

Modelos Econométricos Multiecuacionales de Estimación de Demandas

Autor:

Jorge Mauricio Oviedo ¹

Resumen: En este artículo se efectúa una revisión de los principales Métodos Econométricos para estimar ecuaciones simultáneas de demanda considerando diversos supuestos sobre los términos de errores. Se analizan las propiedades de los estimadores y se los comparan con los de Mínimos Cuadrados. Se extraen sugerencias en cuanto a la utilidad y aplicabilidad de cada uno de ellos.

Palabras clave: Sistemas de Ecuaciones Simultáneas, Mínimos Cuadrados Ordinarios, Mínimos Cuadrados en dos etapas, Mínimos Cuadrados en tres etapas, SUR.

¹ joviedo@eco.unc.edu.ar

1.- Introducción

Los modelos econométricos instructorios aprendidos en los cursos de grado, se caracterizaban por constituir modelos en donde los verdaderos procesos generadores de datos provenían de modelos de una simple ecuación lineal en los parámetros, es decir, los modelos en que únicamente había una variable dependiente Y y una o varias variables explicatorias X_i . A saber:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_n X_{nt} + u_t \quad \text{Matricialmente: } \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

En los mismos se vio, a través del Teorema Gauss-Markov, que los estimadores obtenidos por el Método de los Mínimos Cuadrados Ordinarios (MCO), resultaban ser lineales, insesgados y de varianza mínima (dentro del conjunto de todos los estimadores insesgados posibles) siempre que se cumplieran los supuestos de $E(\mathbf{u}) = \mathbf{0}$, $E(\mathbf{u}'\mathbf{u}) = \sigma^2 \mathbf{I}$, exacta especificación lineal en los parámetros y que las variables independientes o explicativas no son estocásticas (son fijas en muestras repetidas) o, en caso de serlo, se distribuyen independientemente del término de perturbación.

Los Modelos de Ecuaciones Simultaneas surgen para captar la posibilidad más realista que los valores observados (\mathbf{Y}, \mathbf{X}) provengan de un proceso generador de datos en donde éstos son creados en forma simultanea y mutuamente interdependientes vía una interconexión entre ellos. Esto ocurre cuando no solamente la Y es determinada por las X , sino que además algunas de las X son a su vez determinadas por Y . En otras palabras, cuando hay una relación causal en las dos direcciones o una relación simultánea entre Y y algunas de las X , lo cual hace que la distinción entre variable dependiente y variable explicatoria sea de poco valor. Es mejor tener un conjunto de variables que pueden ser determinadas simultáneamente por otras y esto es lo que efectivamente se hace en los modelos de ecuaciones simultaneas.

Los sistemas de ecuaciones simultáneas se distinguen por estar conformadas por varias ecuaciones en las cuales hay un número de variables endógenas o variables determinadas conjuntamente y un número de variables predeterminadas, o determinantes (estas a su vez pueden ser variables exógenas, retardadas o no, y variables endógenas retardadas). En estos modelos se estiman los parámetros de las ecuaciones teniendo en cuenta la información suministrada por todas las ecuaciones del sistema. Un supuesto implícito en estos tipos de modelos es que los valores observados corresponden siempre a situaciones de equilibrio, es decir no se concibe la posibilidad de obtener datos en algún momento de transición hacia el equilibrio.

En un modelo general lineal que contenga M ecuaciones estructurales en M variables endógenas o conjuntamente dependientes y K variables predeterminadas, la relación i -ésima en el momento (observación) t puede escribirse en forma escalar como:

$$\beta_{i1} Y_{1t} + \dots + \beta_{iG} Y_{Gt} + \gamma_{i1} X_{1t} + \dots + \gamma_{iK} X_{Kt} = u_{it} \quad ; \quad (i = 1 \dots M) \quad ; \quad (t = 1 \dots T)$$

o en forma matricial:

$$\mathbf{Y}_t \boldsymbol{\Gamma} + \mathbf{X}_t \mathbf{B} = \mathbf{u}_t \quad t = 1, \dots, n$$

donde :

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \dots & \beta_{kM} \end{bmatrix}; \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{M1} & \gamma_{M2} & \dots & \gamma_{MM} \end{bmatrix};$$

$$\mathbf{Y}_t = [y_{1t} \quad \cdots \quad y_{Mt}]; \quad \mathbf{X}_t = [x_{1t} \quad \cdots \quad x_{Kt}]; \quad \mathbf{u}_t = [u_{1t} \quad \cdots \quad u_{Mt}]$$

Se imponen además las siguientes supuestas al modelo estadístico:

1)

$$E[e_i] = 0, \quad i = 1, \dots, M$$

$$E[e_i e_j'] = \sigma_{ij} I_T, \quad i = 1, \dots, M, \quad j = 1, \dots, M$$

$$E[e, e'] = \Sigma \otimes I_T$$

2) La matriz Γ es no singular

3) $\text{plim}(X'E/T) = 0$

Según lo aprendido hasta ahora uno intentaría aplicar MCO en forma independiente a cada una de las M ecuaciones con respecto a las $M+K-1$ variables explicativas ($M-1$ Y 's y K X 's) para hallar los estimadores \mathbf{B} y Γ , pero si uno efectúa esta acción se encontrará con que dichas estimaciones no solamente serán sesgadas sino también inconsistentes es decir que a medida que el tamaño de la muestra crece indefinidamente los estimadores no convergen al verdadero valor del parámetro, permaneciendo el sesgo. Esto se debe a que las $M-1$ variables endógenas restantes que aparecen en una ecuación cualquiera estarán correlacionas con el término de perturbación de la ecuación considerada, puesto que por ser cada una de las M variables argumentos aleatorios, una perturbación en una, alguna o todas las restantes ($M-1$) afectaran el valor del termino de error de dicha ecuación quien luego influirá en las demás ecuaciones. Es decir el termino de perturbación de cada ecuación depende de los valores que asuman las $G-1$ variables endógenas restantes y viceversa. De esta manera se viola uno de los supuestos del teorema de Gauss-Markov, haciendo indeseable la aplicación directa de los MCO a cada una de las ecuaciones. Para ello, demostraremos dichas afirmaciones a continuación

Consecuencias de la aplicación directa de MCO

Para demostrar que el estimador por mínimos cuadrados es inconsistente y sesgado consideremos la siguiente ecuación:

$$Y\Gamma_i + X\mathbf{B}_i + \mathbf{e}_i = 0$$

Algunos elementos de Γ y \mathbf{B} generalmente pueden tomar el valor de cero mientras que es costumbre seleccionar una variable endógena del lado izquierdo de la ecuación. Esto es llamado normalización y es logrado fijando un coeficiente, digamos γ_{ij} , con el valor -1. Así, con algunas manipulaciones si es necesario, se tiene:

$$\mathbf{y}_i = Y_i \gamma_i + Y_i^* \gamma_i^* + X_i \beta_i + X_i^* \beta_i^* + \mathbf{e}_i$$

$$\mathbf{y}_i = Y_i \gamma_i + X_i \beta_i + \mathbf{e}_i$$

$$\mathbf{y}_i = [Y_i \quad X_i] \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix} + \mathbf{e}_i$$

$$\mathbf{y}_i = Z_i \delta_i + \mathbf{e}_i$$

Donde:

$$\Gamma_i = \begin{bmatrix} 1 \\ \gamma_i \\ \gamma_i^* \end{bmatrix} = \begin{bmatrix} 1 \\ \gamma_i \\ 0 \end{bmatrix}, \quad B_i = \begin{bmatrix} \beta_i \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} \beta_i \\ 0 \end{bmatrix}, \quad \delta_i = \begin{bmatrix} \gamma_i \\ \beta_i \end{bmatrix}$$

$$Y = [y_i \quad Y_i \quad Y_i^*], \quad X = [X_i \quad X_i^*], \quad Z_i = [Y_i \quad X_i]$$

y $M = m_i + m_i^*$, $K = k_i + k_i^*$. La matriz Y_i^* contiene aquellos m_i^* variables endógenas que no aparecen en la i -ésima ecuación; esto es, sus coeficientes asociados γ_i^* son cero. La matriz X_i^* contiene aquellas k_i^* variables predeterminadas que no aparecen en la i -ésima ecuación, o sea, sus coeficientes asociados β_i^* son cero.

El estimador mínimo cuadrático de δ_i es:

$$\hat{\delta}_i = (Z_i' Z_i)^{-1} Z_i' y_i$$

Su valor esperado es

$$E[\hat{\delta}_i] = E[(Z_i' Z_i)^{-1} Z_i' y_i]$$

$$= \delta_i + E[(Z_i' Z_i)^{-1} Z_i' e_i] \neq \delta_i$$

El último término de la expresión anterior no desaparece ya que Z_i contiene variables endógenas que son conjuntamente determinadas con y_i y por ende no son independientes de e_i . Con lo cual el estimador es sesgado.

Adicionalmente, a medida que el tamaño de la muestra crece el estimador no converge en probabilidad al verdadero valor del parámetro ya que

$$\text{plim } \hat{\delta}_i = \delta_i + \text{plim}[Z_i' Z_i / T]^{-1} \text{plim}[Z_i' e_i / T] \neq \delta_i$$

El último término, no converge en probabilidad al vector nulo ya que $Z = [Y_i \quad X_i]$ contiene a Y_i , el cual no es independiente del término de error u , independientemente del tamaño de la muestra.

De esta manera los estimadores mínimo-cuadráticos son sesgados e inconsistentes.

