Sarriko-On

Ecuaciones Simultáneas con aplicaciones en Gretls

ISBN: 978-84-691-9177-4

Marta Regúlez Castillo

05-08

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

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Ecuaciones Simultáneas con aplicaciones en Gretl

Marta Regúlez Castillo
Departamento de Economía Aplicada III (Econometría y Estadística)
Facultad de Ciencias Económicas y Empresariales
Universidad del País Vasco (UPV-EHU)

Contenido

1. El problema de la identificación **1**

1.1. Notación general del modelo 1

1.2. Estructuras observacionalmente equivalentes 2

1.3. Condiciones de Orden y Rango para la identificación 4

2. Estimación MCO en Modelos de Ecuaciones Simultáneas **9**

2.1. Estimación MCO de la Forma Estructural 9

2.2. Estimación MCO de la Forma Reducida 12

2.3. Forma Reducida: Estimación MCG equivalente a MCO 13

3. Métodos de información limitada **15**

3.1. Método de Mínimos Cuadrados Indirectos 15

3.2. Mínimos Cuadrados en dos Etapas 19

3.2.1. MC2E como estimador de Variables Instrumentales 20

3.2.2. Equivalencia entre MC2E y MCI en el caso de exacta identificación 21

3.3. Propiedades de los estimadores MCI y MC2E. Inferencia. 21

3.4. Máxima Verosimilitud con Información Limitada bajo Normalidad 26

4. Métodos de Información Completa **31**

4.1. Mínimos Cuadrados en tres Etapas 31

4.1.1. Propiedades del estimador MC3E 36

4.1.2. Inferencia 37

4.2. Máxima Verosimilitud Información Completa 37

5. Contrastes de Especificación **41**

5.1. Introducción 41
5.2. Contrastes de restricciones de sobreidentificación 43
5.3. Contrastes de exogeneidad 45

6. Ejemplos Empíricos utilizando Gretl **49**
6.1. Ejemplo 1: Inflación y grado de apertura. 49
6.1.1. Forma estructural y análisis de la identificación 50
6.1.2. Estimación por Mínimos cuadrados ordinarios 50
6.1.3. Estimación por Mínimos Cuadrados en dos etapas 51
6.1.4. Análisis de la bondad de los instrumentos 54
6.1.5. Extensiones del modelo original 55
6.2. Ejemplo 2: Oferta y Demanda de trabajo. 57
6.2.1. Identificación 59
6.2.2. Estimación de la Forma Estructural 60
6.2.3. Estimación de la Forma Reducida 65
6.2.4. Cambio de normalización de la segunda ecuación estructural. 67
6.2.5. Contraste de restricciones de sobreidentificación uniecuacional: Estadístico de Sargan 73
6.2.6. Contraste de Hausman como un contraste basado en residuos. (Residual based test) 75
6.3. Ejemplo 3: Modelo de Klein. 77
6.3.1. Estimación con Información Limitada 79
6.3.2. Estimación con Información Completa 84
6.3.3. Comparación de los resultados 86

Capítulo 1

El problema de la identificación

1.1. Notación general del modelo

Consideremos un modelo lineal que contenga G relaciones estructurales. La Forma Estructural nos recoge las relaciones entre todas las variables sean endógenas o exógenas. Las G variables endógenas se determinan conjuntamente dentro del sistema dadas las variables exógenas.

Forma Estructural (F.E.):

$$B\mathbf{y}_t + \Gamma\mathbf{x}_t = \mathbf{u}_t \qquad t = 1, \ldots, T$$

donde B es una matriz $(G \times G)$ de parámetros que acompañan a las variables endógenas, Γ es una matriz $(G \times K)$ que acompañan a las variables exógenas, $\mathbf{y}_t$ es el vector $(G \times 1)$ de variables endógenas en el momento t, $\mathbf{x}_t$ es el vector $(K \times 1)$ de variables exógenas en el momento t y $\mathbf{u}_t$ es el vector $(G \times 1)$ de términos de error de cada ecuación en el momento t de la F.E.

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1G} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2G} \\ \vdots & \vdots & \cdots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GG} \end{bmatrix} \qquad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{G1} & \gamma_{G2} & \cdots & \gamma_{GK} \end{bmatrix}$$

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Gt} \end{bmatrix} \quad \mathbf{x}_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Kt} \end{bmatrix} \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Gt} \end{bmatrix}$$

Podemos escribir la ecuación i-ésima de la F.E. en el momento t como:

$$\beta_{i1}y_{1t} + \beta_{i2}y_{2t} + \cdots + \beta_{iG}y_{Gt} + \gamma_{i1}x_{1t} + \gamma_{i2}x_{2t} + \cdots + \gamma_{iK}x_{Kt} + u_{it} \quad i = 1, \ldots, G \quad t = 1, \ldots, T$$

Suponemos que:

1) $|B| \neq 0$, el sistema es completo, no hay ecuaciones estructurales que son combinación unas de otras.

2) $\mathbf{u}_t \sim NID(0, \Sigma)$ donde

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1G} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2G} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{G1} & \sigma_{G2} & \cdots & \sigma_G^2 \end{bmatrix}$$

3) Las variables exógenas $\mathbf{x}_t$ son independientes de los términos de perturbación u_{it} [1]. La variable x_{1t} normalmente será el término constante $x_{1t} = 1 \quad \forall t$.

Forma Reducida (F.R.)

$$\mathbf{y}_t = \Pi \mathbf{x}_t + \mathbf{v}_t \qquad t = 1, \ldots, T$$

donde

$$\Pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1K} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ \pi_{G1} & \pi_{G2} & \cdots & \pi_{GK} \end{bmatrix} = -B^{-1}\Gamma \qquad \mathbf{v}_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Gt} \end{bmatrix} = B^{-1}\mathbf{u}_t$$

siendo $\mathbf{v}_t \sim NID(0, \Omega)$ donde $\Omega = B^{-1}\Sigma(B^{-1})'$.

Podemos escribir la ecuación i-ésima de la F.R. en el momento t como:

$$y_{it} = \pi_{i1}x_{1t} + \pi_{i2}x_{2t} + \pi_{i3}x_{3t} \cdots + \pi_{iK}x_{Kt} + u_{it} \quad i = 1, \ldots, G \quad t = 1, \ldots, T$$

La matriz Π contiene GK elementos. Las matrices B y Γ contienen como máximo $G^2 + GK$ elementos. Por lo tanto si no hay restricciones sobre los parámetros en B y Γ existe una infinidad de estructuras o valores de los parámetros estructurales en B y Γ que corresponden a los mismos valores de los parámetros en la matriz Π de la forma reducida.

1.2. Estructuras observacionalmente equivalentes

El conocimiento de B, Γ y Σ solo puede venir de Π y de Ω ya que de los datos $(\mathbf{y}_t, \mathbf{x}_t)$ podemos tener conocimiento de Π y de Ω, ya que lo que observamos es la solución de $\mathbf{y}_t$ dado $\mathbf{x}_t$ y $\mathbf{v}_t$, es decir la forma reducida. El problema de la identificación surge cuando no podemos recuperar de forma única los parámetros de la forma estructural a partir de la forma reducida.

Si dos estructuras o valores concretos de los parámetros de la Forma Estructural, por ejemplo $(B_0, \Gamma_0, \Sigma_0)$ y $(B_1, \Gamma_1, \Sigma_1)$, dan lugar a la misma Forma Reducida o valores de Π y Ω, entonces se

[1] Por el momento no consideramos la existencia de variables endógenas retardadas que no serían propiamente exógenas sino predeterminadas. Estas no serían independientes de u_{it} pero sí pueden estar incorrelacionadas con el término de error si estos son independientes en el tiempo

dice que son estructuras observacionalmente equivalentes. En este caso, no podemos distinguir cual de las dos estructuras ha generado los datos observados. Se dice que no están identificadas.

Supongamos que la verdadera estructura que ha generado los datos observados es $(B_0, \Gamma_0, \Sigma_0)$ que tiene asociada la forma reducida (Π_0, Ω_0). Si esta estructura no está identificada es que existen otras estructuras $(FB_0, F\Gamma_0, F\Sigma_0 F')$ donde F es cualquier matriz no singular $|F| \neq 0$, $F \neq I$ tal que:

$$\Pi^* = -(FB_0)^{-1}(F\Gamma_0) = -B_0^{-1}F^{-1}F\Gamma_0 = -B_0^{-1}\Gamma_0 = \Pi_0$$

$$\Omega^* = (FB_0)^{-1}(F\Sigma_0F')((FB_0)^{-1})' = B_0^{-1}F^{-1}F\Sigma_0F'(F')^{-1}(B_0^{-1})' = B_0^{-1}\Sigma_0B_0^{-1} = \Omega_0$$

es decir, $(FB_0, F\Gamma_0, F\Sigma_0F')$ son observacionalmente equivalentes a $(B_0, \Gamma_0, \Sigma_0)$ ya que dan lugar a la misma Forma Reducida.

En términos más generales, sin suponer normalidad de $\mathbf{u}$, sino cualquier función de densidad f($\mathbf{u}$), tenemos el siguiente teorema:

Teorema: Dos estructuras $S_0 = (B_0, \Gamma_0, f_0(\mathbf{u}))$ y $S_1 = (B_1, \Gamma_1, f_1(\mathbf{u}))$ son observacionalmente equivalentes si y solo si $B_0^{-1}\Gamma_0 = B_1^{-1}\Gamma_1$ y $f(B_0^{-1}\mathbf{u}) = f(B_1^{-1}\mathbf{u})$ o equivalentemente si y solo si existe una transformación lineal no singular F tal que $B_1 = FB_0$, $\Gamma_1 = F\Gamma_0$ y $f(\mathbf{u}) = f(\mathbf{Fu})$.

Se llaman **Estructuras admisibles del modelo** a aquellas que satisfacen todas las restricciones que caracterizan el modelo. Se dice que F es una transformación lineal no singular admisible del modelo si la estructura $(FB, F\Gamma, F\Sigma F')$ es admisible del modelo.

¿Cómo identificar un modelo? ¿Bajo qué condiciones no existirán estructuras observacionalmente equivalentes dada la especificación del modelo?

Podemos considerar dos formas de comprobar la identificación del modelo:

1) Ver si se pueden recuperar los parámetros estructurales de los parámetros de la forma reducida.

