimador para β definido por:

$$\hat{d} = \hat{\beta}_0 + \lambda\hat{q} \quad \text{donde} \quad \lambda = \text{cte}$$

$plim\hat{d} = plim\hat{\beta}_0 + \lambda plim\hat{q} = \beta + \lambda \cdot 0 = \beta$ lo que implica que $\hat{d}$ es un estimador consistente de $\beta \quad \forall\lambda$

iii)

$$\begin{aligned} Var(\hat{d}) &= E(\hat{\beta}_0 + \lambda\hat{q})^2 \\ &= Var(\hat{\beta}_0) + \lambda^2 Var(\hat{q}) + 2\lambda Cov(\hat{\beta}_0, \hat{q}) \geq Var(\hat{\beta}_0) \end{aligned}$$

$$\Rightarrow \lambda^2 Var(\hat{q}) + 2\lambda Cov(\hat{\beta}_0, \hat{q}) \geq 0 \quad \forall\lambda \Rightarrow Cov(\hat{\beta}_0, \hat{q}) = 0$$

que es lo que queremos demostrar.
Si suponemos $Cov(\hat{\beta}_0, \hat{q}) > 0$ y $\lambda < 0$

$$\begin{aligned} &\lambda = \tfrac{-2Cov(\hat{\beta}_0, \hat{q})}{Var(\hat{q})} \\ &\Rightarrow \lambda^2 Var(\hat{q}) + 2\lambda Cov(\beta_0, \hat{q}) = \\ &= \tfrac{Cov(\hat{\beta}_0, \hat{q})^2}{(Var(\hat{q}))^2} Var(\hat{q}) - \tfrac{2(Cov(\hat{\beta}_0, \hat{q}))^2}{Var(\hat{q})} < 0 \end{aligned}$$

Si suponemos que $Cov(\hat{\beta}_0, \hat{q}) < 0$ y $\lambda = \frac{Cov(\hat{\beta}_0, \hat{q})}{Var(\hat{q})}$ ocurre lo mismo. Por tanto $Cov(\hat{\beta}_0, \hat{q}) = 0 \quad \forall\lambda$, como $\hat{\beta}_1 = \hat{\beta}_0 + \hat{q}$

$$Var(\hat{\beta}_1) = Var(\hat{\beta}_0) + Var(\hat{q})$$

$$\Rightarrow Var(\hat{q}) = Var(\hat{\beta}_1) - Var(\hat{\beta}_0)$$

c.q.d.

Observaciones:
El estadístico de contraste de Hausman tiene algunos problemas al ser implementado. Hay que buscar $Var(\hat{q})$, sabemos que:

$$\begin{aligned} Var(\hat{q}) &= Var(\hat{\beta}_1) - Var(\hat{\beta}_0) \\ &= \sigma_u^2 (Z'X)^{-1}(Z'Z)((Z'X)^{-1})' - \sigma_u^2 (X'X)^{-1} \end{aligned}$$

i) Si nosotros estimamos:

$$\begin{aligned} \widehat{Var}(\hat{q}) &= \widehat{Var}(\hat{\beta}_1) - \widehat{Var}(\hat{\beta}_0) \\ &= \hat{\sigma}_{u,VI}^2 (Z'X)^{-1}(Z'Z)((Z'X)^{-1})' - \hat{\sigma}_{u,MCO}^2 (X'X)^{-1} \end{aligned}$$

no tendremos problemas ya que la diferencia no será singular ya que ambos sumandos se premultiplican por su estimador consistente de la varianza de la perturbación correspondiente:

$$\hat{\sigma}_{u,VI}^2 = \frac{(Y - X\hat{\beta}_{VI})'(Y - X\hat{\beta}_{VI})}{T-k}$$

$$\hat{\sigma}_{u,MCO}^2 = \frac{(Y - X\hat{\beta}_{MCO})'(Y - X\hat{\beta}_{MCO})}{T-k}$$

pero es incorrecto ya que asintóticamente ambos coinciden y el test es asintótico. Deberíamos utilizar como estimador consistente de σ_u^2 al estimador de VI, $\hat{\sigma}_{u,VI}^2$ ya que así se puede demostrar que la potencia del contraste aumenta.