Forma Reducida

Cuando se expresa a las variables endógenas en términos de las predeterminadas, dicha reexpresión se denomina Forma Reducida. Específicamente, si post multiplicamos el sistema por Γ^{-1} y reacomodando términos se tiene:

$$Y\Gamma + XB = U \quad (t = 1, \dots, T)$$

$$Y = X\Pi + V$$

donde:

$$\Pi = -B\Gamma^{-1} = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1M} \\ \vdots & \ddots & \vdots \\ \pi_{K1} & \cdots & \pi_{KM} \end{bmatrix} = [\pi_1 \quad \cdots \quad \pi_M] \quad (1) \quad y$$

$$V = -U\Gamma^{-1} = \begin{bmatrix} v_{11} & \cdots & v_{1M} \\ \vdots & \ddots & \vdots \\ v_{K1} & \cdots & v_{KM} \end{bmatrix} = [v_1 \quad \cdots \quad v_M]$$

Los supuestos estocásticos sobre V se siguen directamente de los de u . Si u_t' es la i -ésima fila de U y v_t' la i -ésima fila de V , entonces:

$$v_t' = -u_t \Gamma^{-1}$$

Ya que todos los vectores u_t tienen media igual a cero y matriz de covarianzas Σ , se sigue que

$$E[v_t] = E[-u_t \Gamma^{-1}]' = -\Gamma^{-1} E[u_t] = \mathbf{0}$$

$$\begin{aligned} \text{var}(v_t) &= E[v_t v_t'] \\ &= (\Gamma^{-1})' E(u_t u_t') (\Gamma^{-1}) \\ &= (\Gamma^{-1})' \Sigma (\Gamma^{-1}) \\ &= \Omega \end{aligned}$$

Adicionalmente, como $E(u_t u_s') = 0$, entonces $E(v_t v_s') = 0$ para todo t distinto de s . La forma reducida individual será:

$$y_i = X \pi_i + v_i$$

Si X es no estocástica, se verificará:

$$E(X'V) = E[-X'ET] = \mathbf{0}$$

Si X es aleatoria entonces el $\text{plim}(X'V/T) = \text{plim}[-X'U/T] \Gamma^{-1} = 0$

Entonces los Estimadores Mínimo Cuadráticos para la i -ésima ecuación serán

$$\hat{\pi}_i = (X'X)^{-1} X'y_i$$

y para todo el sistema

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

Los mismos serán insesgados y consistentes.

De esta manera, ya que es posible estimar los parámetros de la forma reducida de manera insesgada y consistente y dada la relación entre la forma estructural y la reducida

$$\begin{aligned} \Pi &= -B\Gamma^{-1} \\ \Omega &= (\Gamma^{-1})\Sigma\Gamma^{-1} \end{aligned}$$

La cuestión que emerge es si los parámetros estructurales pueden ser únicamente derivados de las estimaciones de la forma reducida y alguna otra información sobre la forma estructural. Esto nos conduce al problema de la **identificación**.

Este tipo de situaciones pueden obtenerse en un modelo clásico de ofertas y demandas de un bien en donde una alteración en las funciones de Demanda u Oferta ocasionarían alteraciones en el precio y cantidad observada con lo cual el precio no será independiente de las cantidades ofrecidas y demandadas.

En virtud de la no-deseabilidad de la aplicación de MCO a este modelo, lo que se intenta hacer es reexpresar el sistema original (dado en su forma estructural, es decir tal y como surge de la realidad siendo un fiel interprete de la misma) de modo tal de que cada variable endógena muestre su dependencia sólo con respecto a las variables predeterminadas y los términos de perturbación. De esta manera, al ser por definición las variables predeterminadas independientes de los términos de perturbación y manteniendo el supuesto de “ruido blanco” para u_t , sería eficiente estimar este nuevo sistema de ecuaciones por MCO (en forma independiente a cada una de las ecuaciones) ya que los estimadores así obtenidos serán nuevamente los mejores al entrar en vigencia una vez más el Teorema de Gauss-Marcov.

Matricialmente:

$$\begin{aligned} \mathbf{Y}_t \boldsymbol{\Gamma} + \mathbf{X}_t \mathbf{B} &= \mathbf{u}_t \\ \mathbf{Y}_t &= \mathbf{X}_t \boldsymbol{\Pi} + \mathbf{v}_t \end{aligned} \quad (t = 1, \dots, T)$$

donde:

$$\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1} \quad (1) \quad \text{y} \quad \mathbf{v}_t = \mathbf{u}_t \boldsymbol{\Gamma}^{-1}$$

Una vez obtenido los estimadores de $\boldsymbol{\Pi}$ el paso siguiente es intentar recuperar los parámetros de la forma estructural a través de los estimadores de los parámetros de la forma reducida vía la relación $\boldsymbol{\Pi} = -\mathbf{B}\boldsymbol{\Gamma}^{-1}$. Como se puede dilucidar, la matriz $\boldsymbol{\Pi}$ es de orden $M \times K$ y, por lo tanto, contiene GK elementos. Las matrices \mathbf{B} y $\boldsymbol{\Gamma}$ contienen como máximo $M^2 + MK$ elementos. Por ende, existe una infinidad de estructuras de \mathbf{B} y $\boldsymbol{\Gamma}$ que corresponden a cualquier matriz $\boldsymbol{\Pi}$ dada.

Esta imposibilidad de recuperar los parámetros en un caso general conlleva al tratamiento del problema de la identificación es decir el análisis de los casos y las condiciones bajo las cuales este proceso de recupero se torna factible.

Así se pueden presentar tres situaciones:

- Que se pueden obtener estimaciones únicas de los parámetros estructurales en cuyo caso diremos que la ecuación esta *exactamente identificada*
- Que se obtengan mas de una estimación de los parámetros estructurales (pero un numero finito de los mismos) en cuyo caso se dice que la ecuación esta *sobreidentificada*
- Que se obtengan infinitos valores para las estimaciones de los parámetros estructurales y por ende haciendo imposible su recuperación, diciéndose en este caso que la ecuación está *subidentificada*. Esto surge por la falta de información necesaria para construir una función a través de la forma reducida

Supongamos que mercado con ofertas y demandas lineales donde cantidad y precio son endógenos sin tener ninguna variable predeterminada. En este caso, las dos rectas serán subidentificadas porque lo que se obtendrán serán sólo puntos de corte, sin saber si es una misma función de oferta con distintas demandas o viceversa, o si son funciones distintas en cada punto. La situación sería otra si por ejemplo se toma como variable exógena relevante el ingreso de los consumidores explicando su demanda. De esta manera, puede asegurarse que cada punto pertenece a una misma oferta en corte con diferentes demandas según sea el ingreso dado.

A continuación se analizará condetenimiento el problema de la Identificación y se daran definiciones mas precisas de lo enunciado anteriormente..

El problema de la Identificación

Mas precisamente, sea \mathbf{Y} u vector observables de variables aleatorias, una estructura S es una especificación completa de la función de densidad de \mathbf{y} , digamos $f(\mathbf{y})$. El conjunto de todas las estructuras posibles, \mathcal{S} es llamado un Modelo. El problema de la identificación consiste en hacer inferencias sobre S dada \mathcal{S} y las observaciones. Para hacer las cosas mas precisas supongamos que \mathbf{y} es generado por una función de densidad paramétrica:

$$f(\mathbf{y} / S) = f(\mathbf{y} / \alpha)$$

Donde α es un vector real k -dimensional. La función f se supone conocida pero α no. Entonces la estructura es descrita mediante un punto en \mathcal{R}^k mientras que el Modelo como un subconjunto de \mathcal{R}^k .

En general se dice que dos estructuras $\mathbf{S} = \alpha$ y $\mathbf{S} = \alpha^*$ son observacionalmente equivalentes si

$$f(\mathbf{y} / \alpha) = f(\mathbf{y} / \alpha^*), \quad \forall \mathbf{y}$$

En base a esas consideraciones se procederá a analizar las condiciones bajo las cuales dos estructuras de sistemas de ecuaciones son observacionalmente equivalentes.

$$\mathbf{Y}_t \Gamma + \mathbf{X}_t \mathbf{B} + \mathbf{u}_t = 0 \quad t = 1, \dots, T$$

Donde \mathbf{Y}_t , \mathbf{X}_t y \mathbf{u}_t , son la t -ésima filas de \mathbf{Y} , \mathbf{X} y \mathbf{u} respectivamente. Supongamos que \mathbf{u}_t tiene densidad $P(\mathbf{u}_t / \Sigma)$, luego la densidad conjunta asociada a $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_T$ será:

$$\prod_{t=1}^T P(\mathbf{u}_t / \Sigma)$$

Haciendo cambio de variables por sustitución de \mathbf{u}_t y multiplicando por el Jacobiano de la transformación, la densidad conjunta de $\mathbf{y}_1, \dots, \mathbf{y}_T$ condicionada en los parámetros será:

$$\begin{aligned} P(\mathbf{y}_1, \dots, \mathbf{y}_T / \Gamma, \mathbf{B}, \Sigma, \mathbf{X}) &= \|\Gamma\|^T \prod_{t=1}^T P(\mathbf{u}_t / \Sigma) \\ &= \|\Gamma\|^T \prod_{t=1}^T P(-\mathbf{y}_t \Gamma - \mathbf{x}_t \mathbf{B} / \Sigma) \end{aligned}$$