2) Demostrar que no existe ninguna otra estructura lineal que es admisible del modelo, es decir, que la única estructura que satisface todas las restricciones del modelo es $(FB, F\Gamma, F\Sigma F')$ donde $F = I$.

Si no imponemos ninguna restricción tenemos $(G^2 + GK)$ parámetros estructurales desconocidos en B y Γ y $(\frac{G(G+1)}{2})$ en Σ, mientras que los parámetros que tenemos en la forma reducida son GK en Π y $(\frac{G(G+1)}{2})$ en Ω. Luego la diferencia nos da un exceso de

$$G^2 + GK + \frac{G(G+1)}{2} - GK - \frac{G(G+1)}{2} = G^2$$

Información adicional que puede ayudar a identificar:

1) Normalizaciones: el coeficiente de una variable endógena en cada ecuación de la forma estructural sea la unidad.

2) Identidades: sus coeficientes son conocidos y aportan posibles restricciones de exclusión en otras ecuaciones.

3) Exclusión de variables, o equivalentemente restricciones de coeficientes cero, en distintas ecuaciones de la F. E.

4) Restricciones lineales en F.E.

5) Restricciones en Σ

6) No linealidades en variables (o en parámetros aunque en el marco que nos vamos a mover no vamos a tener este caso).

1.3. Condiciones de Orden y Rango para la identificación

Teníamos un exceso de G^2 incógnitas estructurales si no imponíamos ningún tipo de restricción ni normalización. Vamos a imponer que en cada ecuación una de las variables endógenas tenga coeficiente igual a uno, es decir normalizamos con respecto a ese coeficiente y esa sería la variable dependiente de esa ecuación de la forma estructural. Entonces tenemos G incógnitas menos, luego el exceso es ahora de $(G^2 - G) = G(G-1)$.

Para formalizar los criterios de identificación, utilizaremos la siguiente notación para escribir una ecuación cualquiera de la Forma estructural del sistema, llamémosla j. Por conveniencia omitimos el subíndice t:

$$\underbrace{B'_j}_{(1\times G)} \underbrace{\mathbf{y}}_{(G\times 1)} + \underbrace{\Gamma'_j}_{(1\times K)} \underbrace{\mathbf{x}}_{(K\times 1)} = \mathbf{u}_j$$

donde B'_j y Γ'_j son las filas j-ésima de B y Γ respectivamente. En esta ecuación sabemos que:

1) Uno de los elementos de B'_j es uno: Normalización

2) Algunas variables que aparecen en otras ecuaciones del modelo pueden no estar incluidas en esta ecuación: Exclusión de variables endógenas y/o exógenas.

Para incorporar estas restricciones o información vamos a utilizar la siguiente notación para la ecuación j donde además de la variable dependiente y_j tenemos:

	Variables endógenas	**Variables exógenas**
Incluidas	Y_j con G_j variables	X_j con K_j variables
Excluidas	Y_j^* con G_j^* variables	X_j^* con K_j^* variables

Número de ecuaciones o variables endógenas en el modelo: $G_j + G_j^* + 1 = G$
Número de variables exógenas total en el modelo : $K_j + K_j^* = K$
El coeficiente asociado a y_j es la unidad. El símbolo $*$ está asociado con variables excluidas.

Dada esta notación, podemos escribir la ecuación j-ésima como:

$$\left[\begin{array}{ccc} 1 & -\beta_j' & -\beta_j^{*'} \end{array}\right] \left[\begin{array}{c} y_j \\ Y_j \\ Y_j^* \end{array}\right] + \left[\begin{array}{cc} -\gamma_j' & -\gamma_j^{*'} \end{array}\right] \left[\begin{array}{c} X_j \\ X_j^* \end{array}\right] = \mathbf{u}_j$$

esto es,

$$y_j = \beta_j' Y_j + \beta_j^{*'} Y_j^* + \gamma_j' X_j + \gamma_j^{*'} X_j^* + \mathbf{u}_j$$

Las exclusiones implican que $\beta_j^* = 0$ y $\gamma_j^* = 0$. Por lo tanto

$$B_j' = \left[\begin{array}{ccc} 1 & -\beta_j' & \mathbf{0}' \end{array}\right] \text{ y } \Gamma_j' = \left[\begin{array}{cc} -\gamma_j' & \mathbf{0}' \end{array}\right]$$

La matriz de coeficientes de la forma reducida es:

$$\Pi = -B^{-1}\Gamma \quad \Rightarrow B\Pi = -\Gamma$$

Para la ecuación j-ésima de este sistema: $B_j'\Pi = -\Gamma_j'$ esto es

$$\left[\begin{array}{ccc} 1 & \underbrace{-\beta_j'}_{(1\times G_j)} & \underbrace{\mathbf{0}'}_{1\times G_j^*)} \end{array}\right] \left[\begin{array}{cc} \underbrace{\pi_j'}_{(1\times K_j)} & \underbrace{\pi_j*'}_{(1\times K_j^*)} \\ \underbrace{\Pi_{1j}}_{(G_j\times K_j)} & \underbrace{\Pi_{1j}^*}_{(G_j\times K_j^*)} \\ \underbrace{\Pi_{2j}}_{(G_j^*\times K_j)} & \underbrace{\Pi_{2j}^*}_{(G_j^*\times K_j^*)} \end{array}\right] = \left[\begin{array}{cc} \underbrace{\gamma_j'}_{(1\times K_j)} & \underbrace{\mathbf{0}'}_{(1\times K_j^*)} \end{array}\right]$$

De este sistema extraemos las siguientes K ecuaciones:

$$\pi_j - \Pi_{1j}'\beta_j = \gamma_j \qquad K_j \quad \texttt{ecuaciones} \tag{1.1}$$

$$\pi_j^* - \Pi_{1j}^{*'}\beta_j = \mathbf{0} \qquad K_j^* \quad \texttt{ecuaciones} \tag{1.2}$$

De (1.2) $\Longrightarrow \Pi_{1j}^{*'}\beta_j = \pi_j^*$ por lo tanto, conocidas Π_{1j}^* y π_j^* tenemos un sistema de ecuaciones lineal no homogéneo con K_j^* ecuaciones en G_j incógnitas. Si podemos resolver para β_j, podemos utilizar esta solución en (1.1) para obtener γ_j y por lo tanto, los parámetros estructurales de la ecuación j estarán identificados. Diremos entonces que la ecuación j estará identificada.

Una **condición necesaria pero no suficiente** para que (1.2) tenga solución es que *deberá de haber al menos tantas ecuaciones como incógnitas*, lo cual nos lleva a la siguiente condición necesaria para la identificación de la j-ésima ecuación:

Condición de orden: $K_j^* \geq G_j$, esto es, el número de variables **exógenas excluidas** de la ecuación j-ésima deberá de ser **al menos igual o mayor** que el número de variables **endógenas incluidas** en la ecuación j-ésima.

La condición de orden es necesaria pero no es suficiente para la identificación de la ecuación j-ésima, ya que establece la condición para que (1.2) tenga al menos una solución pero no asegura que sea única. La **condición necesaria y suficiente** para que exista y sea única esta solución es (1.2) la siguiente:

Condición de Rango: $rango(\Pi_{1j}^{*'}) = G_j$. Si se satisface esta condición por la ecuación j-ésima, asegura una única solución para los parámetros estructurales de esa ecuación dados los parámetros de la forma reducida.

Una aproximación alternativa al problema de la identificación es la de utilizar restricciones a priori sobre B y Γ para eliminar otras estructuras que sean observacionalmente equivalentes (que ya no fueran admisibles al modelo). Veamos en este sentido las condiciones de identificación.

Escribamos la matriz A de parámetros estructurales $(G \times (G+K))$ que representa una estructura como

$$A = \left[\begin{array}{cc} B & \Gamma \end{array} \right]$$

y dada una matriz $(G \times G)$ cualquiera F no singular tal que

$$FA = \left[\begin{array}{cc} FB & F\Gamma \end{array} \right]$$

representa una nueva estructura lineal. Esta nueva estructura es admisible del modelo, o equivalentemente, F es una matriz de transformación admisible, si FA satisface todas las restricciones a priori en relación con A. Así, la identificabilidad de la ecuación j-ésima implica que la j-ésima ecuación de cada una de las estructuras admisibles sea un múltiplo escalar de la verdadera estructura.

Denotemos por:

$\mathbf{i}'_j$: vector fila que tiene un 1 en la j-ésima posición y ceros el resto de elementos.

$\mathbf{a}'_j = \mathbf{i}'_j A = \left[\begin{array}{cc} B'_j & \Gamma'_j \end{array} \right]$

Las restricciones de exclusión, $\beta_j^* = \mathbf{0}$ y $\gamma_j^* = \mathbf{0}$, las podemos expresar como

$$\underbrace{\mathbf{a}'_j}_{1\times(G+K)} \underbrace{\Phi_j}_{(G+K)\times R} = \underbrace{\mathbf{0'}}_{1\times R}$$

donde $R = G_j^* + K_j^*$ y Φ_j es una matriz de constantes con unos y ceros que selecciona de $\mathbf{a}'_j$ solamente β_j^* y γ_j^*.

De la misma forma podemos considerar a FA y denotar $\tilde{\mathbf{a}}'_j = \mathbf{i}'_j FA = \mathbf{f}'_j A$, donde $\mathbf{f}'_j$ es la fila j-ésima de F. Luego para que F sea una transformación admisible del modelo, $\tilde{\mathbf{a}}'_j$ tiene que satisfacer las mismas restricciones que $\mathbf{a}'_j$, por lo tanto

$$\underbrace{\tilde{\mathbf{a}}'_j}_{1\times(G+K)} \underbrace{\Phi_j}_{(G+K)\times R} = \underbrace{\mathbf{0'}}_{1\times R} \quad \Leftrightarrow \quad \mathbf{f}'_j(A\Phi_j) = \mathbf{0'} \quad \Leftrightarrow \quad \underbrace{(A\Phi_j)'}_{(R\times G)} \underbrace{\mathbf{f}_j}_{G\times 1)} = \underbrace{\mathbf{0}}_{(R\times 1)} \tag{1.3}$$

Por otro lado

$$\mathbf{a}'_j \Phi_j = \mathbf{0'} \quad \Leftrightarrow \quad \mathbf{i}'_j(A\Phi_j) = \mathbf{0'} \quad \Leftrightarrow \quad \underbrace{(A\Phi_j)'}_{(R\times G)} \underbrace{\mathbf{i}_j}_{(G\times 1)} = \underbrace{\mathbf{0}}_{(R\times 1)} \tag{1.4}$$

Tanto (1.3) como (5.4) son un sistema de ecuaciones lineal homogéneo. La identificación de la ecuación estructural j-ésima requiere que $\mathbf{f}_j$ sea un múltiplo escalar de $\mathbf{i}_j$, esto es, un vector con todos sus elementos igual a cero excepto el elemento j-ésimo que es igual a una constante c.