ii) Sin embargo, si buscamos

$$Var(\hat{q}) = \sigma_u^2((Z'X)^{-1}(Z'Z)((Z'X)^{-1})' - (X'X)^1)$$

esta diferencia es singular y no podremos implementar el estadístico si ocurre alguna de las cosas siguientes:

1. Existe columna de unos en el modelo y/o
2. Z y X tienen alguna columna en común.

en este caso

$$\hat{\sigma}_{u,MCO}^2 = \frac{(Y - X\hat{\beta}_{MCO})'(Y - X\hat{\beta}_{MCO})}{T-k}$$

y la solución sería utilizar el modelo en desviaciones cuando Z y X no tengan más columnas en común. Como en el ejemplo 1, que ilustra esta situación. O alternativamente conformar el test para aquellas variables que puedan estar correlacionadas con el término de error. Esta es la alternativa utilizada por Hausman y Wu (1978) quienes sugieren escribir el modelo como:

$$Y = X\beta + u = Y_1\alpha + Z_1\delta + u$$

donde: Y_1 incluye las r variables explicativas que pueden estar correlacionadas con el término de error.
Z_1 incluye las variables cuya ortogonalidad a u no se cuestiona.
proceso de contraste:

1. Estimamos el modelo por MCO y obtenemos $\hat{u}_0'\hat{u}_{0,MCO}$

2. Estimamos las regresiones auxiliares Y_1/instrumentos y se obtiene $\hat{Y}_1$ que se sustituye en el modelo inicial reestimándolo por MCO guardando $\hat{u}_1'\hat{u}_{1,MCO}$. Se computa:

$$(\hat{\beta}_{MCO} - \hat{\beta}_{VI})' \left[Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MCO})\right]^{-1} (\hat{\beta}_{MCO} - \hat{\beta}_{VI}) = \frac{\hat{u}_0'\hat{u}_{0,MCO} - \hat{u}_1'\hat{u}_{1,MCO}}{\hat{\sigma}_u^2}$$

$$\frac{\hat{u}'\hat{u}_{0,MCO} - \hat{u}_1'\hat{u}_{1,MCO}}{\hat{\sigma}_u^2} \sim \chi_r^2$$

bajo la hipótesis nula de que todas las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría el supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.

Los ejemplos 1 y 2 conforman el test de Hausman para un modelo en desviaciones donde Z y X no tienen nada en común excepto la columna correspondiente al término independiente. Si los modelos no se tratasen en desviaciones a la media la matriz de covarianzas $\sigma_u^2((Z'X)^{-1}(Z'Z)((Z'X)^{-1})' - (X'X)^1)$ sería singular.

Ejercicio 1
Sea:

$$Y_t = \alpha + \beta X_t + u_t \qquad u_t \sim N(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T$$

y queremos contrastar la incorrelación entre X y u.

$$H_0 : E(X_t u_t) = 0$$
$$H_1 : E(X_t u_t) \neq 0$$

$$\begin{aligned}
&\hat{\beta}_{0,MCO} = \tfrac{\sum x_t y_t}{\sum x_t^2} \\
&\hat{\beta}_{1,VI} = \tfrac{\sum y_t z_t}{\sum x_t z_t} = \beta + \tfrac{\sum u_t z_t}{\sum x_t z_t} \\
&Var(\hat{\beta}_0) = \tfrac{\sigma_u^2}{\sum x_t^2} \\
&Var(\hat{\beta}_1) = \sigma_u^2 \tfrac{\sum z_t^2}{(\sum x_t z_t)^2} \\
&\hat{q} = \hat{\beta}_1 - \hat{\beta}_0
\end{aligned}$$

$$\begin{aligned}
Var(\hat{q}) &= Var(\hat{\beta}_1) - Var(\hat{\beta}_0) \\
&= \sigma_u^2 \left(\tfrac{\sum z_t^2}{(\sum x_t z_t)^2} - \tfrac{1}{\sum x_t^2} \right) = \\
&= \sigma_u^2 \tfrac{\sum x_t^2}{\sum x_t^2} \left(\tfrac{\sum z_t^2}{(\sum x_t z_t)^2} - \tfrac{1}{\sum x_t^2} \right) = \\
&= \tfrac{\sigma_u^2}{\sum x_t^2} \left(\tfrac{\sum x_t^2 \sum z_t^2}{(\sum x_t z_t)^2} - 1 \right) = \\
&= \tfrac{\sigma_u^2}{\sum x_t^2} \left(\tfrac{1}{r_{XZ}^2} - 1 \right) = \\
&= Var(\hat{\beta}_0) \left(\tfrac{1 - r_{XZ}^2}{r_{XZ}^2} \right)
\end{aligned}$$