Si pre-multiplicamos el sistema original por una matriz no singular arbitraria \mathbf{F} , cada ecuación será reemplazada por una combinación lineal de cada una de las M ecuaciones originales. Esto es:

$$\mathbf{Y}_t (\Gamma \mathbf{F}) + \mathbf{X}_t (\mathbf{B} \mathbf{F}) + \mathbf{u}_t \mathbf{F} = \mathbf{Y}_t \Gamma^* + \mathbf{X}_t \mathbf{B}^* - \boldsymbol{\omega}_t = \mathbf{0} \quad t = 1, \dots, T$$

Donde $\boldsymbol{\omega}$ es $\mathbf{u}_t \mathbf{F}$,

$$P(\boldsymbol{\omega}_t / \mathbf{F}' \Sigma \mathbf{F}) = \|\mathbf{F}\|^{-1} P(\mathbf{u}_t / \Sigma)$$

la densidad conjunta de $\mathbf{y}_1, \dots, \mathbf{y}_T$ condicionada bajo la nueva estructura será:

$$\begin{aligned} \|\Gamma \mathbf{F}\|^T \prod_{t=1}^T P(\boldsymbol{\omega}_t / \mathbf{F}' \Sigma \mathbf{F}) &= \|\Gamma\|^T \|\mathbf{F}\|^T \|\mathbf{F}\|^{-T} \prod_{t=1}^T P(\mathbf{u}_t / \Sigma) \\ &= \|\Gamma\|^T \prod_{t=1}^T P(\mathbf{u}_t / \Sigma) \end{aligned}$$

Con lo cual las funciones de densidad de ambas estructuras son idénticas en consecuencia observacionalmente equivalentes. Obsérvese que eligiendo $F = B^{-1}$, una estructura observacionalmente equivalente es la forma reducida del sistema de ecuaciones. A raíz de esto, todas las estructuras creadas por una postmultiplicación de la estructura original por una matriz original tendrán la misma forma reducida.

De esta manera las condiciones para equivalencia observacional se pueden establecer así:

$$1) \Gamma^{-1}B = \Gamma^{*-1} B^*, \quad \Gamma^{-1}\Sigma\Gamma = \Gamma^{*-1} \Sigma^* \Gamma^*$$

2) existe F no singular,

$$\begin{bmatrix} \Gamma^* \\ B^* \end{bmatrix} = F \begin{bmatrix} \Gamma \\ B \end{bmatrix}$$

Si no hay restricciones a priori en los parámetros del sistema cualquier matriz no singular F podría ser *admissible* en el sentido que la estructura transformada satisface las restricciones del modelo. Si sobre la base de la teoría económica podemos establecer restricciones a priori cualquier modelo transformado debe satisfacer las mismas restricciones si la transformación se aún admisible. Se dice que la i -ésima ecuación del sistema está identificada si y solo si todas las transformaciones admisibles F tienen la siguiente estructura:

$$F = \begin{bmatrix} 0 & & \\ & \vdots & \\ \cdots & c_{ij} & \cdots \\ & \vdots & \\ & 0 & \end{bmatrix}$$

Es decir la i -ésima columna de F debe ser un múltiplo escalar del vector unitario cuyos elementos son todos ceros excepto la i -ésima columna que contiene un uno. Obsérvese que cuando postmultiplicamos B y G por F sus i -ésimas columnas solo son cambiadas por un múltiplo escalar.

Extendiendo dicha definición al sistema completo, se tiene que el sistema es identificado si y solo si todas las ecuaciones están identificadas, o equivalentemente la matriz admisible de transformación es diagonal.

A los efectos de determinar las condiciones para identificar una ecuación, definamos:

$$A = \begin{bmatrix} \Gamma \\ B \end{bmatrix}$$

De modo tal que a_i , la i -ésima columna de A contenga los parámetros de la i -ésima ecuación. Sea R_i y r_i $J \times (M + K)$ y $(J \times 1)$ matrices de constantes, respectivamente, tal que toda la información a priori acerca de la ecuación i -ésima del modelo, incluyendo la restricción de normalización, pueda ser escrita como:

$$R_i a_i = r_i$$

Así se puede hacer la siguiente observación: Sabemos que la i -ésima columna de AF es Af_i , donde f_i es la i -ésima columna de F . Ya que F es admisible, los coeficientes estructurales resultantes satisfacen la restricción $R_i(Af_i) = r_i$. Sin embargo, para que la i -ésima ecuación esté identificada, f_i debe ser igual a j_i , el i -ésimo vector unitario. Efectivamente j_i es una solución de $R_i A j_i = r_i$ porque $R_i A j_i = R_i a_i = r_i$. Entonces para que $(R_i A) f_i = r_i$ tenga solución única igual a $f_i = j_i$, se requiere que $\text{Rango}(R_i A) = M$.

Teorema: Condición de Rango:

La i -ésima ecuación está identificada si y solo si $\text{Rango}(R_iA) = M$

Corolario: Condición de Orden

Una condición necesaria para la identificación de la i -ésima ecuación es que J , el número de restricciones lineales incluida la normalización, debe ser mayor o igual que el número de variables endógenas del sistema M . Si R_i excluye la condición de normalización, entonces el rango de R_iA debe ser $M - 1$ y el orden de R_i debe ser $J > M - 1$.

Si la normalización incluye las restricciones, entonces el teorema y el corolario conducen a las siguientes definiciones:

- 1.- Los parámetros de la i -ésima ecuación están sin identificar o *subidentificados* si $\text{Rango}(R_iA) < M$.
- 2.- Los parámetros de la i -ésima ecuación están *perfectamente identificados* si $\text{Rango}(R_iA) = M$ y $\text{Rango}(R_i) = M$
- 3.- Los parámetros de la i -ésima ecuación están *sobreidentificados* si $\text{Rango}(R_iA) = M$ y $\text{Rango}(R_i) > M$

Una versión alternativa y más práctica de la condición de rango es la siguiente:

Una ecuación está identificada sólo si se puede construir por lo menos un determinante diferente de cero de orden $(M-1) \times (M-1)$, a partir de los coeficientes de las variables endógenas y predeterminadas excluidas de esa ecuación, pero incluidas en las demás².

Métodos de Estimación

Habiendo ya señalado las condiciones que hacen factible la recuperabilidad de los parámetros se pasará a continuación a dar un breve comentario acerca de los métodos existentes que permiten solucionar el problema de la inconsistencia que presentan la aplicación directa de los MCO. Estos métodos surgen a partir de la búsqueda de soluciones alternativas cuando se viola el supuesto de no aleatoriedad en las variables explicativas y , a su vez, de la no existencia de correlación igual a cero entre ellas y las perturbaciones.

Métodos uniecuacionales o de información limitada: En estos métodos, se estima cada ecuación separadamente, utilizando sólo la información sobre los coeficientes contenida en dicha ecuación. Entre estos métodos están:

Mínimos Cuadrados Ordinarios (MCO): Como ya se señaló antes, los estimadores que surgen de aplicar el método **no son consistentes**. Sin embargo, este método es más robusto frente a los errores de especificación que otros métodos, y además las predicciones de modelos estimados por él pueden ser mejores que las correspondientes a modelos estimados por métodos de ecuaciones simultáneas, por lo que puede resultar conveniente presentar estimaciones MCO de las ecuaciones estructurales junto a las de otros métodos como referencia o norma de comparación.

Mínimos Cuadrados Indirectos (MCI): consiste en estimar por MCO los coeficientes de la forma reducida y luego recuperar los estimadores de los parámetros de la forma estructural vía un sistema de ecuaciones. Sólo es aplicable cuando las ecuaciones del modelo están exactamente identificadas, ya que sólo en éste caso se puede obtener valores únicos para los coeficientes de la forma reducida. Los estimadores de los parámetros así obtenidos heredan todas las propiedades asintóticas de los estimadores de la forma reducida, asegurándonos así de que sean consistentes (y pueden ser asintóticamente eficientes si las perturbaciones se distribuyen normalmente) pero no gozan de éstas propiedades para muestras chicas.

² Ver Gujarati, Cap. 17 Sección 3 Pág: 362

Mínimos Cuadrados en Dos Etapas (MC2E): Es útil para modelos cuyas ecuaciones están sobreidentificadas, ya que presenta una forma de ponderar las soluciones múltiples. La idea básica que subyace en el método es la de reemplazar las variables endógenas, que están correlacionadas con las perturbaciones, por funciones lineales de variables exógenas; así, ya que estas variables no presentan correlación con dichas perturbaciones, las estimaciones de los parámetros que conseguiremos serán **consistentes**. Se elige cada función lineal de manera que esté lo más altamente correlacionada posible con la variable endógena que reemplaza.

El método requiere dos aplicaciones sucesivas de MCO.

Primeramente se regresan las variables endógenas contra todas las variables predeterminadas del sistema. La estimación obtenida para la variable endógena se utiliza como dato en las ecuaciones y se aplica nuevamente MCO sobre la ecuación que incluye la estimación. Los estimadores así obtenidos son consistentes.

Además se puede aplicar tanto para ecuaciones sobreidentificadas como exactamente identificadas. En este último caso los estimadores serán idénticos a los obtenidos por Mínimos Cuadrados Indirectos.

Se puede demostrar que éste método es una aplicación particular del método de las variables instrumentales que busca reemplazar los regresores por variables altamente correlacionados con ellos y escasamente con los términos de perturbación.

Máxima Verosimilitud con Información Limitada (MVIL)

Este método consiste en elegir como estimadores de los parámetros estructurales aquellos valores que hacen máxima la función de verosimilitud de la ecuación respectiva sujeto a elegir valores que cumplan con la condición de rango. Es decir este método ante los caso de sobreidentificación elige dentro del grupo de valores posibles aquellos que maximicen la función de verosimilitud de la ecuación siendo un requisito previo el conocimiento de la distribución de densidad del término de perturbación de la ecuación.