Si rango$(A\Phi_j)' =$ rango$(A\Phi_j) = G - 1$ entonces $\mathbf{i}_j$ es solución y también $c\mathbf{i}_j$ para cualquier constante c. Por lo tanto, tenemos la siguiente **condición necesaria y suficiente** para la identificación de la j-ésima ecuación estructural, en términos de los parámetros estructurales[2]:

Condición de Rango: **rango**$(A\Phi_j) = G - 1$

Entonces, se dice que la ecuación estructural j-ésima:

- **No está identificada**, si bien no se satisface la condición de orden, esto es se tiene que

$$K_j^* < G_j \Leftrightarrow G_j^* + K_j^* < G - 1$$

o bien se satisface la condición de orden pero no se satisface la de rango, esto es

$$K_j^* \geq G_j \quad \texttt{pero} \quad rango(A\Phi_j) < G - 1$$

- **Exactamente identificada**, si se satisface la condición de rango y la condición de orden con igualdad, esto es

$$K_j^* = G_j \Leftrightarrow G_j^* + K_j^* = G - 1 \quad \texttt{y} \quad rango(A\Phi_j) = G - 1$$

- **Sobreidentificada**, si se satisface la condición de rango y la condición de orden con exceso, esto es

$$K_j^* > G_j \Leftrightarrow G_j^* + K_j^* > G - 1 \quad \texttt{y} \quad rango(A\Phi_j) = G - 1$$

[2]Notar que la condición necesaria de identificación de esa ecuación es que $K_j^* \geq G_j$ lo que implica que

$$G_j^* + K_j^* \geq G_j^* + G_j \Leftrightarrow R \geq G - 1$$

Capítulo 2

Estimación MCO en Modelos de Ecuaciones Simultáneas

2.1. Estimación MCO de la Forma Estructural

Forma Estructural (F.E.):

$$\underbrace{B}_{GxG}\underbrace{\mathbf{y}_t}_{Gx1} + \underbrace{\Gamma}_{GxK}\underbrace{\mathbf{x}_t}_{Kx1} = \underbrace{\mathbf{u}_t}_{Gx1} \qquad t = 1, \ldots, T \qquad \mathbf{u}_t \sim NID(0, \Sigma)$$

Si denotamos

$$\underbrace{Y}_{TxG} = \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_T \end{bmatrix} \quad \underbrace{X}_{TxK} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_T \end{bmatrix} \quad \underbrace{U}_{TxG} = \begin{bmatrix} \mathbf{u}'_1 \\ \mathbf{u}'_2 \\ \vdots \\ \mathbf{u}'_T \end{bmatrix}$$

podemos escribir la Forma Estructural como

$$\underbrace{Y}_{TxG}\underbrace{B'}_{GxG} + \underbrace{X}_{TxK}\underbrace{\Gamma'}_{KxG} = \underbrace{U}_{TxG}$$

Consideremos la j-ésima ecuación estructural del Modelo de Ecuaciones Simultáneas, después de haberle impuesto todas las restricciones a priori y el convenio de normalización[1]

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

[1]Por lo tanto $G = G_j + G_j^* + 1$ y $K = K_j + K_j^*$.

donde $Z_j = [Y_j \;\; X_j]$ $\quad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$. El estimador MCO de δ_j se define como:

$$\hat{\delta}_j^{MCO} = (Z_j' Z_j)^{-1} Z_j' \mathbf{y}_j = \delta_j + (Z_j' Z_j)^{-1} Z_j' \mathbf{u}_j$$

Teorema: Dado que $plim\, \left(\frac{1}{T} X_j' U\right) = 0$ y $plim\, \left(\frac{1}{T} X' X\right) = M$ matriz (KxK) finita y definida positiva, si $plim\, \left(\frac{1}{T} Y_j' \mathbf{u}_j\right) \neq 0$ entonces $plim\, \hat{\delta}_j^{MCO} \neq \delta_j$ por lo que el estimador MCO de δ_j no es consistente[2]

Demostración:

$$plim\, \hat{\delta}_j^{MCO} = \delta_j + \left[plim \left(\frac{1}{T} Z_j' Z_j\right)\right]^{-1} plim \left(\frac{1}{T} Z_j' \mathbf{u}_j\right)$$

Por un lado tenemos

$$plim \left(\frac{1}{T} Z_j' Z_j\right) = \begin{bmatrix} plim \left(\frac{1}{T} Y_j' Y_j\right) & plim \left(\frac{1}{T} Y_j' X_j\right) \\ \\ plim \left(\frac{1}{T} X_j' Y_j\right) & plim \left(\frac{1}{T} X_j' X_j\right) \end{bmatrix}$$

Si particionamos las siguientes matrices:

$$Y = \left[\underbrace{Y_j}_{TxG_j} \vdots \underbrace{Y_*}_{Tx(G_j^*+1)}\right] \quad \Pi' = \left[\underbrace{\Pi_j'}_{KxG_j} \vdots \underbrace{\Pi_*'}_{Kx(G_j^*+1)}\right] \quad V = \left[\underbrace{V_j}_{TxG_j} \vdots \underbrace{V_*}_{Tx(G_j^*+1)}\right] \quad D = \left[\underbrace{D_j}_{GxG_j} \vdots \underbrace{D_*}_{Gx(G_j^*+1)}\right]$$

tenemos que en la Forma Reducida

$$Y_j = X\Pi_j' + V_j$$

donde, como $V = UD$ podemos expresar $V_j = \underbrace{U}_{TxG} \underbrace{D_j}_{GxG_j}$. Por lo tanto tenemos los siguientes resultados:

$$\begin{aligned} Y_j' Y_j &= \Pi_j X' X \Pi_j' + V_j' V_j + \Pi_j X_j' V_j + V_j' X \Pi_j' \\ Y_j' X_j &= \Pi_j X' X_j + V_j' X_j \end{aligned}$$

Tomando límites en probabilidad una vez multiplicamos ambos lados de las igualdades por $\frac{1}{T}$ tenemos

$$\begin{aligned} plim \left(\tfrac{1}{T} Y_j' Y_j\right) &= \Pi_j\, plim \left(\tfrac{1}{T} X' X\right) \Pi_j' + plim \left(\tfrac{1}{T} V_j' V_j\right) + \Pi_j\, plim \left(\tfrac{1}{T} X_j' V_j\right) + plim \left(\tfrac{1}{T} V_j' X\right) \Pi_j' \\ plim \left(\tfrac{1}{T} Y_j' X_j\right) &= \Pi_j\, plim \left(\tfrac{1}{T} X' X_j\right) + plim \left(\tfrac{1}{T} V_j' X_j\right) \end{aligned}$$

[2] En general, al incluir la matriz de regresores Z_j a variables endógenas Y_j que son regresores estocásticos no independientes de $\mathbf{u}_j$, el estimador $\hat{\delta}_j^{MCO}$ es no lineal y sesgado, siendo desconocida su distribución exacta para muestras finitas.

Dado que

$$\begin{aligned} plim\left(\frac{1}{T}X_j'V_j\right) &= plim\left(\frac{1}{T}X_j'U\right)D_j = 0 \\ plim\left(\frac{1}{T}V_j'V_j\right) &= \Omega_j \end{aligned}$$

donde Ω_j es la submatriz de varianzas y covarianzas de la forma reducida asociada a Y_j.

Entonces

$$\begin{aligned} plim\left(\frac{1}{T}Y_j'Y_j\right) &= \Pi_j\, M\, \Pi_j' + \Omega_j \\ plim\left(\frac{1}{T}Y_j'X_j\right) &= \Pi_j\, M_j \\ plim\left(\frac{1}{T}X_j'X_j\right) &= M_{jj} \end{aligned}$$

donde $\underbrace{M_j}_{KxK_j}$ y $\underbrace{M_{jj}}_{K_jxK_j}$ son las submatrices correspondientes de M. Luego obtenemos que

$$plim\left(\frac{1}{T}Z_j'Z_j\right) = \begin{bmatrix} \Pi_j\, M\, \Pi_j' + \Omega_j & \Pi_j\, M_j \\ M_j'\, \Pi_j' & M_{jj} \end{bmatrix}$$

Por otro lado

$$plim\left(\frac{1}{T}Z_j'\mathbf{u}_j\right) = \begin{bmatrix} plim\left(\frac{1}{T}Y_j'\mathbf{u}_j\right) \\ plim\left(\frac{1}{T}X_j'\mathbf{u}_j\right) \end{bmatrix}$$

Dado que $plim\left(\frac{1}{T}X_j'\,U\right) = 0$ entonces $plim\left(\frac{1}{T}X_j'\mathbf{u}_j\right) = 0$.

El otro límite en probabilidad lo podemos obtener como sigue:

$$Y_j'\mathbf{u}_j = \Pi_j X'\mathbf{u}_j + V_j'\mathbf{u}_j$$

Multiplicando ambos lados de la igualdad por $\frac{1}{T}$ y tomando límites en probabilidad tenemos:

$$plim\left(\frac{1}{T}Y_j'\mathbf{u}_j\right) = \Pi_j\, \underbrace{plim\left(\frac{1}{T}X'\mathbf{u}_j\right)}_{=0} + plim\left(\frac{1}{T}V_j'\mathbf{u}_j\right)$$

$$plim\left(\frac{1}{T}V_j'\mathbf{u}_j\right) = D_j'\, plim\left(\frac{1}{T}U'\mathbf{u}_j\right) = D_j'\,\sigma_{.j}$$

donde $\sigma_{.j}$ es la j-ésima columna de la matriz de varianzas y covarianzas de las perturbaciones estructurales, Σ. Esta columna recoge la varianza de u_{jt} y las covarianzas de u_{jt} con el resto de perturbaciones estructurales. Por otro lado D_j' es la submatriz (G_jxG) de la matriz $D' \equiv B^{-1}$, matriz de coeficientes estructurales que acompañan a las variables endógenas.