de donde el estadístico de Hausman, que denotamos por m sería:

$$m = \frac{\hat{q}^2}{\widehat{Var}(\hat{q})} = \frac{(\hat{\beta}_{0,MCO} - \hat{\beta}_{1,VI})^2}{\widehat{Var}(\hat{\beta}_0) \left(\frac{1 - r_{XZ}^2}{r_{XZ}^2} \right)}$$

Operando:

$$m = \frac{(\hat{\beta}_{0,MCO} - \hat{\beta}_{1,VI})^2 r^2_{XZ}}{\widehat{Var}(\hat{\beta}_0)(1 - r^2_{XZ})} \sim \chi^2_{(1)}$$

donde

$$\widehat{Var}(\hat{\beta}_0) = \frac{\hat{\sigma}^2_u}{\sum x_t^2} \quad \text{y} \quad \hat{\sigma}^2_u = \hat{\sigma}^2_{u,VI}$$

Ejercicio 2
Se propone la siguiente especificación para la función de demanda de vino de un país:

$$Q_t = \alpha + \beta P_t + u_t \qquad t = 1, 2, \ldots, T$$

donde $u_t \sim (0, 0{,}0921)$. dado que el precio se determina simultáneamente con la cantidad Q_t, se sospecha que P_t pueda estar correlacionada con u_t. Se dispone de datos de un índice de costes de almacenamiento, S_t que se determina exógenamente, por lo que se considera independiente de u_t. Dados los siguientes datos para los años de 1955-1975:

$$\begin{array}{ll} \sum s_t q_t = 1,78037 & \sum s_t^2 = 2,1417 \\ \sum p_t^2 = 0,507434 & \sum p_t s_t = 0,500484 \\ \sum s_t q_t = 2,75474 & \end{array}$$

Queremos contrastar la incorrelación entre P_t y u_t:

$$H_0 : E(P_t u_t) = 0$$

$$H_1 : E(P_t u_t) \neq 0$$

$$\hat{\beta}_{0,MCO} = \frac{\sum p_t q_t}{\sum p_t^2} = \frac{1,78037}{0,507434} = 3,5085$$

$$\hat{\beta}_{1,VI} = \frac{\sum s_t q_t}{\sum s_t p_t} = \frac{2,75474}{0,500484} = 5,4862$$

de donde como instrumento para P_t se usa el índice de coste de existencias de almacén S_t

$$\hat{q} = \hat{\beta}_{1,VI} - \hat{\beta}_{0,MCO} = 1,9777$$

$$r^2_{PS} = \frac{(\sum p_t q_t)^2}{\sum p_t^2 \sum s_t^2} = 0,2304$$

$$\hat{\sigma}^2 = 0,09217$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2_u}{\sum p_t^2} = \frac{0,09217}{0,507434} = 0,18164$$

$$m = \frac{\hat{q}^2 r^2_{P_t S_t}}{Var(\hat{\beta}_0)(1 - r^2_{P_t S_t})} = \frac{(1,9777)^2 (0,2304)^2}{0,18164(1 - 0,2304)} = 6,5721$$

$\chi^2_{0,05(1)} = 3,841$
$6,5721 > 3,841$ por tanto rechazo la hipótesis nula para $\alpha = 5\,\%$ y P_t y u_t no son incorreladas.

Nota: El resultado del problema es exactamente el mismo que si nosotros buscamos directamente el estadístico m como:

$$m = \frac{(\hat{\beta}_{VI} - \hat{\beta}_{MCO})^2}{Var(\hat{\beta}_{VI}) - Var(\hat{\beta}_{MCO})}$$

donde:

$$Var(\hat{\beta}_{VI}) = \sigma_u^2 \frac{\sum s_t^2}{(\sum s_t p_t)^2} = \frac{0,0921 \cdot 2,1417}{(0,500484)^2} = 0,78747$$

$$m = \frac{(1,9777)^2}{0,78747 - 0,18164} = 6,5721$$

ya que en este caso σ_u^2 es conocida.