Además se puede aplicar tanto para ecuaciones sobreidentificadas como exactamente identificadas. En este último caso los estimadores serán idénticos a los obtenidos por Mínimos Cuadrados Indirectos.

Métodos de sistemas o de información completa: En estos métodos, a diferencia de los anteriores, se estiman en conjunto todas las ecuaciones del sistema, usando las restricciones sobre los parámetros de todas ellas. Son asintóticamente más eficientes que los métodos con información limitada en la medida que la especificación del modelo sea la correcta.

Mínimos Cuadrados en Tres Etapas: consiste en estimar cada una de las ecuaciones del modelo por el método de los MC2E, calcular los residuos para estimar la matriz de varianzas y covarianzas (aquí se suponen ruido blanco en cada término de perturbación de cada ecuación y se permite la existencia de correlación contemporánea y heterocedasticidad entre las ecuaciones) para finalmente aplicar mínimos cuadrados generalizados factibles al modelo completo. Este método es más eficiente asintóticamente que el de MC2E en la medida que la especificación del modelo sea la correcta y tendrá igual eficiencia en el caso de que no exista correlación contemporánea ni heterocedasticidad entre los errores de cada una de las ecuaciones, o en el caso que todas las ecuaciones estén exactamente identificadas.

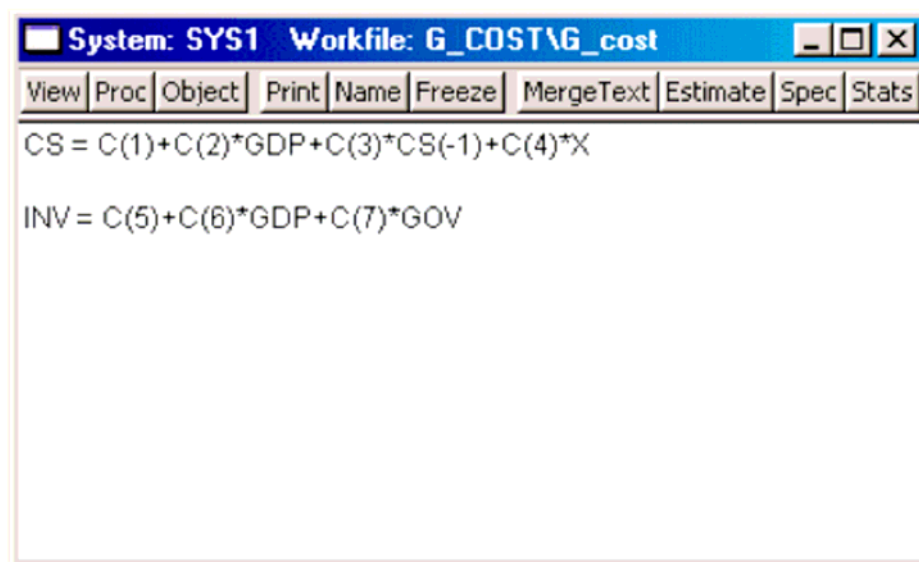
Máxima verosimilitud con Información completa: Aquí lo que se pretende es hallar el conjunto de estimadores de la forma estructural que hacen máxima la probabilidad de ocurrencia de los valores muestrales de las variables endógenas y predeterminadas, es decir se intenta maximizar una función de verosimilitud para todo el modelo simultáneamente. Este método es caro computacionalmente en el sentido que arroja sistemas de ecuaciones no lineales y sumado al hecho de que los estimadores que se obtienen bajo una distribución normal multivariante son exactamente iguales a los de los de MC3E hacen que este método sea menos preferido que éste último.

Además se puede aplicar tanto para ecuaciones sobreidentificadas como exactamente identificadas. En este último caso los estimadores serán idénticos a los obtenidos por Mínimos Cuadrados Indirectos.

Implementación en Eviews:

Creando el Objeto System

Para estimar los parámetros de un Sistema de Ecuaciones en Eviews, se debe primero crear un nuevo objeto y especificar el sistema de ecuaciones. Para ello se debe hacer Clic en Objet/New Objet/System. Cuando se cree el sistema la ventana aparecerá en blanco y ahí se deberá especificar el sistema describiendo con texto las ecuaciones y potencialmente algunas líneas describiendo los instrumentos y los valores iniciales de los parámetros en caso de modelos no lineales.



Declaración de las ecuaciones

Para ingresar las ecuaciones por fórmulas se debe emplear las expresiones estándares de Eviews. Las ecuaciones deben ser ecuaciones de comportamiento con coeficientes desconocidos y términos de error implícitos. Las ecuaciones pueden ser no lineales en sus variables, coeficientes o ambas, pudiéndose agregar restricciones cruzadas en los coeficientes de las ecuaciones usando los mismos coeficientes en diferentes ecuaciones. Por ejemplo:

$$y = c(1)*x1 + c(2)*x2 + c(3)*x3$$

$$z = c(3) + c(2)*z + (1-c(2))*x$$

Si se desea agregar la restricción $C(1)+C(2)+C(3)=1$, la misma se puede imponer especificando la ecuación como:

$$y = c(1)*x1 + c(2)*x2 + (1-c(1)-c(2))*x3$$

Si la ecuación no posee el signo igual por ejemplo

$$(c(1)*x + c(2)*y + 4)^2$$

Eviews interpretará la expresión completa igual al término de error.

No deben incluirse en la especificación ecuaciones que representen identidades. De ser necesario se deberá resolver el modelo para eliminar de él las identidades y especificar en Eviews el modelo resuelto libre de ellas.

Variables Instrumentales

Si se planea estimar el sistema usando Mínimos Cuadrados en 2 Etapas, 3 Etapas o GMM, se deben especificar las variables a ser usadas en la estimación. Se pueden especificar los mismos instrumentos para todas las ecuaciones o instrumentos distintos para cada ecuación.

De esta manera existen dos maneras de especificar los instrumentos. Si uno desea usar los mismos instrumentos para todas las ecuaciones se debe incluir una línea de comando cuya sintaxis comience con @INST o INST seguida por una lista de variables exógenas a ser usadas como instrumentos. Por ejemplo:

@inst gdp(-1 to -4) x gov

Indica a Eviews usar las seis variables como instrumentos para todas las ecuaciones que se listarán a continuación.

Para especificar instrumentos distintos para cada ecuación se debe agregar al fin de cada ecuación: @, seguido por la lista de instrumentos. Por ejemplo:

cs = c(1)+c(2)*gdp+c(3)*cs(-1) @ cs(-1) inv(-1) gov

inv = c(4)+c(5)*gdp+c(6)*gov @ gdp(-1) gov

La primer ecuación usa CS(-1), INV(-1), GOV, y una constante como instrumentos mientras que la segunda usa GDP(-1), GOV, y una constante.

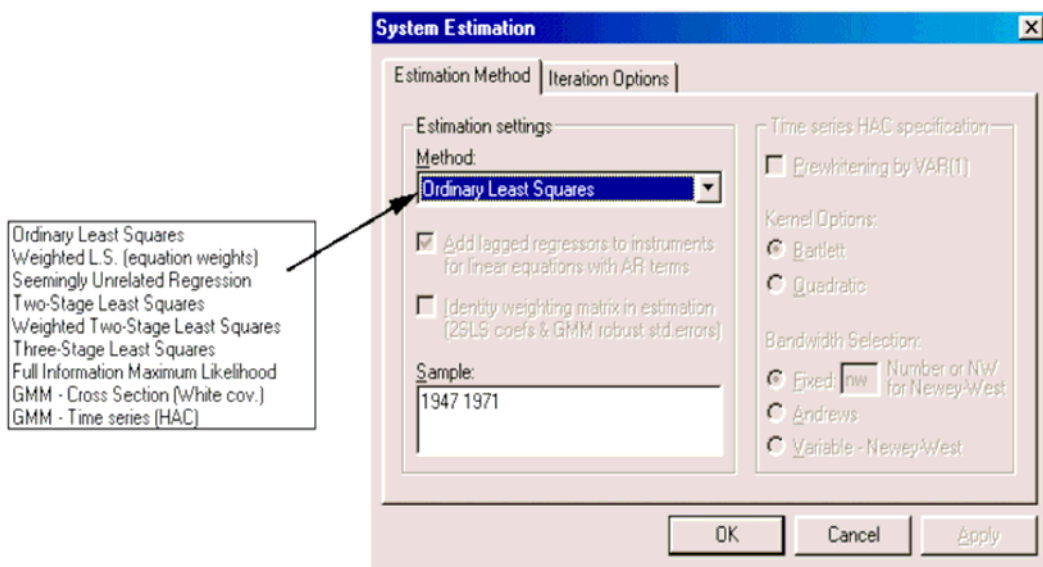
Identificación

Los criterios de identificación exigen que debe haber al menos tantos instrumentos, incluyendo constantes, en cada ecuación como variables existen del lado derecho de cada una de ellas.

Si dichos requerimientos no se cumplen, Eviews mostrará un cuadro de error.