Por lo tanto,

$$plim\,\hat{\delta}_j^{MCO} = \delta_j + \left[plim\left(\frac{1}{T}Z_j'Z_j\right)\right]^{-1} plim\left(\frac{1}{T}Z_j'\,\mathbf{u}_j\right) =$$

$$= \delta_j + \left[\begin{array}{cc} \Pi_j\,M\,\Pi_j' + \Omega_j & \Pi_j\,M_j \\ & \\ M_j'\,\Pi_j' & M_{jj} \end{array}\right]^{-1} \left[\begin{array}{c} D_j'\,\sigma_{.j} \\ \\ \mathbf{0} \end{array}\right] \neq \delta_j \;\; \texttt{si} \;\; D_j'\,\sigma_{.j} \neq 0$$

2.2. Estimación MCO de la Forma Reducida

Forma Reducida (F.R.):

$$\mathbf{y}_t = \Pi \mathbf{x}_t + \mathbf{v}_t \qquad t = 1, \ldots, T \qquad \mathbf{v}_t \sim NID(0, \Omega)$$

donde $\underbrace{\Pi}_{GxK} = -B^{-1}\Gamma$ y $\mathbf{v}_t = B^{-1}\mathbf{u}_t$ y $\Omega = B^{-1}\Sigma(B^{-1})'$.

Utilizando la notación anterior también podemos expresar la Forma Reducida como:

$$\underbrace{Y}_{TxG} = \underbrace{X}_{TxK}\underbrace{\Pi'}_{KxG} + \underbrace{V}_{TxG}$$

donde $\underbrace{\Pi'}_{KxG} = -\Gamma'\,(B')^{-1}$, y $V = U\,(B')^{-1} \equiv UD$

Si escribimos todas las ecuaciones de la Forma Reducida de la forma:

$$\underbrace{Y}_{TxG} = \underbrace{X}_{TxK}\underbrace{\Pi'}_{KxG} + \underbrace{V}_{TxG}$$

El estimador MCO de Π' es

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

Equivalentemente podemos obtener cada columna de $\hat{\Pi}'$ de la regresión ecuación por ecuación. Esto es, considerando la ecuación j-ésima de la Forma Reducida:

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{X}_{TxK}\underbrace{\pi_j}_{Kx1} + \mathbf{v}_j$$

donde

$$\pi_j = \left[\begin{array}{c} \pi_{j1} \\ \pi_{j2} \\ \vdots \\ \pi_{jK} \end{array}\right]$$

El estimador $\hat{\pi}_j^{MCO} = (X'X)^{-1}X'\mathbf{y}_j$ es la columna j-ésima de $\hat{\Pi}'_{MCO} = \left[\hat{\pi}_1^{MCO} \;\cdots\; \hat{\pi}_G^{MCO}\right]$.

Teorema:

Dado que $plim\ \left(\frac{1}{T}X'\,U\right) = 0$ para y $plim\ \left(\frac{1}{T}X'\,X\right) = M$ matriz (KxK) finita y definida positiva, el estimador MCO de Π' es consistente.

Demostración:

$$plim\hat{\Pi}'_{MCO} = \Pi' + \left[\underbrace{plim\,\frac{1}{T}\,(X'\,X)}_{M}\right]^{-1}\underbrace{plim\,\left(\frac{1}{T}X'V\right)}_{=0} = \Pi'$$

dado que como $V = UD$ donde D es una matriz de constantes

$$plim\,\left(\frac{1}{T}X'\,U\right) = 0 \Rightarrow plim\,\left(\frac{1}{T}X'V\right) = 0$$

2.3. Forma Reducida: Estimación MCG equivalente a MCO.

La Forma Reducida no restringida es un sistema de ecuaciones aparentemente no relacionadas (SURE) donde en cada ecuación tenemos la misma matriz de datos de las variables exógenas.

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{bmatrix} = \begin{bmatrix} X & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & X & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & X \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_G \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_G \end{bmatrix} \quad \Leftrightarrow \underbrace{\mathbf{y}}_{GTx1} = \underbrace{(I_G \bigotimes X)}_{GTxGK}\underbrace{\Psi}_{GKx1} + \underbrace{\mathbf{v}}_{GTx1}$$

donde $\mathbf{v} \sim (\mathbf{0}, (\Omega \bigotimes I_T))$.

$$E(\mathbf{v}\mathbf{v}') = (\Omega \bigotimes I_T) = \begin{bmatrix} w_1^2 I_T & w_{12}I_T & \cdots & w_{1G}I_T \\ w_{12}I_T & w_2^2 I_T & \cdots & w_{2G}I_T \\ \vdots & \vdots & \ddots & \vdots \\ w_{G1}I_T & w_{G2}I_T & \cdots & w_G I_T \end{bmatrix}$$

Por esa razón, estimar el sistema por un método que tenga en cuenta la matriz de varianzas y covarianzas del vector de perturbaciones, como es MCG(F), no aporta nada en términos de eficiencia ya que es idéntico a estimar por MCO cada ecuación por separado.

$$\begin{aligned} \Psi_{MCG} &= \left((I_G \bigotimes X)'(\Omega \bigotimes I_T)^{-1}(I_G \bigotimes X)\right)^{-1}\left((I_G \bigotimes X)'(\Omega \bigotimes I_T)^{-1}\mathbf{y}\right) = \\ &= \left((\Omega^{-1} \bigotimes X'X)\right)^{-1}\left((\Omega^{-1} \bigotimes X')\mathbf{y})\right) = \left(I_G \bigotimes (X'X)^{-1}X'\right)\mathbf{y} = \Psi_{MCO} \end{aligned}$$

Capítulo 3

Métodos de información limitada

3.1. Método de Mínimos Cuadrados Indirectos

El método de Mínimos Cuadrados Indirectos es un método de estimación de Modelos de Ecuaciones Simultáneas con información limitada. Información limitada se refiere a que se estima una ecuación aislada de la Forma Estructural sin hacer uso de toda la información contenida en la especificación detallada del resto del modelo. La información se limita a la especificación de esa ecuación considerando tanto las variables endógenas y exógenas que están incluidas, y cuales son las excluidas de ella pero pertenecientes al modelo. No se utiliza la especificación concreta del resto de ecuaciones, ni tampoco de la matriz de varianzas y covarianzas del sistema.

Como veremos, esta técnica de estimación está indicada cuando la ecuación está exactamente identificada y se puede considerar como un método de estimación por Variables Instrumentales.

Este método consiste en:

- Estimar la matriz de coeficientes de la forma reducida Π por MCO ecuación por ecuación,

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{X}_{TxK}\underbrace{\pi_j}_{Kx1} + \mathbf{v}_j$$

$$\hat{\pi}_j^{MCO} = (X'X)^{-1}X'\mathbf{y}_j \qquad j = 1, \ldots, G$$

que es equivalente a estimar el sistema:

$$\underbrace{Y}_{TxG} = \underbrace{X}_{TxK}\underbrace{\Pi'}_{KxG} + \underbrace{V}_{TxG}$$

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

$\hat{\Pi}'_{MCO} = \left[\hat{\pi}_1^{MCO} \ \cdots \ \hat{\pi}_G^{MCO}\right]$, donde $\hat{\pi}_j^{MCO} \quad j = 1, \ldots G$ son las columnas de $\hat{\Pi}'_{MCO}$.

- Obtener las estimaciones de los coeficientes estructurales de la ecuación de interés, B_j y Γ_j, a partir del sistema de ecuaciones que relaciona Π con B_j y Γ_j.

Consideremos la j-ésima ecuación estructural del Modelo de Ecuaciones Simultáneas, después de haberle impuesto todas las restricciones a priori y el convenio de normalización[1]

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1}$$

o de forma más completa, equivalentemente:

$$\begin{bmatrix} \mathbf{y}_j & Y_j & Y_j^* & X_j & X_j^* \end{bmatrix} \begin{bmatrix} 1 \\ -\beta_j \\ \mathbf{0} \\ -\gamma_j \\ \mathbf{0} \end{bmatrix} = \mathbf{u}_j$$

donde Y_j^* y X_j^* son las matrices de observaciones de las G_j^* variables endógenas y las K_j^* variables exógenas o predeterminadas que se han excluido de la ecuación estructural j-ésima.

Las relaciones entre los coeficientes estructurales de esa ecuación y los coeficientes de la forma reducida vienen dados por

$$\Pi' B_j = -\Gamma_j \quad \Leftrightarrow \quad \Pi' \begin{bmatrix} 1 \\ -\beta_j \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \gamma_j \\ \mathbf{0} \end{bmatrix}$$

Utilizando la estimación de Π' en este sistema y denotando los estimadores MCI de los parámetros estructurales por $\hat{\beta}_j^{MCI}$ y $\hat{\gamma}_j^{MCI}$ como los vectores que se obtienen de resolver el sistema de K ecuaciones:

$$(X'X)^{-1}X'Y \begin{bmatrix} 1 \\ -\hat{\beta}_j^{MCI} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_j^{MCI} \\ \mathbf{0} \end{bmatrix}$$

La cuestión fundamental es si existen unos vectores solución $\hat{\beta}_j^{MCI}$, $\hat{\gamma}_j^{MCI}$ que sean únicos.

Reescribiendo el sistema como

$$(X'X)^{-1}X' \begin{bmatrix} \mathbf{y}_j & Y_j & Y_j^* \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\beta}_j^{MCI} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_j^{MCI} \\ \mathbf{0} \end{bmatrix}$$

$$(X'X)^{-1}X'\mathbf{y}_j - (X'X)^{-1}X'Y_j\hat{\beta}_j^{MCI} = \begin{bmatrix} \hat{\gamma}_j^{MCI} \\ \mathbf{0} \end{bmatrix}$$

Premultiplicando por $(X'X)$

$$X'\mathbf{y}_j - X'Y_j\hat{\beta}_j^{MCI} = (X'X) \begin{bmatrix} \hat{\gamma}_j^{MCI} \\ \mathbf{0} \end{bmatrix}$$

[1]Por lo tanto $G = G_j + G_j^* + 1$ y $K = K_j + K_j^*$.