5.5. Errores de medida en variables

Hasta ahora hemos supuesto que las variables utilizadas en el proceso de estimación se medían sin error. En la práctica es muy posible que existan errores de medida en las variables o que simplemente las variables a utilizar no sean sino estimaciones de conceptos teóricos que no se observan en la realidad, por ejemplo el stock de capital, el PIB, o las variables de Contabilidad Nacional. Estas situaciones alterarán las propiedades de los estimadores de los parámetros, en concreto introduciendo sesgos en las estimaciones y generando estimadores de MCO inconsistentes. Estudiamos tres casos:

1. Variable endógena medida con error.
2. Variable exógena medida con error.
3. Variable exógena y endógena medidas con error.

5.5.1. Variable endógena medida con error

Sea

$$Y_t = \alpha + \beta X_t + u_t \qquad u_t \sim N(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T$$

el verdadero modelo. Pero por alguna razón la variable endógena disponible no es Y_t sino $Y_t^\star = Y_t + \epsilon_t$ y por tanto el modelo que vamos a estimar es:

$$Y_t = \alpha + \beta X_t + u_t \qquad u_t \sim N(0, \sigma_u^2) \qquad t = 1, 2, \ldots, T$$

$$\text{donde} \quad Y_t^\star = Y_t + \epsilon_t \quad \text{y} \quad \begin{array}{l} \epsilon_t \sim iid(0, \sigma_\epsilon^2) \\ Cov(X_t, \epsilon_t) = 0 \\ Cov(u_t, \epsilon_t) = 0 \end{array}$$

por lo que la auténtica relación a estimar es:

$$\begin{array}{rl} Y_t^\star - \epsilon_t = \alpha + \beta X_t + u_t & \Rightarrow Y_t^\star = \alpha + \beta X_t + (u_t + \epsilon_t) \\ & \Rightarrow Y_t^\star = \alpha + \beta X_t + u_t^\star \end{array}$$

El error de medida en la variable endógena se acumula en la perturbación original, con lo cual debemos preocuparnos por las propiedades de la nueva perturbación:

$$\begin{array}{l} E(u_t^\star) = E(u_t + \epsilon_t) = E(u_t) + E(\epsilon_t) = 0 \\ Var(u_t^\star) = E(u_t^\star - E(u_t^\star))^2 = E(u_t^\star)^2 = E(u_t + \epsilon_t)^2 = E(u_t^2) + E(\epsilon_t^2) + 2E(u_t \epsilon_t) = \sigma_u^2 + \sigma_\epsilon^2 \\ Cov(u_t^\star, u_s^\star) = E((u_t^\star - E(u_t^\star))(u_s^\star - E(u_s^\star))) = E(u_t^\star u_s^\star) = E(u_t + \epsilon_t, u_s + \epsilon_s) = 0 \end{array}$$

Por tanto la perturbación del modelo a estimar es homocedástica y no autocorrelada. Resumiendo, si:

$$u_t \sim N(0, \sigma_u^2) \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad \text{y} \quad Cov(u_t, \epsilon_t) = 0 \quad \forall t \quad \Rightarrow u_t^\star \sim N(0, \sigma_u^2 + \sigma_\epsilon^2)$$

Como conclusión podemos decir que en presencia de errores de medida en la variable endógena exclusivamente y dado que $Cov(X_t, \epsilon_t) = 0$ y $Cov(X_t, u_t) = 0$ y la perturbación del modelo estimable tiene propiedades esféricas, los MCO son apropiados y tienen buenas propiedades.