Estimación

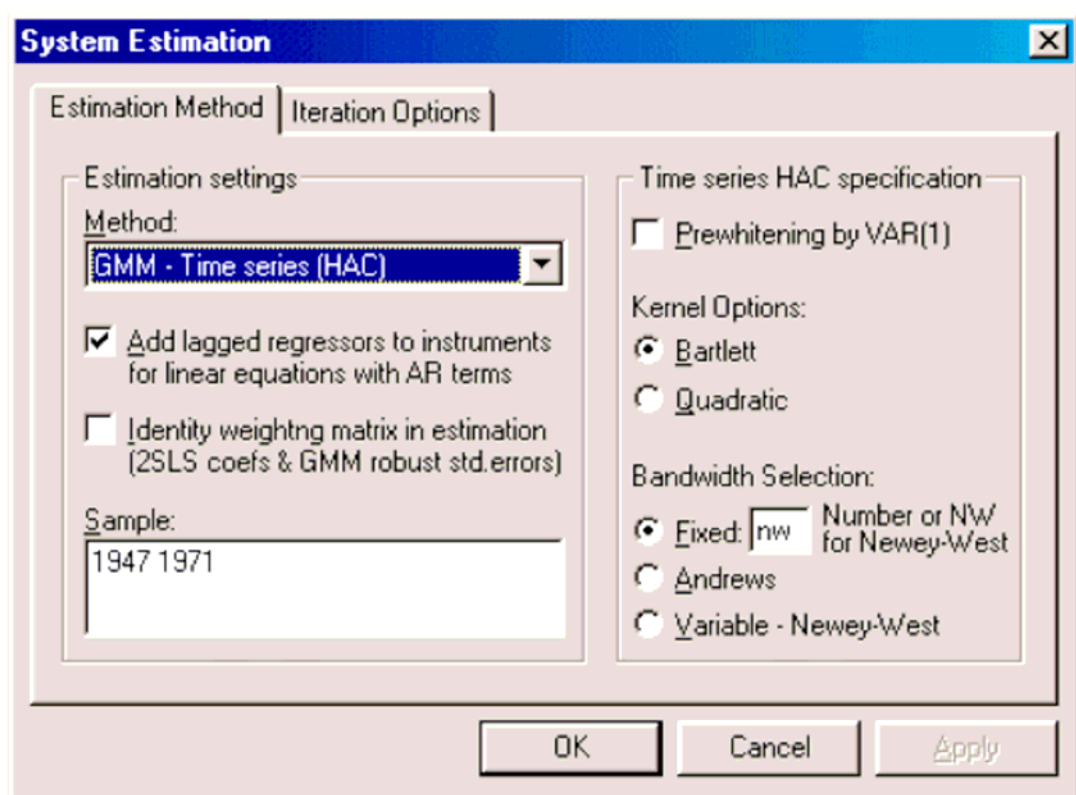
Una vez creado y especificado el sistema, uno puede presionar el botón de estimación para seleccionar el método de estimación por medio de un cuadro de dialogo



Si uno selecciona alguno de los métodos de estimación por GMM, Eviews solicitará algunas opciones específicas para este método. Si uno selecciona la opción rotulada como Identity

weighting matrix, Eviews estimará el modelo usando idénticos ponderadores y usará los coeficientes estimados y la especificación GMM que se indique para computar una matriz de covarianzas que es robusta a heterocedasticidad en las secciones cruzadas (White) o a heterocedasticidad y correlación contemporánea. Si dicha opción no se selecciona Eviews usará los ponderadores GMM en la estimación y cómputo de lo matriz de covarianzas.

Cuando uno selecciona la opción GMM-Time series (HAC), el cuadro de diálogo adicionalmente mostrará opciones para especificar la matriz de ponderación y opciones para el Kernel para determinar la forma funcional del nucleo usado para ponderar las autocovarianzas al computar la matriz de ponderación.



Salidas de las Estimaciones

Las salidas de la estimación del sistema contiene los parámetros estimados, sus errores estándares y los estadísticos t para cada uno de los coeficientes en el sistema. Adicionalmente Eviews reportará el determinante de la matriz de residuos y en el caso del método de Máxima Verosimilitud con Información Completa, el valor maximizado de la verosimilitud. También se incluyen un resumen de los estadísticos de cada ecuación como los errores estándares de la regresión, suma de cuadrados de residuos, etc.

APÉNDICE

1.- Mínimo Cuadrados Generalizados

Hasta ahora se han tratado casos de estimación de modelos lineales en donde los supuestos en los que se generaban los datos respondían a todas las condiciones que hacían aplicables el Teorema de Gauss-Markov arrojando por ende estimadores insesgados y eficientes. Dichos supuestos eran:

- 1- Las X 's son fijas en muestras repetidas (no estocásticas).
- 2- El rango de la matriz X es de rango completo (no existe dependencia lineal entre las observaciones)
- 3- Correcta especificación del modelo en el sentido de que el verdadero proceso generador de datos provenía de una relación exactamente lineal entre las variables y además que no se cometen errores por incluir o excluir variables de más.
- 4- $E(\mathbf{u}) = \mathbf{0}$
- 5- $V(\mathbf{u}) = \sigma^2 \mathbf{I}$ es decir, errores no autocorrelacionados y con varianza constante e igual para todos (homoscedásticos).

Cuando este último supuesto no se cumple la aplicación directa de los MCO conducen a estimadores que dejan de ser eficientes. Además cuando se comete el error de no percatarse o de hacer caso omiso a la existencia de autocorrelación y/u homoscedasticidad se incurrirá además en el error de subestimar la varianza del estimador conllevando esto último a predicciones erróneas en cuanto a intervalos de confianza y test de hipótesis. A continuación se desarrollarán estas ideas de un lenguaje más formal.

Consideremos un modelo como del ejercicio en donde ahora

$$V(\mathbf{u}) = \sigma^2 \mathbf{\Omega}$$

aplicando directamente MCO al modelo se tendrá

$$\hat{\beta}_{mco} = (X'X)^{-1} X'Y, \text{ reemplazo por el VM :}$$

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

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

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

$$E(\hat{\beta}_{mco}) = \beta$$

que como claramente se observa los estimadores conservan aun su insesgabilidad

Con respecto a su varianza se tiene

$$V(\hat{\beta}_{mco}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

$$V(\hat{\beta}_{mco}) = E[(X'X)^{-1} X'uu'X(X'X)^{-1}]$$

$$V(\hat{\beta}_{mco}) = (X'X)^{-1} X'E(uu')X(X'X)^{-1}$$

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

esta expresión para la varianza de los estimadores. A continuación se verá que existen otros estimadores de los parámetros que poseen una varianza menor.

Para que el teorema de Gauss-Markov continúe en vigencia lo que habría que hacer intentar hallar una transformación de los datos originales de modo tal que los errores asociados a dichos datos transformados no exhiban autocorrelación ni heteroscedasticidad. La existencia de dicha transformación será el hecho que marque la entrada en vigencia del Teorema de Gauss-Markov en el sentido de que la aplicación de MCO a los datos transformados generará estimadores insesgados y eficientes de los

parámetros. Ya que Ω es una matriz definida positiva, por ser esta una matriz de varianzas y covarianzas, siempre existirá otra matriz tal que permita expresarla como sigue

$$\Omega = HH' \rightarrow \Omega^{-1} = (H')^{-1}H^{-1} \text{ ya que la existencia de } H \text{ garantiza la existencia de } H^{-1}$$

Si ahora se premultiplica el modelo por H^{-1} se tendrá :

$$H^{-1}Y = H^{-1}X\beta + H^{-1}u$$

$$Y^* = X^*\beta + u^*$$

$$E(u^*) = H^{-1}E(u) \rightarrow E(u^*) = \theta$$

$$V(u^*) = E(u^*u^{*\prime})$$

$$V(u^*) = E[H^{-1}uu'(H')^{-1}] = H^{-1}E(uu')(H')^{-1}$$

$$V(u^*) = \sigma^2 H^{-1}\Omega(H')^{-1} = \sigma^2 H^{-1}HH'(H')^{-1}$$

$$V(u^*) = \sigma^2 I$$

con lo cual se logra hallar la transformación de los datos adecuada para eliminar la heteroscedasticidad y la autocorrelación, habilitando la aplicación de MCO a los datos transformados para obtener estimadores de los parámetros insesgados y eficientes tal como lo predice el Teorema de Gauss-Markov. Estos nuevos estimadores son denominados estimadores por mínimos cuadrados generalizados (MCG).

La expresión para los estimadores (MCG), la prueba de que siguen siendo insesgados y la expresión que adoptan su matriz de varianzas y covarianzas se muestra como sigue:

$$\hat{\beta}_{mcg} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$$

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

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

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

$$E(\hat{\beta}_{mcg}) = \beta \rightarrow \text{sigue siendo insesgado}$$

$$V(\hat{\beta}_{mcg}) = E[(\hat{\beta}_{mcg} - \beta)(\hat{\beta}_{mcg} - \beta)']$$

$$V(\hat{\beta}_{mcg}) = E[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}uu'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}]$$

$$V(\hat{\beta}_{mcg}) = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}E(uu')\Omega^{-1}X(X'\Omega^{-1}X)^{-1}$$

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

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

Esta expresión a la que se arriba es la varianza correspondiente a los estimadores de varianza mínima (por el teorema de Gauss-Markov) por ende será inferior que la de aplicar MCO a los datos sin transformar. Este hecho se puede demostrar vía el cálculo de la diferencia entre ambas arrojando esto una matriz semidefinida positiva.

Para hallar la matriz H que transforme los datos se pueden seguir varias maneras. Una de ellas sería aplicando una descomposición de Cholesky a Ω , otra sería planteando un sistema de ecuaciones. Notese además que es esencial conocer con exactitud Ω para en función de esto obtener H ya que sería imposible estimar todos sus elementos al mismo tiempo si estos fuesen todos distintos en virtud de que sería más los parámetros a estimar que los datos con que se cuentan para hacerlo.