Si particionamos $\underbrace{X}_{TxK} = \left(\underbrace{X_j}_{TxK_j} \vdots \underbrace{X_j^*}_{TxK_j^*} \right)$ entonces $\underbrace{X'}_{KxT} = \begin{pmatrix} \underbrace{X_j'}_{K_jxT} \\ \underbrace{X_j^{*\prime}}_{K_j^*xT} \end{pmatrix}$ y tenemos

$$\begin{pmatrix} \underbrace{X_j'\mathbf{y}_j}_{K_jx1} \\ \\ \underbrace{X_j^{*\prime}\mathbf{y}_j}_{K_j^*x1} \end{pmatrix} - \begin{pmatrix} \underbrace{X_j'Y_j}_{K_jxG_j} \\ \\ \underbrace{X_j^{*\prime}Y_j}_{K_j^*xG_j} \end{pmatrix} \underbrace{\hat{\beta}_j^{MCI}}_{G_jx1} = \begin{pmatrix} \underbrace{X_j'X_j}_{K_jxK_j} & \underbrace{X_j'X_j^*}_{K_jxK_j^*} \\ \\ \underbrace{X_j^{*\prime}X_j}_{K_j^*xK_j} & \underbrace{X_j^{*\prime}X_j^*}_{K_j^*xK_j^*} \end{pmatrix} \begin{bmatrix} \underbrace{\hat{\gamma}_j^{MCI}}_{K_jx1} \\ \underbrace{\mathbf{0}}_{K_j^*x1} \end{bmatrix}$$

reorganizando se obtiene:

$$\begin{array}{ccccc} (X_j'Y_j)\,\hat{\beta}_j^{MCI} & + & (X_j'X_j)\,\hat{\gamma}_j^{MCI} & = & X_j'\mathbf{y}_j \\ (X_j^{*\prime}Y_j)\,\hat{\beta}_j^{MCI} & + & (X_j^{*\prime}X_j)\,\hat{\gamma}_j^{MCI} & = & X_j^{*\prime}\mathbf{y}_j \end{array}$$

En total son K ecuaciones con $(G_j + K_j)$ incógnitas. Puesto que la condición necesaria para la **identificación exacta** es

$$K_j^* = G_j \Leftrightarrow \underbrace{K_j^* + K_j}_{K} = G_j + K_j$$

resulta que el número de ecuaciones y de incógnitas es el mismo, de forma que, **si se satisface la condición de rango**, el sistema tiene una única solución para $\hat{\beta}_j^{MCI}$ y $\hat{\gamma}_j^{MCI}$. Si la ecuación está sobreidentificada, y se satisface la condición de rango, entonces hay más ecuaciones que incógnitas

$$K_j^* > G_j \Leftrightarrow \underbrace{K_j^* + K_j}_{K} > G_j + K_j$$

Esto implica que obtenemos distintas estimaciones para los mismos parámetros estructurales. La cuestión es que la sobreidentificación implica restricciones no lineales sobre los coeficientes de Π y esta se ha estimado sin imponer tales restricciones. Por lo tanto $\hat{\Pi}^{MCO}$ no las satisface dando lugar a este problema. Por este motivo, este método es adecuado para ecuaciones exactamente identificadas.

Si la ecuación no está identificada bien porque no se satistace la condición necesaria

$$K_j^* < G_j \Leftrightarrow \underbrace{K_j^* + K_j}_{K} < G_j + K_j$$

lo que implicaría menos ecuaciones que incógnitas, o no se satisface la de rango aún satisfaciéndose la de orden, no se podría resolver de forma única el sistema de ecuaciones. Habría infinitas soluciones.

MCI como Estimador de Variables Instrumentales.
Si volvemos a la ecuación estructural j-ésima que consideramos estimar

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

donde $Z_j = \left[\underbrace{Y_j}_{TxG_j} \quad \underbrace{X_j}_{TxK_j}\right] \qquad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$.

Si la ecuación está exactamente identificada el número de variables exógenas excluidas, K_j^*, será igual al número de endógenas incluidas en la ecuación j-ésima, G_j.

Por lo tanto, podemos utilizar como matriz de instrumentos a $X = (X_j \vdots X_j^*)$.

Estas variables exógenas están incorreladas con $\mathbf{u}_j$ y, están correladas con las variables endógenas, por lo que son buenos instrumentos[2]. Además al estar la ecuación exactamente identificada, la matriz $(X'Z_j)$ es una matriz cuadrada , ya que $K = G_j + K_j$, y podemos obtener el estimador de Variables Instrumentales de δ_j como

$$\hat{\delta}_j^{VI} = (X'Z_j)^{-1}X'\,\mathbf{y}_j$$

Este estimador es la solución al sistema de ecuaciones

$$(X'\,Z_j)\hat{\delta}_j^{VI} = X'\,\mathbf{y}_j$$

o equivalentemente

$$\begin{pmatrix} X_j'Y_j & X_j'X_j \\ X_j^{*\prime}Y_j & X_j^{*\prime}X_j \end{pmatrix}\begin{pmatrix} \hat{\beta}_j^{VI} \\ \hat{\gamma}_j^{VI} \end{pmatrix} = \begin{pmatrix} X_j'\mathbf{y}_j \\ X_j^{*\prime}\mathbf{y}_j \end{pmatrix}$$

Desarrollando obtenemos

$$\begin{array}{rcrcl} (X_j'Y_j)\,\hat{\beta}_j^{VI} & + & (X_j'X_j)\,\hat{\gamma}_j^{VI} & = & X_j'\mathbf{y}_j \\ (X_j^{*\prime}Y_j)\,\hat{\beta}_j^{VI} & + & (X_j^{*\prime}X_j)\,\hat{\gamma}_j^{VI} & = & X_j^{*\prime}\mathbf{y}_j \end{array}$$

Este es el mismo sistema de ecuaciones del que es solución el estimador $\hat{\beta}_j^{MCI}$ y $\hat{\gamma}_j^{MCI}$.

[2]Notar que en la Forma reducida cada una de las variables endógenas viene determinada por todas las exógenas del sistema

3.2. Mínimos Cuadrados en dos Etapas

El método de Mínimos Cuadrados en dos Etapas (MC2E) también es un método de estimación de Modelos de Ecuaciones Simultáneas con información limitada. Información limitada se refiere a que se estima una ecuación aislada de la Forma Estructural sin hacer uso de toda la información contenida en la especificación detallada del resto del modelo. La información se limita a la especificación de esa ecuación considerando tanto las variables endógenas y exógenas que están incluidas, y cuales son las excluidas de ella pero pertenecientes al modelo. No se utiliza la especificación concreta del resto de ecuaciones, ni tampoco de la matriz de varianzas y covarianzas del sistema.

Como veremos, esta técnica de estimación está indicada tanto si la ecuación está exactamente identificada como si está sobreidentificada. Se puede considerar como un método de estimación por Variables Instrumentales y en el caso de exacta identificación MC2E coincide con MCI.

Consideremos la j-ésima ecuación estructural del Modelo de Ecuaciones Simultáneas, después de haberle impuesto todas las restricciones a priori y el convenio de normalización[3]

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1}$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

donde $Z_j = \left[\underbrace{Y_j}_{TxG_j} \;\; \underbrace{X_j}_{TxK_j}\right] \qquad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$.

El método de MC2E, como su propio nombre indica, se puede llevar a cabo en dos etapas realizando una estimación por MCO en cada etapa:

- **1ª Etapa**: Se realiza la regresión de cada variable endógena en Y_j sobre **todas** las variables exógenas o predeterminadas del modelo. Esto es, se estima la forma reducida para Y_j:

  $$\underbrace{Y_j}_{TxG_j} = \underbrace{X}_{TxK}\underbrace{\Pi_j'}_{KxG_j} + \underbrace{V_j}_{TxG_j}$$

  $$\hat{\Pi}'_{j\,MCO} = (X'X)^{-1}X'Y_j$$

  $\hat{\Pi}'_{j\,MCO} = \left[\hat{\pi}_1^{MCO} \cdots \hat{\pi}_{G_j}^{MCO}\right]$, donde las columnas de $\hat{\Pi}'_{j\,MCO}$ son respectivamente los coeficientes de la regresión de cada variable en Y_j sobre X.

  Se obtienen los valores ajustados de Y_j dada la estimación de la forma reducida

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

[3]Por lo tanto $G = G_j + G_j^* + 1$ y $K = K_j + K_j^*$.

- **2ª Etapa** Se realiza la regresión de $\mathbf{y}_j$ sobre $\hat{Y}_j$ y X_j. Si denotamos la matriz de regresores de esta segunda regresión como $\hat{Z}_j = \left(\hat{Y}_j \;\; X_j\right)$, el estimador resultante de δ_j sería solución al sistema de ecuaciones normales $\left(\hat{Z}_j'\hat{Z}_j\right)\hat{\delta}_j^{MC2E} = \hat{Z}_j'\mathbf{y}_j$

$$\begin{pmatrix} \hat{Y}_j'\hat{Y}_j & \hat{Y}_j'X_j \\ X_j'\hat{Y}_j & X_j'X_j \end{pmatrix} \begin{pmatrix} \hat{\beta}_j^{MC2E} \\ \hat{\gamma}_j^{MC2E} \end{pmatrix} = \begin{pmatrix} \hat{Y}_j\mathbf{y}_j \\ X_j'\mathbf{y}_j \end{pmatrix}$$

En la matriz $\hat{Z}_j = \left(\hat{Y}_j \; X_j\right)$ todas las columnas son funciones lineales de las K columnas de X. Si la ecuación satisface la condición de rango para la identificación, y por lo tanto también la de orden ($K \geq G_j + K_j$), la matriz $\left(\hat{Z}_j'\hat{Z}_j\right)$ será de rango completo e invertible. El sistema tendrá una única solución,

$$\hat{\delta}_j^{MC2E} = \left(\hat{Z}_j'\hat{Z}_j\right)^{-1}\hat{Z}_j'\mathbf{y}_j$$

Si la ecuación no está identificada, bien porque no se satisface la de orden o, aunque esta se satisfaga, no se satisface la de rango, $(\hat{Z}_j'\hat{Z}_j)$ sería de rango reducido, no existiría su inversa y $\hat{\delta}_j^{MC2E}$ no estaría definido.