5.5.2. Variable exógena medida con error

Sea el verdadero modelo de regresión:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1, 2, \ldots, T$$

donde la variable X_t es una variable fija pero inobservable, pero observamos $X_t^\star = X_t + v_t$, variable aleatoria que incorpora el efecto de v_t, aún en el caso de que X_t sea fija. Además hacemos las siguientes hipótesis:

$$\begin{array}{l} u_t \sim iid(0, \sigma_u^2) \\ v_t \sim iid(0, \sigma_v^2) \\ Cov(u_t, v_t) = Cov(u_t, v_s) = Cov(u_s, v_t) = 0 \end{array}$$

En esta situación el modelo que efectivamente se estima es el siguiente:

$$\begin{array}{l} Y_t = \alpha + \beta(X_t^\star - v_t) + u_t \\ Y_t = \alpha + \beta X_t^\star + (u_t - \beta v_t) \\ Y_t = \alpha + \beta X_t^\star + u_t^\star \end{array}$$

- Propiedades de la nueva perturbación $u_t^\star$:

$$\begin{array}{ll} E(u_t^\star) & = E(u_t - \beta v_t) = E(u_t) - \beta E(v_t) = 0 \\ Var(u_t^\star) & = E(u_t^\star - E(u_t^\star))^2 = E(u_t^\star)^2 = E(u_t - \beta v_t)^2 \\ & = E(u_t^2) + \beta^2 E(v_t^2) - 2\beta E(u_t v_t) = \sigma_u^2 + \beta^2 \sigma_v^2 - 2\beta \cdot 0 \\ & = \sigma_u^2 + \beta^2 \sigma_v^2 \quad \text{homocedástica} \\ Cov(u_t^\star, u_s^\star) & = E((u_t^\star - E(u_t^\star))(u_s^\star - E(u_s^\star))) = E(u_t - \beta v_t, u_s - \beta v_s) = \\ & = E(u_t u_s) - \beta E(v_t u_s) - \beta E(u_t v_s) + \beta^2 E(v_t v_s) = 0 \quad \text{no autocorrelada} \end{array}$$

- Además necesitamos conocer la relación entre el regresor y el error, es decir $X_t^\star$ y $u_t^\star$:

$$\begin{array}{ll} Cov(X_t^\star, u_t^\star) & = E(X_t^\star u_t^\star) = E((X_t + v_t)(u_t - \beta v_t)) = \\ & = E(X_t u_t) + E(v_t u_t) - \beta E(X_t v_t) - \beta E(v_t^2) = -\beta \sigma_v^2 \end{array}$$

  ya que al ser X_t una variable fija $E(X_t u_t) = E(X_t v_t) = 0$. Al ser la covarianza entre $X_t^\star$ y $u_t^\star$ distinta de cero existe correlación contemporánea entre la variable exógena y la perturbación.

  Buscamos ahora la existencia de correlación no contemporánea:

$$\begin{array}{ll} Cov(X_s^\star, u_t^\star) & = E(X_s^\star u_t^\star) == E((X_s + v_s)(u_t - \beta v_t) = \\ & = E(X_s u_t) + E(v_s u_t) - \beta E(X_s v_t) - \beta E(v_s v_t) = 0 \end{array}$$

  y por lo tanto no existe correlación no contemporánea entre la variable exógena y la perturbación del modelo estimable.

- Lo que sabemos es:
  El modelo a estimar es:

$$Y_t = \alpha + \beta X_t^\star + u_t^\star \qquad t = 1, 2, \ldots, T$$

$$\text{donde:} \quad \begin{array}{l} u_t^\star \sim iid(0, \sigma_u^2 + \beta^2\sigma_v^2) \\ Cov(X_t^\star, u_t^\star) = -\beta\sigma_v^2 \\ Cov(X_s^\star, u_t^\star) = 0 \end{array}$$

El estimador MCO del parámetro β en el modelo es:

$$\hat{\beta}_{MCO} = \frac{\sum y_t x_t^\star}{\sum x_t^{\star 2}} = \beta + \frac{\sum x_t^\star u_t^\star}{\sum x_t^{\star 2}}$$

Buscamos sus propiedades en muestras pequeñas:

$$E(\hat{\beta}_{MCO}) = \beta + E\left(\frac{\sum x_t^\star u_t^\star}{\sum x_t^{\star 2}}\right) \neq \beta$$

por lo tanto el estimador es sesgado y estaremos interesados en sus propiedades asintóticas. Para ello tendremos que introducir hipótesis sobre la relación en el límite entre las diferentes variables. Los supuestos, aplicados en términos generales, son:

1. Los errores de medida en $X^\star$ no están correlacionados en el límite con los verdaderos regresores $X \quad \Rightarrow plim\left(\frac{1}{T}X'V\right) = 0$.