Una consideración importante sobre el tema es que cuando se comete el error de no percatarse o de hacer caso omiso a la existencia de autocorrelación y/u homoscedasticidad se incurrirá además en

el error de subestimar la varianza del estimador conllevando esto último a predicciones erróneas en cuanto a intervalos de confianza y test de hipótesis que se hagan sobre los estimadores.

$$V(\hat{\beta}_{mco} \text{ Falsa}) = \sigma^2 (X'X)^{-1} < V(\beta_{mcg}) < V(\beta_{mco})$$

como se dijo en el párrafo anteprecedente ante la imposibilidad de obtener estimaciones sobre todos los parámetros de la matriz ω surge la necesidad de hacer algunas suposiciones adicionales sobre el tipo de autocorrelación como la que se realiza en el ejercicio.

a) También en estos casos puede suceder que el estimador de la varianza muestral sea sesgado con respecto al valor del parámetro. Generalmente lo subestima, convirtiéndose esto en una grave peligro a la hora de realizar las inferencias. A continuación se efectúa una demostración del mismo

$$e = Y - X \hat{\beta}$$

$$e = Y - X (X'X)^{-1} X'Y$$

$$e = [I - X (X'X)^{-1} X'] (X\beta + u),$$

donde $[I - X (X'X)^{-1} X'] = M$, simétrica e idempotente

$$e = MX\beta + Mu = Mu \text{ ya que puede demostrarse que } MX = [I - X (X'X)^{-1} X']X = 0$$

$$e'e = u'M'Mu = u'Mu$$

$$E(e'e) = E[tr(u'Mu)] = E[tr(uu'M)]$$

$$E(e'e) = tr[E(uu')M] = \sigma^2 tr(\Omega M)$$

$$E\left(\frac{e'e}{n-k}\right) = E(\hat{\sigma}^2) = \sigma^2 \frac{tr(\Omega M)}{n-k}$$

La traza de (ΩM) es distinta de $(n-k)$ siempre que $\Omega \neq I$, por lo tanto el estimador de la varianza residual $(e'e/(n-k))$ es sesgado.

Para concluir vale destacar que prevenirse de la autocorrelación y/o heteroscedasticidad (varianzas de los errores distintas entre sí) es una tarea de suma importancia (no trivial), porque permitirá imponer barreras a todos los peligros de la estimación clásica por MCO. Para esto, hay tener en cuenta que son diversas las causas que llevan a la pérdida de la propiedad de ruido blanco a los errores. Entre estas están: el error en la especificación de la forma funcional, omisión o inclusión variables relevantes, etc. Para conocer si es lógico o no el supuesto de ruido blanco de los errores se puede contar con la observación de los digramas de dispersión, con los correlogramas y con algunos test de hipótesis como el de Durbin Watson.

Para el caso de esquemas autorregresivos de primer orden, existe un estadístico que permite determinar si éste está presente entre las perturbaciones o no; dicho estadístico es el de Durbin-Watson y se define así:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

Podemos definirlo alternativamente mediante algunos pasos algebraicos como:

$$d = \frac{\sum e_t^2 + \sum e_{t-1}^2 - 2\sum e_t e_{t-1}}{\sum e_t^2}$$

$$d = 2 \left(1 - \frac{\sum e_t e_{t-1}}{\sum e_t^2} \right)$$

y siendo

$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2}$$

entonces

$$\boxed{d = 2(1 - \hat{\rho})}$$

Hay que tener en cuenta que el estadístico se basa en algunos supuestos sobre el modelo, a saber, que el mismo incluye un término independiente, que las variables X son fijas en muestras repetidas, que las perturbaciones u se generan mediante un proceso autorregresivo de primer orden y se distribuyen en forma normal multivariante y por último que el modelo no incluye valores rezagados de la variable dependiente como una de las variables explicativas.

Para poder rechazar o no la hipótesis nula de que no hay autocorrelación entre las perturbaciones, se han tabulado valores críticos máximos y mínimos para poder tomar una decisión. Estos límites dependen del número de observaciones y del número de variables explicativas del modelo. No desarrollará aquí el criterio de decisión para estas pruebas de hipótesis pero sí se pueden reparar brevemente en los límites de variación del estadístico. Dependiendo de los valores que asuma $\hat{\rho}$, (y siendo que $-1 < \hat{\rho} < 1$):

$$0 < d < 4$$

2.- El método SUR

Ante determinadas situaciones el investigador puede verse en la necesidad de modelar y estimar conjuntamente varias ecuaciones que en apariencia no representen simultaneidad entre las mismas. Sin embargo, los errores aleatorios pueden presentar algún grado de correlación contemporánea en la medida que involucren a factores comunes no medibles y/o no observables y será esta correlación no percibida la que haga que resulte más eficiente estimar todas las ecuaciones simultáneamente y no una por una por Mínimos Cuadrados Ordinarios.

En la teoría económica existen innumerables ejemplos de estimaciones conjuntas. Se puede tratar de ajustar la demanda de determinado bien a lo largo del tiempo y en diferentes regiones. Se podrían, a su vez, encontrar las demandas de diferentes bienes que están relacionados (ya bien como sustitutos o complementarios), caso éste del ejercicio a realizar. La inversión en el tiempo de distintas firmas es otro caso típico.

La correlación del término de perturbación de distintas ecuaciones en un momento del tiempo es conocida como correlación contemporánea (distinta a la autocorrelación, que es la correlación en el tiempo en una misma ecuación).

La técnica apropiada de estimación conjunta es conocida como Seemingly Unrelated Regressions -Ecuations- (SUR), o regresiones aparentemente no relacionadas.

El método SUR es utilizado por lo general en la estimación de un conjunto de ecuaciones utilizando series de tiempo, pero es igualmente útil para datos de corte transversal. El método Sur puede ser considerado como un método para combinar datos de series de tiempo y corte transversal.

De una manera más formal la situación de modelización que propicia la aplicación de este nuevo método es:

$$y_i = X_i \beta_i + e_i \quad \text{con } i = 1, 2, \dots, M, \text{ donde } M = \text{cantidad de ecuaciones}$$

donde claramente se observa las diferencias con los problemas tradicionales de estimación, ya que ahora se debe relacionar series de tiempo con cortes transversales, existiendo no sólo la posibilidad de autocorrelación entre los errores para cada observación de una ecuación “i”, sino que además pueden ocurrir interrelaciones entre las perturbaciones de dos ecuaciones (vector e_i contra el vector e_j).

MÉTODO SUR

Dado $y_i = X_i \beta_i + e_i$ con $i = 1, 2, \dots, M$. Donde cada y_i es $(T \times 1)$, la matriz X_i es de $(T \times K_i)$, los β_i son de $(K_i \times 1)$ y las perturbaciones de $(T \times 1)$. Dados los tamaños, el modelo se puede reducir a:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix} \rightarrow Y_{(TM \times 1)} = X_{(TM \times K)} \beta_{(K \times 1)} + e_{(TM \times 1)}$$

Donde K surge de la suma de los K_i de cada ecuación.

Cada ecuación, como se dijo anteriormente, posee ruido blanco en los disturbios, pudiéndose estimar por MCO separadamente. El problema surge cuando al hacer esto, no consideramos las relaciones entre las ecuaciones (correlación contemporánea). Para encontrar entonces la matriz de covarianzas de todo el sistema, a partir de la matriz (Σ) que contiene las varianzas de los disturbios de cada ecuación (σ_{ii}) , se deben recordar la propiedad de homoscedasticidad, y las covarianzas entre las ecuaciones $(\sigma_{ij}$ con $i \neq j$), que reflejan las correlaciones contemporáneas, con lo cual se tiene

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}$$

Haciendo el producto de kronecker (\otimes) entre la matriz anterior y una identidad de orden (T), se obtiene lo siguiente:

$$\Sigma \otimes I_T = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 & \sigma_{12} & 0 & \cdots & 0 & \cdots & \sigma_{1M} & 0 & \cdots & 0 \\ 0 & \sigma_{11} & \cdots & 0 & 0 & \sigma_{12} & \cdots & 0 & \cdots & 0 & \sigma_{1M} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{11} & 0 & 0 & \cdots & \sigma_{12} & \cdots & 0 & 0 & \cdots & \sigma_{1M} \\ \sigma_{12} & 0 & \cdots & 0 & \sigma_{22} & 0 & \cdots & 0 & \cdots & \sigma_{2M} & 0 & \cdots & 0 \\ 0 & \sigma_{12} & \cdots & 0 & 0 & \sigma_{22} & \cdots & 0 & \cdots & 0 & \sigma_{2M} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{12} & 0 & 0 & \cdots & \sigma_{22} & \cdots & 0 & 0 & \cdots & \sigma_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{1M} & 0 & \cdots & 0 & \sigma_{2M} & 0 & \cdots & 0 & \cdots & \sigma_{MM} & 0 & \cdots & 0 \\ 0 & \sigma_{1M} & \cdots & 0 & 0 & \sigma_{2M} & \cdots & 0 & \cdots & 0 & \sigma_{MM} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{1M} & 0 & 0 & \cdots & \sigma_{2M} & \cdots & 0 & 0 & \cdots & \sigma_{MM} \end{bmatrix}$$

A esta matriz la llamaremos Φ y representará la matriz de V y COV relevante. Si nosotros conocemos a Φ , podemos estimar los coeficientes por MCG, deduciendo del álgebra matricial que:

$$\hat{\beta} = [X'(\Sigma^{-1} \otimes I)X]^{-1} X'(\Sigma^{-1} \otimes I)y \Rightarrow \hat{\beta} = [X'\Phi^{-1}X]^{-1} X'\Phi^{-1}y$$

El nuevo estimador posee una menor varianza ya que tiene en cuenta la correlación contemporánea entre los distintos vectores de las perturbaciones de las diferentes ecuaciones

Sin embargo existen dos casos particulares en que se obtienen resultados idénticos al aplicar MCO sobre cada ecuación y al aplicar SUR y no se producen ganancias al tratar a las ecuaciones como sistema.