3.2.1. MC2E como estimador de Variables Instrumentales

Podemos escribir el estimador $\hat{\delta}_j^{MC2E}$ como un estimador de VI donde $\hat{Z}_j = \left(\hat{Y}_j \; X_j\right)$ es la matriz de instrumentos. Denotamos $\hat{V}_j$ la matriz (TxG_j) donde cada columna tiene el vector de residuos de la regresión de cada variable en Y_j sobre X, esto es, $Y_j = \hat{Y}_j + \hat{V}_j$. Entonces se satisface que cada columna de $\hat{V}_j$ es ortogonal a X, esto es:

$$\hat{V}_j'X = X'\hat{V}_j = 0$$

Por lo tanto:

$$\begin{array}{rclcl} \hat{Y}_j'\hat{V}_j & = & Y_j'X(X'X)^{-1}X'\hat{V}_j & = & 0 \\ \hat{Y}_j'\hat{Y}_j & = & \hat{Y}_j'(Y_j - \hat{V}_j) & = & \hat{Y}_j'Y_j \\ \hat{Y}_j'X_j & = & (Y_j - \hat{V}_j)'X_j & = & Y_j'X_j \end{array}$$

Así pues

$$\left(\hat{Z}_j'\hat{Z}_j\right) = \begin{pmatrix} \hat{Y}_j'\hat{Y}_j & \hat{Y}_j'X_j \\ X_j'\hat{Y}_j & X_j'X_j \end{pmatrix} = \begin{pmatrix} \hat{Y}_j'Y_j & \hat{Y}_j'X_j \\ X_j'Y_j & X_j'X_j \end{pmatrix} = \left(\hat{Z}_j'Z_j\right)$$

Por lo tanto

$$\hat{\delta}_j^{MC2E} = \left(\hat{Z}_j'\hat{Z}_j\right)^{-1}\hat{Z}_j'\mathbf{y}_j = \left(\hat{Z}_j'Z_j\right)^{-1}\hat{Z}_j'\mathbf{y}_j$$

por lo que $\hat{\delta}_j^{MC2E}$ es un estimador de VI donde la matriz de instrumentos es $\hat{Z}_j$.

Denotamos por $M_X = I_T - X\,(X'X)^{-1}\,X' = I_T - P_X$. Tanto M_X como P_X son matrices simétricas e idempotentes[4]. Dado que $M_X X = (I - X\,(X'X)^{-1}\,X')X = \mathbf{0}$ entonces

$$M_X(X_j \vdots X_j^*) = (\mathbf{0} \ \vdots \mathbf{0}) \Leftrightarrow M_X X_j = \mathbf{0} \quad M_X X_j^* = \mathbf{0}$$

Por lo tanto

$$X_j'(I - M_X) = X_j' \Leftrightarrow X_j' P_X = X_j' \Leftrightarrow X_j' X\left(X'X\right)^{-1} X' = X_j'$$

Así mismo tenemos que

$$\hat{Z}_j = \left[\hat{Y}_j \ \ X_j\right] = \left[X(X'X)^{-1}X'Y_j \ \ X_j\right] = \left[P_X Y_j \ \ P_X X_j\right] = P_X\left[Y_j \ \ X_j\right] = P_X Z_j$$

Utilizando este resultado podemos obtener el resultado anterior,

$$\hat{\delta}_j^{MC2E} = \left(\hat{Z}_j'\hat{Z}_j\right)^{-1}\hat{Z}_j'\mathbf{y}_j = \left(Z_j' \underbrace{P_X}_{P_X P_X} Z_j\right)^{-1} Z_j' \underbrace{P_X}_{P_X P_X} \mathbf{y}_j = \left(\hat{Z}_j' Z_j\right)^{-1}\hat{Z}_j'\mathbf{y}_j$$

y la siguiente forma de obtener el estimador MC2E de δ_j,

$$\hat{\delta}_j^{MC2E} = \left(Z_j'X\left(X'X\right)^{-1}X'Z_j\right)^{-1} Z_j'X\left(X'X\right)^{-1}X'\mathbf{y}_j$$

3.2.2. Equivalencia entre MC2E y MCI en el caso de exacta identificación

Si la ecuación está exactamente identificada, además de satisfacerse la condición de rango, la condición de orden es tal que $K = G_j + K_j$. Esto implica que $(Z_j'X)$ es una matriz cuadrada (KxK), de rango completo y existe su inversa. Entonces

$$\left(Z_j'X\left(X'X\right)^{-1}X'Z_j\right)^{-1} = \left(X'Z_j\right)^{-1}\left(X'X\right)\left(Z_j'X\right)^{-1}$$

y sustituyendo en la expresión anterior para el estimador MC2E obtenemos,

$$\hat{\delta}_j^{MC2E} = \left(X'Z_j\right)^{-1}\left(X'X\right)\left(Z_j'X\right)^{-1}\left(Z_j'X\right)\left(X'X\right)^{-1}X'\mathbf{y}_j = \left(X'Z_j\right)^{-1}X'\mathbf{y}_j = \hat{\delta}_j^{MCI}$$

3.3. Propiedades de los estimadores MCI y MC2E. Inferencia.

La ecuacion de interés de la forma estructural a estimar era

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$$

[4]Entonces $M_X = M_X' \quad M_X M_X = M_X \quad P_X = P_X' \quad y P_X P_X = P_X$

donde $Z_j = [Y_j \;\; X_j]$ $\quad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$.

Si en la ecuación aparecen variables endógenas como variables explicativas, la matriz Z_j es estocástica ya que las columnas correspondientes a Y_j se determinan en el sistema a partir de los errores estructurales, por lo que no son independientes de estos y en general estarán correlacionados con el término de perturbación de la ecuación. Las variables exógenas X son independientes de los errores estructurales.

Hemos obtenido los estimadores de MCI y MC2E de los parámetros δ_j de interés como estimadores de Variables Instrumentales:

$$\hat{\delta}_j^{MCI} = \left(X'Z_j\right)^{-1} X'\,\mathbf{y}_j = \delta_j + \left(X'Z_j\right)^{-1} X'\,\mathbf{u}_j$$

donde la matriz de instrumentos es X, la matriz de datos de las variables exógenas del sistema.

$$\hat{\delta}_j^{MC2E} = \left(\hat{Z}_j'Z_j\right)^{-1} \hat{Z}_j'\mathbf{y}_j = \delta_j + \left(\hat{Z}_j'Z_j\right)^{-1} \hat{Z}_j'\mathbf{u}_j$$

donde la matriz de instrumentos es $\hat{Z}_j = \left[\hat{Y}_j \;\; X_j\right] = \left[X(X'X)^{-1}X'Y_j \quad X_j\right] = [P_X Y_j \quad X_j]$. En el caso de exacta identificación ambos estimadores coinciden.

Las propiedades de estos estimadores para muestras finitas en general son desconocidas ya que son estimadores no lineales. Notar que aparece la matriz Z_j, donde está Y_j, de forma no lineal, por lo tanto estos estimadores son función no lineal de las perturbaciones estructurales. En general serán estimadores sesgados, ya que si en Z_j aparece Y_j, estas variables no son independientes de $\mathbf{u}_j$, por lo que $E(\mathbf{u}_j \setminus Z_j, X) \neq E(\mathbf{u}_j) = 0$. Podemos obtener propiedades asintóticas, para tamaños de muestra grande.

Para analizar la consistencia consideramos el límite en probabilidad del estimador,

$$plim\ \hat{\delta}_j^{MCI} = \delta_j + \left(plim\,\frac{1}{T}\,X'Z_j\right)^{-1} plim\ \left(\frac{1}{T}\,X'\,\mathbf{u}_j\right)$$

$$plim\ \hat{\delta}_j^{MC2E} = \delta_j + \left(plim\,\frac{1}{T}\hat{Z}_j'Z_j\right)^{-1} plim\ \left(\frac{1}{T}\,\hat{Z}_j'\mathbf{u}_j\right)$$

Y para analizar la distribución asintótica analizaremos la convergencia en distribución de

$$\sqrt{T}\left(\hat{\delta}_j^{MCI} - \delta_j\right) = \left(\frac{X'Z_j}{T}\right)^{-1} \frac{X'\mathbf{u}_j}{\sqrt{T}}$$

$$\sqrt{T}\left(\hat{\delta}_j^{MC2E} - \delta_j\right) = \left(\frac{\hat{Z}_j'Z_j}{T}\right)^{-1} \frac{\hat{Z}_j\mathbf{u}_j}{\sqrt{T}}$$

Por el teorema de Mann-Wald, dado que

i) $E(\mathbf{u}_j) = 0$, $E(\mathbf{u}_j\mathbf{u}_j') = \sigma_j^2 I_T$.

ii) $E(X_i'u) = 0$, $i = 1, 2, ..., K$ donde X_i es la columna i-ésima de la matriz X.

iii) $plim\left(\frac{1}{T}X'X\right)$ es una matriz finita, simétrica y definida positiva, llamémosla Q_{XX}.

Entonces se tiene que:

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

b) $\frac{1}{\sqrt{T}}X'\mathbf{u}_j \xrightarrow{d} N\left(\mathbf{0}, \sigma_j^2 Q_{XX}\right)$

Bajo las condiciones del teorema de Mann-Wald, y suponiendo que

iv) $plim\left(\frac{1}{T}X'Z_j\right)$ es una matriz finita, y definida positiva, llamémosla Q_{XZ_j}

Entonces, el estimador $\hat{\delta}_j^{MCI}$ es consistente,

$$plim\,\hat{\delta}_j^{MCI} = \delta_j + \left(plim\,\frac{1}{T}X'Z_j\right)^{-1} plim\left(\frac{1}{T}X'\mathbf{u}_j\right) = \delta_j + Q_{Z_jX}^{-1}\;\mathbf{0} \;=\; \delta_j$$

En cuanto a su distribución asintótica, utilizando el teorema de Cramer

$$\sqrt{T}\left(\hat{\delta}_j^{MCI} - \delta_j\right) = \underbrace{\left(\frac{X'Z_j}{T}\right)^{-1}}_{\xrightarrow{p} Q_{XZ_j}^{-1}} \quad \underbrace{\frac{1}{\sqrt{T}}X'\mathbf{u}_j}_{\xrightarrow{d} N(\mathbf{0},\ \sigma_j^2 Q_{XX})} \quad \xrightarrow{d} N\left(\mathbf{0},\ \sigma_j^2 Q_{XZ_j}^{-1} Q_{XX} Q_{Z_jX}^{-1}\right)$$

$$\sqrt{T}\left(\hat{\delta}_j^{MCI} - \delta_j\right) \xrightarrow{d} N\left(\mathbf{0},\ \sigma_j^2 Q_{XZ_j}^{-1} Q_{XX} Q_{Z_jX}^{-1}\right)$$

Un estimador consistente de la matriz de varianzas y covarianzas asintótica es[5]

$$\widehat{Va}(\sqrt{T}\hat{\delta}_j^{MCI}) = \hat{\sigma}_j^2 \left(\frac{1}{T}X'Z_j\right)^{-1}\left(\frac{1}{T}X'X\right)\left(\frac{1}{T}Z_j'X\right)^{-1} = T\,\hat{\sigma}_j^2\,(X'Z_j)^{-1}\,(X'X)\,(Z_j'X)^{-1}$$

siendo $\hat{\sigma}_j^2 = \frac{\hat{\mathbf{u}}_j\hat{\mathbf{u}}_j'}{T} = \frac{1}{T}\left(\mathbf{y}_j - Z_j\hat{\delta}_j^{MCI}\right)'\left(\mathbf{y}_j - Z_j\hat{\delta}_j^{MCI}\right)$. Esto mismo aplica para el caso del estimador MC2E con identificación exacta, ya que ambos estimadores en ese caso coinciden.