2. $plim\left(\frac{1}{T}X'X\right) = Q_{XX}$.

3. $plim\left(\frac{1}{T}V'V\right) = \Omega_v$

   Bajo estos supuestos [1]+[2]+[3] tenemos:

$$\begin{aligned} plim\left(\frac{1}{T}X^\star X^\star\right) &= plim\left(\frac{1}{T}(X+V)'(X+V)\right) = \\ &= plim\left(\frac{1}{T}(X'X + V'X + X'V + V'V)\right) = \\ &= plim\left(\frac{1}{T}X'X\right) + plim\left(\frac{1}{T}V'X\right) + plim\left(\frac{1}{T}X'V\right) + plim\left(\frac{1}{T}V'V\right) = \\ &= Q_{XX} + \Omega_v \end{aligned}$$

4. La perturbación no está correlacionada en el límite ni con X ni con el error de medida en X, así:

$$\begin{array}{l} plim\left(\frac{1}{T}V'U\right) = 0 \\ plim\left(\frac{1}{T}X'U\right) = 0 \end{array}$$

Buscamos la consistencia del estimador, para ello le escribimos como:

$$\hat{\beta}_{MCO} = \beta + \frac{1/T\sum x_t^\star u_t^\star}{1/T\sum x_t^{\star 2}}$$

de donde:

$$plim\hat{\beta}_{MCO} = plim\beta + plim\left(\frac{1}{T}\sum x_t^{\star 2}\right)^{-1} + plim\left(\frac{1}{T}\sum x_t^\star u_t^\star\right)$$

$$\begin{aligned} plim\left(\frac{1}{T}\sum x_t^{\star 2}\right) &= plim\left(\frac{1}{T}\sum (x_t + v_t)^2\right) = \\ & plim\left(\frac{1}{T}\sum x_t^2\right) + plim\left(\frac{1}{T}\sum v_t^2\right) + 2\left(plim\left(\frac{1}{T}\sum x_t v_t\right)\right) = \\ &= \sigma_X^2 + \sigma_v^2 \end{aligned}$$

$$\begin{aligned} plim\left(\frac{1}{T}\sum x_t^\star u_t^\star\right) &= plim\left(\frac{1}{T}\sum (x_t+v_t)(u_t-\beta v_t)\right) = \\ & plim\left(\frac{1}{T}\sum x_t u_t\right) + plim\left(\frac{1}{T}\sum v_t u_t\right) - \beta plim\left(\frac{1}{T}\sum x_t v_t\right) - \beta\left(plim\frac{1}{T}\sum v_t^2\right) = \\ &= 0+0-\beta\cdot 0-\beta\sigma_v^2 = -\beta\sigma_v^2 \end{aligned}$$

De donde:

$$plim\hat{\beta}_{MCO} = \beta - \frac{\beta\sigma_v^2}{\sigma_X^2+\sigma_v^2} \neq \beta$$

y por lo tanto **inconsistente**.

$$\text{Sesgo asintótico} = plim\hat{\beta} - \beta = \frac{\beta\sigma_v^2}{\sigma_X^2+\sigma_v^2}$$

Aunque el error de medida afecta sólo a X la inconsistencia se traslada a todos los parámetros estimados por MCO. Así para el término independiente tenemos:

$$\hat{\alpha}_{MCO} = \overline{Y} - \hat{\beta}\overline{X}^\star = \alpha + \beta\overline{X}^\star + \overline{u}^\star - \hat{\beta}\overline{X}^\star = \alpha - (\hat{\beta}-\beta)\overline{X}^\star + \overline{u}^\star$$

$$plim\hat{\alpha} = \alpha - plim(\hat{\beta}-\beta)plim\overline{X}^\star + plim\overline{u}^\star = \alpha + \frac{\beta\sigma_v^2}{\sigma_X^2+\sigma_v^2}\mu_X^\star + 0 \neq \alpha$$

por lo tanto como conclusión podemos decir que un error de medida en la variable exógena tal que $E(X_t^\star u_t^\star) \neq 0$ implica que los estimadores MCO son sesgados e inconsistentes. Si el error de medida fuese una constante no se producirían sesgos en la estimación de los parámetros. El modelo debería ser estimado por el Método de Variables Instrumentales.