- a) Cuando las correlaciones contemporáneas son iguales a "0". Ello es obvio, ya que es la existencia de dicha correlación lo que provoca que las ecuaciones estén relacionadas.

$$\sigma_{11} = \sigma_{22} = \sigma_{33} = 0$$

b) Cuando las variables explicatorias de cada ecuación son las mismas.

$$X_1 = X_2 = X_3 = \bar{X}$$

Las variables explicativas de todas las ecuaciones son las mismas, por lo que no se produce una ganancia al tratarlas en forma conjunta, ya que no se agrega ninguna variable explicativa a cada variable dependiente.

Por lo tanto, la estimación a través de MCO nos proporciona la misma solución que el método SUR. Ello se puede demostrar partiendo del estimador SUR, utilizando algunas propiedades del producto de Kronecker y observando que, al ser todas las matrices X_i iguales, entonces $X = (I \otimes \bar{X})$. Por lo tanto:

$$\begin{aligned} \hat{\beta}_{SUR} &= (X'(\Sigma^{-1} \otimes I)X)^{-1} X \odot (\Sigma^{-1} \otimes I) y \\ &= ((I \otimes \bar{X})'(\Sigma^{-1} \otimes I)(I \otimes \bar{X}))^{-1} (I \otimes \bar{X})'(\Sigma^{-1} \otimes I) y \\ &= (\Sigma^{-1} \otimes (\bar{X}'\bar{X}))^{-1} (\Sigma^{-1} \otimes I) y \\ &= (\Sigma \otimes (\bar{X}'\bar{X})^{-1})(\Sigma^{-1} \otimes I) y \\ &= (I \otimes (\bar{X}'\bar{X})^{-1} \bar{X}') y \\ &= ((I \otimes \bar{X})'(I \otimes \bar{X}))^{-1} (I \otimes \bar{X})' y \\ &= (XX)^{-1} X \odot y = \hat{\beta}_{MCO} \end{aligned}$$

Cuando no se conoce a la matriz de V y COV (Φ), se opta por construirla a través de los errores muestrales de la siguiente forma:

$$\hat{\beta}_{MCOi} = (X \odot_i X_i)^{-1} X \odot_i y_i \Rightarrow \hat{e}_i = y_i - X_i \hat{\beta}_{MCOi}$$

$$\hat{\sigma}_{ij} = \frac{1}{T} \hat{e}_i \hat{e}_j = \frac{1}{T} \sum_{t=1}^T \hat{e}_{it} \hat{e}_{jt}$$

Aunque sea sesgada en muestras pequeñas. Para evitar el sesgo, se suele dividir el sumatorio por $(T - K/M)$. La dificultad aparece cuando el modelo incluye ecuaciones con diferentes números de parámetros, ya que en dicho caso no se cuenta con iguales grados de libertad para las distintas ecuaciones, en especial al calcular las covarianzas entre dos

ecuaciones de distintos grados de libertad. Una alternativa interesante es utilizar como divisor $T - \bar{k}$, donde \bar{k} es la media del número de coeficientes de las ecuaciones.

Dicho divisor posee la ventaja de que si todas las ecuaciones poseen el mismo número de coeficientes, provee de estimadores insesgados. Si definimos a la matriz de los estimadores de las varianzas y covarianzas del modelo como $\hat{\Sigma}$, el estimador MCG del modelo cuando la matriz de varianzas y covarianzas es desconocida, es:

$$\hat{\beta}_{SUR} = (X'(\hat{\Sigma} \otimes I)X)^{-1} X'(\hat{\Sigma} \otimes I)y$$

Este estimador es conocido como el estimador Sur de Zellner y es el que se utiliza por lo general en la práctica.

Restricciones

1- Condición de Engel:

$$w_1\beta_{14} + w_2\beta_{24} + w_3\beta_{34} = 1$$

La condición de Engel establece que la suma de las elasticidades ingreso de cada bien, ponderadas por la participación en el gasto total ($p \times q / y$), debe ser igual a uno. Si se desarrollamos esta suma:

$$\left(\frac{p_1 q_1}{y} * \frac{\partial q_1}{\partial y} * \frac{y}{q_1} \right) + \left(\frac{p_2 q_2}{y} * \frac{\partial q_2}{\partial y} * \frac{y}{q_2} \right) + \left(\frac{p_3 q_3}{y} * \frac{\partial q_3}{\partial y} * \frac{y}{q_3} \right) = 1$$

$$\left(p_1 * \frac{\partial q_1}{\partial y} \right) + \left(p_2 * \frac{\partial q_2}{\partial y} \right) + \left(p_3 * \frac{\partial q_3}{\partial y} \right) = 1$$

Debemos observar si **la suma del gasto que en cada bien se produce, luego de un incremento en el ingreso, es igual a uno**. Es decir, que el incremento del gasto en los n bienes que se consumen deben agotar la totalidad del incremento en el ingreso.

$$\left(p_1 * \frac{\partial q_1}{\partial y} \right) dy + \left(p_2 * \frac{\partial q_2}{\partial y} \right) dy + \left(p_3 * \frac{\partial q_3}{\partial y} \right) dy = dy$$

2- Homogeneidad:

$$\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4} = 0 \quad \text{para } i = 1, 2, 3$$

Recordando que los coeficientes representan las elasticidades precio ($\beta_{i1}, \beta_{i2}, \beta_{i3}$) e ingreso (β_{i4}); debemos comprobar la propiedad de “**homogeneidad de grado cero**” de las funciones de demanda de este tipo para cada ecuación. La importancia de esta característica radica en su definición: Al multiplicar todos los precios por un valor “ δ ”

arbitrario, distinto de cero, la cantidad demandada no variará si también multiplicamos nuestro ingreso por el mismo valor. Económicamente la condición de homogeneidad indica que el consumidor no padece “ilusión monetaria” en el sentido de que a la hora de consumir solo se fija en precios reales y no nominales. Lo dicho anteriormente se puede traducir formalmente como sigue:

$$\begin{aligned}
 q_{it} &= A * p_{1t}^{\beta_{i1}} * p_{2t}^{\beta_{i2}} * p_{3t}^{\beta_{i3}} * y_t^{\beta_{i4}} \\
 q'_{it} &= A * (\delta \times p_{1t})^{\beta_{i1}} * (\delta \times p_{2t})^{\beta_{i2}} * (\delta \times p_{3t})^{\beta_{i3}} * (\delta \times y_t)^{\beta_{i4}} \\
 q'_{it} &= A * \delta^{(\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4})} * p_{1t}^{\beta_{i1}} * p_{2t}^{\beta_{i2}} * p_{3t}^{\beta_{i3}} * y_t^{\beta_{i4}}
 \end{aligned}$$

Se demuestra que al cumplirse la propiedad, el exponente de “ δ ” se anula dejando como consecuencia una identidad entre q_{it} y q'_{it} ; es decir, las cantidades demandadas de ese bien “i” en el momento “t” no variaron.

3- Simetría:

La condición de simetría establece que la elasticidad cruzada de un bien i con respecto a otro j, expresada como proporción de la participación en el ingreso del bien j, más la elasticidad-renta del bien i, debe ser igual a la elasticidad cruzada del bien j con respecto al bien i, expresada como proporción de la participación en el ingreso del bien i, más la elasticidad-renta del bien j., o dicho formalmente:

$$\frac{\beta_{ij}}{w_j} + \beta_{i4} = \frac{\beta_{ji}}{w_i} + \beta_{j4} \quad \text{para } i, j = 1, 2, 3 (i \neq j)$$

Donde: $w_i = p_i q_i / y$ y siendo $i = (1, 2, 3)$ representa la participación en el presupuesto del bien i-ésimo. Se debe utilizar: $\bar{w}_i = T^{-1} \sum_{t=1}^T (p_i q_{it} / y_t)$

SUR Iterativo

La metodología que se expuso para obtener los estimadores SUR cuando la matriz de varianzas y covarianzas era desconocida consistía en calcular en una primera etapa los estimadores MCO de cada una de las demandas, y utilizarlos para obtener los vectores de errores:

$$\hat{\beta}_{MCOi} = (X_i' X_i)^{-1} X_i' y_i \Rightarrow \hat{e}_i = y_i - X_i \hat{\beta}_{MCOi}$$

Con dichos vectores obtenemos el estimador de las varianzas y covarianzas:

$$\hat{\sigma}_{ij} = \frac{\hat{e}_i' \hat{e}_j}{T - k} = \frac{\sum_{t=1}^T \hat{e}_i \hat{e}_j}{T - k}$$

Formando una matriz estimada de las varianzas y covarianzas ($\hat{\Sigma}$) a partir de los estimadores anteriores, estamos en condiciones de obtener el estimador SUR, dado por:

$$\hat{\beta}_{SUR} = (X'(\hat{\Sigma} \otimes I)X)^{-1} X'(\hat{\Sigma} \otimes I)y$$

Suele utilizarse también otro estimador SUR a través de un proceso iterativo con los estimadores de las varianzas y covarianzas y los estimadores de los coeficientes del supermodelo. Una vez obtenido los estimadores SUR de los coeficientes, se los utiliza para estimar nuevamente la matriz de varianzas y covarianzas. A partir de la nueva estimación de las varianzas y covarianzas, aplicamos nuevamente la definición anterior y obtenemos en una segunda etapa un nuevo estimador SUR.