En general, bajo las condiciones anteriores,

$$\bullet\; plim\,\frac{1}{T}\hat{Z}_j'\mathbf{u}_j = \begin{pmatrix} plim\,\frac{1}{T}\hat{Y}_j'\mathbf{u}_j \\ \\ plim\,\frac{1}{T}X_j'\mathbf{u}_j \end{pmatrix} = \begin{pmatrix} \Pi_j\,Q_{XX}\,Q_{XX}^{-1}\,\mathbf{0} \\ \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \\ \mathbf{0} \end{pmatrix}$$

siendo

$$plim\,\frac{1}{T}\hat{Y}_j'\mathbf{u}_j = plim\,\frac{1}{T}(X(X'X)^{-1}X'Y_j)'\mathbf{u}_j = \underbrace{plim\,\frac{1}{T}Y_j'X}_{\Pi_j\,Q_{XX}}\;\underbrace{(plim\,\frac{1}{T}X'X)^{-1}}_{Q_{XX}^{-1}}\;\underbrace{plim\,\frac{1}{T}X'\mathbf{u}_j}_{\mathbf{0}}$$

[5]Bajo las condiciones i) a iv).

$$plim\,\frac{1}{T}Y_j'X = plim\,\frac{1}{T}(X\Pi_j' + V_j)'X = \Pi_j plim\,\frac{1}{T}X'X + plim\,\frac{1}{T}V_j'X = \Pi_j\,Q_{XX} + \mathbf{0}$$

Por otro lado,

$$\bullet\; plim\,\frac{1}{T}\hat{Z}_j'Z_j = \begin{pmatrix} plim\,\frac{1}{T}\hat{Y}_j'Y_j & plim\,\frac{1}{T}\hat{Y}_j'X_j \\ plim\,\frac{1}{T}X_j'Y_j & plim\,\frac{1}{T}X_j'X_j \end{pmatrix} = \begin{pmatrix} \Pi_j Q_{XX}\Pi_j' & \Pi_j\,Q_{XX_j} \\ Q_{XX_j}'\,\Pi_j' & Q_{X_jX_j} \end{pmatrix} = Q_{\hat{Z}_jZ_j}$$

siendo

$$plim\,\frac{1}{T}\hat{Y}_j'Y_j = plim\,\frac{1}{T}\left[(X\hat{\Pi}_j')'(X\Pi_j' + V_j)\right] = \underbrace{plim\,\hat{\Pi}_j\,plim\,\frac{1}{T}(X'X)\Pi_j'}_{\Pi_jQ_{XX}\Pi_j'} + \underbrace{plim\,\hat{\Pi}_j\,plim\,\frac{1}{T}(X'V_j)}_{\mathbf{0}}$$

$$plim\,\frac{1}{T}\hat{Y}_j'X_j = plim\,\frac{1}{T}\left[(X\hat{\Pi}_j')'X_j\right] = plim\,\hat{\Pi}_j\,plim\,\frac{1}{T}(X'X_j) = \Pi_j\,Q_{XX_j}$$

Por lo tanto, si la ecuación está identificada $plim\,\frac{1}{T}\hat{Z}_j'Z_j$ será invertible y el estimador MC2E consistente,

$$plim\,\hat{\delta}_j^{MC2E} = \delta_j + \left(plim\,\frac{1}{T}\,\hat{Z}_j'Z_j\right)^{-1} plim\,\left(\frac{1}{T}\,\hat{Z}_j'\,\mathbf{u}_j\right) = \delta_j + Q_{\hat{Z}_jZ_j}^{-1}\;\mathbf{0} \;=\; \delta_j$$

En cuanto a la distribución asintótica del estimador, tenemos que, aplicando Cramer y resultados anteriores,

$$\frac{1}{\sqrt{T}}\hat{Z}_j'\mathbf{u}_j = \underbrace{\left(\frac{1}{T}\,Z_j'X\right)}_{\xrightarrow{p} Q_{Z_jX}} \underbrace{\left(\frac{1}{T}\,X'X\right)^{-1}}_{\xrightarrow{p} Q_{XX}^{-1}} \underbrace{\left(\frac{1}{\sqrt{T}}\,X'\mathbf{u}_j\right)}_{\xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2 Q_{XX}\right)} \xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2\;\underbrace{Q_{Z_jX}Q_{XX}^{-1}Q_{XZ_j}}_{Q_{Z_jX}Q_{XX}^{-1}Q_{XX}Q_{XX}^{-1}Q_{XZ_j}}\right)$$

$$\frac{1}{\sqrt{T}}\hat{Z}_j'\mathbf{u}_j \xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2 Q_{\hat{Z}_jZ_j}\right)$$

Utilizando de nuevo el Teorema de Cramer,

$$\sqrt{T}\left(\hat{\delta}_j^{MC2E} - \delta_j\right) = \underbrace{\left(\frac{\hat{Z}_j'Z_j}{T}\right)^{-1}}_{\xrightarrow{p} Q_{\hat{Z}_jZ_j}^{-1}} \quad \underbrace{\frac{\hat{Z}_j\mathbf{u}_j}{\sqrt{T}}}_{\xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2 Q_{\hat{Z}_jZ_j}\right)} \quad \xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2 Q_{\hat{Z}_jZ_j}^{-1}\right)$$

$$\sqrt{T}\left(\hat{\delta}_j^{MC2E} - \delta_j\right) \xrightarrow{d} N\left(\mathbf{0},\;\sigma_j^2 Q_{\hat{Z}_jZ_j}^{-1}\right)$$

Un estimador consistente de la matriz de varianzas y covarianzas asintótica es[6]

$$\widehat{Va}(\sqrt{T}\hat{\delta}_j^{MC2E}) = \hat{\sigma}_j^2\left(\frac{1}{T}\,\hat{Z}_j'\hat{Z}_j\right)^{-1} = \hat{\sigma}_j^2\left(\frac{1}{T}\hat{Z}_j'Z_j\right)^{-1} = T\,\hat{\sigma}_j^2\left[\left(Z_j'X\right)\left(X'X\right)^{-1}\left(X'Z_j\right)\right]^{-1}$$

[6] Bajo las condiciones i) a iv).

siendo $\hat{\sigma}_j^2 = \frac{\hat{\mathbf{u}}_j \hat{\mathbf{u}}_j'}{T} = \frac{1}{T} \left(\mathbf{y}_j - Z_j \hat{\delta}_j^{MC2E}\right)' \left(\mathbf{y}_j - Z_j \hat{\delta}_j^{MC2E}\right)$

Comentario:

Es incorrecto calcular $\hat{\sigma}_j^2$ utilizando la suma de cuadrados residual de la segunda etapa, ya que en ese caso se obtendría $\left(\mathbf{y}_j - \hat{Z}_j \hat{\delta}_j^{MC2E}\right)' \left(\mathbf{y}_j - \hat{Z}_j \hat{\delta}_j^{MC2E}\right)$.

Dado que $\hat{Z}_j \neq Z_j$ esta suma de cuadrados residual no sería igual a $\left(\mathbf{y}_j - Z_j \hat{\delta}_j^{MC2E}\right)' \left(\mathbf{y}_j - Z_j \hat{\delta}_j^{MC2E}\right)$ siendo esta última la adecuada para que el estimador de σ_j^2 sea consistente.

Inferencia sobre δ_j utilizando MCI o MC2E.
Utilizando los resultados anteriores sobre distribución asintótica y estimación de las matrices de covarianzas asintóticas de estos estimadores, se puede derivar el estadístico de Wald para contrastar restricciones lineales sobre δ_j. En general, podemos escribir la hipótesis nula y la alternativa como:

$$H_0 : \underset{(q \times (G_j + K_j))}{R} \cdot \underset{((G_j + K_j) \times 1)}{\delta_j} = \underset{(q \times 1)}{r}$$

$$H_A : R\delta_j \neq r$$

siendo q el número de restricciones bajo la hipótesis nula y $G_j + K_j$ el número de parámetros en la ecuación estructural de interés $\mathbf{y}_j = Z_j \, \delta_j + \mathbf{u}_j$. La hipótesis alternativa implicaría que **al menos una** de las igualdades no se satisface[7].