5.5.3. Variable exógena y variable endógena medidas con error

Sea el verdadero modelo:

$$Y_t = \alpha + \beta X_t + u_t \qquad t = 1,2,\ldots,T$$

donde:
$Y_t^\star = Y_t + \epsilon_t$ es la variable endógena disponible.
$X_t^\star = X_t + v_t$ es la variable exógena disponible.
$u_t \sim iid(0,\sigma_u^2)$ $v_t \sim iid(0,\sigma_v^2)$ $\epsilon_t \sim N(0,\sigma_\epsilon^2)$ $Cov(u_t,\epsilon_t) = Cov(u_t,v_t) = Cov(\epsilon_t,v_t) = Cov(X_t,\epsilon_t) = 0$
El modelo a estimar sería:

$$\begin{aligned} Y_t^\star - \epsilon_t &= \alpha + \beta(X_t^\star - v_t) + u_t \qquad t=1,2,\ldots,T \\ Y_t^\star &= \alpha + \beta X_t^\star + (u_t + \epsilon_t - \beta v_t) \qquad t=1,2,\ldots,T \end{aligned}$$

Llamamos $u_t^\star = u_t + \epsilon_t - \beta v_t \sim iid(0,\sigma_u^2+\sigma_\epsilon^2+\beta^2\sigma_v^2)$ homocedástica y no autocorrelada.

Existe correlación contemporánea ya que:

$$\begin{aligned} E(X_t^\star u_t^\star) &= E((X_t+v_t)(u_t+\epsilon_t-\beta v_t) = \\ &= E(X_t u_t) + E(v_t u_t) + E(X-t\epsilon_t) + E(v_t\epsilon_t) - \beta E(X_t v_t) - \beta E(v_t^2) = \\ &= -\beta\sigma_v^2 \end{aligned}$$

No existe correlación no contemporánea ya que $\quad E(X_s^\star u_t^\star) = 0$

El error de medida en Y_t implica un incremento en la varianza de la perturbación del modelo estimable mientras que el error de medida en X_t implica que los estimadores MCO de α y β serán sesgados e inconsistentes. El modelo a estimar (bajo todos los supuestos realizas anteriormente) debería ser estimado por el Método de Variables Instrumentales.

Bibliografía

[1] Alegre, J., J. Arcarons, C. Bolancé y L. Díaz, (1995), Ejercicios y problemas de Econometría, Ed. AC, Colección Plan Nuevo, Madrid.

[2] Alonso, A., F.J. Fernández e I. Gallastegui (2005), Econometría, Prentice Hall, Madrid.

[3] Aznar, A. y A. García (1984), Problemas de Econometría, Pirámide, Madrid.

[4] Fernández, A., P. González, M. Regúlez, P. Moral y M. V. Esteban (2005), Ejercicios de Econometría, 2ª edn., MacGraw-Hill, serie Schaum, Madrid.

[5] Greene, W. (1998), Análisis Econométrico, 3ª edn., Prentice Hall, New Jersey.

[6] Gujarati, D. (1990), Econometría, 2ª edn., MacGraw-Hill, Madrid.

[7] Hill, R. C., W.E. Griffiths, y G. G. Judge (2001), Undergraduate Econometrics, 2ª edn., John Wiley and Sons, Inc., England.

[8] Johnston, J y J. Dinardo (2001), Métodos de Econometría, Vicens Vives, Barcelona.

[9] Novales, A. (1993), Econometría, Edición revisada, McGraw-Hill, Madrid.

[10] Ramanathan, R. (2002), Introductory Econometrics with applications, 5th. edition, Ed. South-Western, Mason, Ohio.

[11] Stock, J. y M. Watson (2003), Introduction to Econometrics, Addison-Wesley, Boston.

[12] Uriel, E., D. Contreras, L. Moltó y A. Peiro (1990), Econometría. El modelo lineal, Ed. AC, Madrid.

[13] Wooldridge, J. M. (2003), Introductory Econometrics: A modern Approach, 2nd. edition, Thomson Learning, Mason, Ohio.