Este estimador SUR de segunda etapa puede ser utilizado para estimar nuevamente la matriz de varianzas y covarianzas. A partir de dicha nueva matriz, podemos obtener un estimador SUR en una tercera etapa; y así de forma consecutiva hasta que se produzca una convergencia a un valor estable. Dicha convergencia llevara a que dos iteraciones sucesivas sean similares.

No se repite el output proporcionado por los programas ya que resulta exactamente el mismo reproducido en la sección anterior. No hay ganancia al utilizar el proceso iterativo ya que los estimadores MCO que permiten obtener el estimador de la matriz de varianzas y covarianzas son los mismos que luego se obtienen al estimar el supermodelo. Por lo tanto, los estimadores de los parámetros de las demandas que se utilizarían para obtener en una segunda etapa el nuevo estimador de la matriz de varianzas y covarianzas son los mismos que se utilizaron en el primer paso. Es decir, las nuevas estimaciones por SUR en la segunda etapa serán los mismos que los de la primera etapa.

SUR RESTRINGIDO

Para estimar el modelo planteado teniendo en cuenta restricciones sobre las funciones de demanda, se debe estudiar con detenimiento cómo se construyen los estimadores del SUR Restringido.

La ecuación (1) desarrollaba el estimador por MCG cuando la matriz de Varianzas y Covarianzas (Φ), se definía: $\Phi = \Sigma \otimes I_T$. Cuando se tiene inserto las restricciones en términos del sistema de ecuaciones $R\beta = r$, se obtiene un tercer vector de coeficientes alternativos al de

MCO y MCG libre. Cabe aclarar que este vector se forma con el supuesto de que la H_0) $R\beta = r$ es cierta.

La optimización ahora surge de minimizar la suma de los cuadrados de los errores sujeto a las siete limitaciones:

$$\text{Min } (Y - X\beta)' (\Sigma^{-1} \otimes I)(Y - X\beta)$$

$$\text{sujeto a: } R\beta = r$$

Resolviendo la CPO se obtiene:

$$\hat{\beta}^* = \hat{\beta}_{mcg} + CR'(RCR')^{-1}(r - R\hat{\beta}_{mcg}), \text{ donde } C = (X' \Phi^{-1} X)^{-1}$$

El problema que se nos plantea es el de no poder contar con la verdadera matriz de varianzas y covarianzas y por ello debemos utilizar MCGE (estimados) construido con la matriz

$$\hat{\Sigma} = [\hat{\sigma}_{ij}]_{K \times K} \text{ de donde cada elemento } \hat{\sigma}_{ij} = \frac{1}{(T - K/M)} \sum_{t=1}^T \hat{e}_{it} \hat{e}_{jt} .$$

De esta manera, se puede hallar también a $\hat{\Phi} = \hat{\Sigma} \otimes I_M$ con lo que el vector de coeficientes estimados para H_0 cierta queda como sigue:

$$\hat{\beta}^* = \hat{\beta}_{mcg} + \hat{C}R'(\hat{R}\hat{C}R')^{-1}(r - R\hat{\beta}_{mcg}).$$

TEST DE CORRELACION CONTEMPORANEA

En la primera parte del trabajo, al describir las características del modelo SUR, enfatizamos la existencia de correlación no nula entre los errores de cada uno de los grupos o ecuaciones del sistema. La misma se define como contemporánea justamente porque relaciona perturbaciones en un mismo período de tiempo (t). Saber si existe correlación contemporánea o no es importante ya que si no existe, aplicar MCO a cada ecuación por separado es tan eficiente como SUR. Testear si existe correlación contemporánea es útil para determinar si es o no es necesario aplicar SUR.

Para el ejercicio de demandas,

$$\ln q_{it} = \beta_{i0} + \beta_{i1} \ln p_{1t} + \beta_{i2} \ln p_{2t} + \beta_{i3} \ln p_{3t} + \beta_{i4} \ln y_t + e_{it}$$

$$i = 1, 2, 3$$

se quiere establecer si influye en el consumo del bien “i” el comportamiento del bien “j”. De esta manera, del sistema:

Se puede deducir que las correlaciones se deben buscar contrastando los e_{it} con los e_{jt} para un mismo “t”. Sabiendo que: $\sigma_{ij} = \frac{1}{(T - K / M)} \sum_{t=1}^T e_{it} e_{jt}$; lo que se quiere probar es simplemente si los σ_{ij} son nulos o no.

$$H_0) \sigma_{ij} = 0 \quad \forall i \neq j$$

H_A) Al menos una covarianza igual a cero.

Construimos el estadístico del Multiplicador de Lagrange. El mismo se arma en base a los coeficientes de correlación muestral (r):

$$\lambda = T \sum_{i=2}^M \sum_{j=1}^{i-1} r_{ij}^2$$

con M = número de ecuaciones. El estadístico, en el límite, se distribuye asintóticamente Chi cuadrado con M(M-1)/2 grados de libertad (la cantidad de covarianzas distintas existentes en la matriz Σ).

TEST DE LA VALIDEZ DE LAS RESTRICCIONES

Para verificar la Hipótesis nula $R\beta = r$, debemos encontrar un estadístico capaz de seguir una distribución asintótica aproximada que sea conocida. Para esto, partiendo de la noción de que: $\hat{\beta} \sim N(\beta; C)$; por lo tanto, por propiedad de las matrices: $R\hat{\beta} \sim N(r; RCR')$ bajo la hipótesis nula **cierta**. Sabiendo las características del exponente de una Normal Multivariada, podemos descifrar que:

$$g = (R\hat{\beta} - r)' (RCR')^{-1} (R\hat{\beta} - r) \sim \chi^2(J).$$

Donde “J” es el número de restricciones. El último paso antes de utilizarlo para el contraste de hipótesis tiene que ver con la imposibilidad de contar con la verdadera varianza y, por ello, se construirá un \hat{g} que surge de cambiar C por \hat{C} y $\hat{\beta}$ por $\hat{\hat{\beta}}$. De esta manera,

$$\hat{g} \xrightarrow{d} \chi^2(J)$$

$$H_0) R\beta = r$$

$$H_A) R\beta \neq r$$

$$\alpha = 5\%$$

Existe otro test construido sobre la base de un estadístico con distribución que tiende a una F. Se parte de la noción de las sumas de cuadrados libres y restringidos:

$$\hat{\lambda}_F = \frac{\hat{g} / J}{\left(y - X\hat{\hat{\beta}} \right)' \left(\hat{\Sigma}^{-1} \otimes I \right) \left(y - X\hat{\hat{\beta}} \right) / (MT - K)} \xrightarrow{d} F(J; MT - K)$$

Puede demostrarse que el denominador converge en probabilidad a uno, por lo que puede ser omitido:

$$\hat{\lambda}_F = \frac{\hat{g}}{J} \sim F_{(J, MT-K)}$$

O de forma equivalente:

$$\hat{\lambda}_F = \frac{(e'_r e_r - e'e)/J}{e'e/(T - K/M)}$$

Y el valor observado de este estadístico debe ser comparado con el valor crítico teórico de una distribución F con $J, (MT-K)$ grados de libertad para un nivel de confianza α . Si el valor observado es mayor que el crítico, se rechaza la hipótesis nula planteada; en caso contrario, no se puede rechazar la hipótesis nula. ¿Cual es el significado en cada caso? Si la reducción en la suma de cuadrados de los residuos debido a la estimación libre (sin restricciones) no es suficientemente grande, no habrá pruebas para rechazar la hipótesis nula de que las restricciones son válidas. Por el contrario, si esta reducción es suficientemente grande, la hipótesis nula de que las restricciones son válidas se rechazará ya que al no considerarlas la suma de cuadrados de los residuos disminuye considerablemente.

Ambos estadísticos son asintóticamente equivalentes y en muestras finitas el estadístico F llevará al rechazo de la hipótesis nula en un numero menor de casos que el estadístico X^2 , por lo que responde a una aproximación más conservadora.

Por último, cabe considerar el efecto de utilizar el divisor $T-(K/M)$ en el estimador de Σ . Los estadísticos F y X^2 serán menores que si utilizáramos el divisor T, y por lo tanto la hipótesis nula será rechazada en un número menor de veces.

BIBLIOGRAFÍA:

- **CHIANG, Alpha:** *Economía Matemática*. Ed. McGraw-Hill. 1998
- **GREENE,** *Econometrics Analisis*. Prentice Hall. 1995
- **GUJARATI, Damodar :** *Econometría Básica*. Ed. McGraw-Hill. 1991
- **JOHNSTON, J. :** *Métodos Econométricos*. Ed. Vicens Vives. 1987. Tercera Edición.
- **JUDGE, GRIFFITHS, HILL, LUTKEPOHL:** *The Theory and Practice Of Econometrics*. Second Edition. Wiley Series. 1985
- **MADDALA, G. S. :** *Econometría*. Ed. McGraw-Hill. 1985.
- **WOOLDRIDGE, J.** *Econometrics Analisis Of Cross Sections and Pnale Data*. 2002. The MIT Press