El estadístico de contraste y su distribución asintótica bajo H_0 son:

$$T(R\,\hat{\delta}_j - r)' \, [R\,\widehat{Va}(\sqrt{T}\hat{\delta}_j)\,R']^{-1} \, (R\,\hat{\delta}_j - r) \xrightarrow{d,H_0} \chi_q^2$$

Si $q = 1$ se puede utilizar también el estadístico

$$\frac{\sqrt{T}(R\,\hat{\delta}_j - r)}{\sqrt{R\,\widehat{Va}(\sqrt{T}\hat{\delta}_j)\,R'}} \xrightarrow{d,H_0} \mathcal{N}(0,1)$$

Para el caso de MCI o MC2E con exacta identificación, la expresión del estadístico para $q \geq 1$ quedaría

$$(R\,\hat{\delta}_j^{MCI} - r)' \left[R\,\hat{\sigma}_j^2 \left(X'Z_j\right)^{-1} \left(X'X\right) \left(Z_j'X\right)^{-1} R'\right]^{-1} (R\,\hat{\delta}_j^{MCI} - r) \xrightarrow{d,H_0} \chi_q^2$$

[7] En la ecuación se satisfacen las condiciones anteriores para la derivación de los resultados de consistencia y distribución asintótica.

y para MC2E en general,

$$(R\,\hat{\delta}_j^{MC2E} - r)' \left[R\,\hat{\sigma}_j^2 \underbrace{\left[\left(Z_j'X\right) \left(X'X\right)^{-1} \left(X'Z_j\right) \right]^{-1}}_{\left(\hat{Z}_j'\hat{Z}_j\right)^{-1}} R' \right]^{-1} (R\,\hat{\delta}_j^{MC2E} - r) \xrightarrow{d,H_0} \chi_q^2$$

Para $q = 1$, el estadístico en cada caso quedaría,

$$\frac{(R\,\hat{\delta}_j^{MCI} - r)}{\sqrt{R\,\hat{\sigma}_j^2\ (X'Z_j)^{-1} (X'X) \left(Z_j'X\right)^{-1}\ R'}} \xrightarrow{d,H_0} \mathcal{N}\,(0,1)$$

$$\frac{(R\,\hat{\delta}_j^{MC2E} - r)}{\sqrt{R\,\hat{\sigma}_j^2\ \left[\left(Z_j'X\right) (X'X)^{-1} (X'Z_j) \right]^{-1}\ R'}} \xrightarrow{d,H_0} \mathcal{N}\,(0,1)$$

Dado un nivel de significación, la regla de decisión del contraste en cada caso será la habitual, bien utilizando el valor crítico en la distribución asintótica bajo H_0 o el valor p.

3.4. Máxima Verosimilitud con Información Limitada bajo Normalidad

El método de estimación por Máxima Verosimilitud Información Limitada (MVIL) considera la estimación de los parámetros estructurales de una ecuación, teniendo en cuenta solamente la función de verosimilitud de las variables endógenas de esa ecuación y las restricciones de identificación correspondientes a la ecuación a estimar. Esto es Información Limitada, no se tiene en cuenta en la verosimilitud el resto de variables endógenas del sistema ni las restricciones de identificación impuestas en otras ecuaciones.

Consideremos la ecuación j-ésima a estimar antes de imponer la normalización

$$\beta_{jj} \underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j} \underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j} \underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1}$$

$$\underbrace{\begin{bmatrix} \mathbf{y}_j & Y_j \end{bmatrix}}_{Tx(G_j+1)} \underbrace{\begin{bmatrix} \beta_{jj} \\ -\beta_j \end{bmatrix}}_{(G_j+1)x1} = \underbrace{X_j}_{TxK_j} \underbrace{\gamma_j}_{K_jx1} + \mathbf{u}_j \qquad \Leftrightarrow \qquad Y_j^+ \beta_j^+ = X_j\gamma_j + \mathbf{u}_j$$

Las ecuaciones de la Forma Reducida para Y_j^+ son las siguientes

$$\underbrace{Y_j^+}_{Tx(G_j+1)} = \underbrace{X}_{TxK} \underbrace{\Pi_j^{+\prime}}_{Kx(G_j+1)} + \underbrace{V_j^+}_{Tx(G_j+1)} = \underbrace{X_j}_{TxK_j} \underbrace{\Pi_{j1}^{+\prime}}_{K_jx(G_j+1)} + \underbrace{X_j^*}_{TxK_j^*} \underbrace{\Pi_{j2}^{+\prime}}_{K_j^*x(G_j+1)} + \underbrace{V_j^+}_{Tx(G_j+1)}$$

Denotemos por $\mathbf{v}_{jt}^{+\prime}$ al vector $(1x(G_j+1))$ correspondiente a la fila t-ésima de la matriz V_j^+. Dado el supuesto sobre la distribución de los errores estructurales $\mathbf{u}_t \sim NID(\mathbf{0}, \Sigma)$, el vector de errores de la forma reducida correspondiente a las ecuaciones de las variables Y_j^+ en el momento t, se distribuye $\mathbf{v}_{jt}^+ \sim NID(\mathbf{0}, \Omega_j^+)$ donde Ω_j^+ es la submatriz correspondiente de Ω.

Si denotamos por $\mathbf{y}_{jt}^{+\prime}$ y $\mathbf{x}_{jt}^{+\prime}$ a la fila t-ésima de Y_j^+ y de X respectivamente

$$\mathbf{y}_{jt}^{+\prime} = \mathbf{x}_{jt}^{+\prime}\Pi_j^{+\prime} + \mathbf{v}_{jt}^{+\prime} \qquad \Longrightarrow \qquad \mathbf{y}_{jt}^{+} = \Pi_j^{+\prime}\mathbf{x}_{jt}^{+} + \mathbf{v}_{jt}^{+} \qquad \Longleftrightarrow \qquad \mathbf{y}_{jt}^{+} - \mathbf{x}_{jt}^{+}\Pi_j^{+} = \mathbf{v}_{jt}^{+}$$

La función de verosimilitud de Y_j^+ es:

$$L_j = L(\Pi_j^+, \Omega_j^+; Y_j^+, X) = \prod_{t=1}^T L(\Pi_j^+, \Omega_j^+; \mathbf{y}_{jt}^+, \mathbf{x}_{jt}) =$$

$$= \prod_{t=1}^{T} \Big((2\pi)^{-(G_j+1)/2} |\Omega_j^+|^{-1/2} \exp\left(-1/2\right)(\mathbf{y}_{jt}^+ - \mathbf{x}_{jt}^+\Pi_j^+)'(\Omega_j^+)^{-1}(\mathbf{y}_{jt}^+ - \mathbf{x}_{jt}^+\Pi_j^+) \Big)$$

Tomando logaritmos, obtenemos

$$lnL_j = -(T(G_j+1)/2)\, ln\, 2\pi - T/2\, ln\, |\Omega_j^+| - 1/2 \left[\sum_{t=1}^{T} (\mathbf{y}_{jt}^+ - \mathbf{x}_{jt}^+\Pi_j^+)'(\Omega_j^+)^{-1}(\mathbf{y}_{jt}^+ - \mathbf{x}_{jt}^+\Pi_j^+) \right]$$

El estimador MVIL maximiza esta función sujeto a las restricciones que relacionan los parámetros estructurales de la ecuación j-ésima con los parámetros de la forma reducida.

De las restricciones de exclusión de la ecuación j-ésima y la relación entre los parámetros de la Forma Estructural y la Reducida tenemos las siguientes K ecuaciones: Utilizando esta notación,

$$\Pi_j^+ = \begin{bmatrix} \underbrace{\pi_j'}_{(1\times K_j)} & \underbrace{\pi_j*'}_{(1\times K_j^*)} \\ \underbrace{\Pi_{1j}}_{(G_j\times K_j)} & \underbrace{\Pi_{1j}^*}_{(G_j\times K_j^*)} \end{bmatrix} = \begin{bmatrix} \Pi_{j1}^+ & \Pi_{j2}^+ \end{bmatrix}$$

$$B_j'\Pi = -\Gamma_j' \quad \Leftrightarrow \quad \begin{bmatrix} \underbrace{1}_{\beta_{jj}} & \underbrace{-\beta_j'}_{(1\times G_j)} & \underbrace{\mathbf{0}'}_{1\times G_j^*)} \end{bmatrix} \begin{bmatrix} \underbrace{\pi_j'}_{(1\times K_j)} & \underbrace{\pi_j*'}_{(1\times K_j^*)} \\ \underbrace{\Pi_{1j}}_{(G_j\times K_j)} & \underbrace{\Pi_{1j}^*}_{(G_j\times K_j^*)} \\ \underbrace{\Pi_{2j}}_{(G_j^*\times K_j)} & \underbrace{\Pi_{2j}^*}_{(G_j^*\times K_j^*)} \end{bmatrix} = \begin{bmatrix} \underbrace{\gamma_j'}_{(1\times K_j)} & \underbrace{\mathbf{0}'}_{(1\times K_j^*)} \end{bmatrix}$$

o equivalentemente

$$\begin{bmatrix} \underbrace{\pi_j}_{(K_j x1)} & \underbrace{\Pi'_{1j}}_{(K_j \times G_j)} & \underbrace{\Pi'_{2j}}_{(K_j x G_j^*)} \\ \underbrace{\pi_j *}_{(K_j^* x1)} & \underbrace{\Pi^{*\prime}_{1j}}_{(K_j^* \times G_j)} & \underbrace{\Pi^{*\prime}_{2j}}_{(K_j^* \times G_j^*)} \end{bmatrix} \begin{bmatrix} \underbrace{1}_{\beta_{jj}} \\ \underbrace{-\beta_j}_{(G_j x1)} \\ \underbrace{\mathbf{0}}_{(G_j^* x1)} \end{bmatrix} = \begin{bmatrix} \underbrace{\gamma_j}_{(K_j x1)} \\ \underbrace{\mathbf{0}}_{(K_j^* x1)} \end{bmatrix}$$

$$\Longleftrightarrow$$

$$\begin{bmatrix} \underbrace{\Pi^{+\prime}_{j1}}_{(K_j \times (G_j+1))} & \underbrace{\Pi'_{2j}}_{(K_j x G_j^*)} \\ \underbrace{\Pi^{+\prime}_{j2}}_{(K_j \times G_j^*)} & \underbrace{\Pi^{*\prime}_{2j}}_{(K_j^* \times G_j^*)} \end{bmatrix} \begin{bmatrix} \underbrace{\beta_j^+}_{((G_j+1)x1)} \\ \underbrace{\mathbf{0}}_{(G_j^* x1)} \end{bmatrix} = \begin{bmatrix} \underbrace{\gamma_j}_{(K_j x1)} \\ \underbrace{\mathbf{0}}_{(K_j^* x1)} \end{bmatrix}$$

Por lo tanto obtenemos las siguientes ecuaciones:

$$\Pi^{+'}_{j1}\beta_j^+ = \gamma_j \qquad K_j \quad \texttt{ecuaciones} \tag{3.1}$$

$$\Pi^{+'}_{j2}\beta_j^+ = \mathbf{0} \qquad K_j^* \quad \texttt{ecuaciones} \tag{3.2}$$

Las primeras K_j ecuaciones sólo muestran cómo obtener γ_j dado β_j^+. Si podemos resolver para β_j^+ de (3.2), podemos utilizar esta solución en (3.1) para obtener γ_j y por lo tanto, los parámetros estructurales de la ecuación j estarán identificados. La condición de Rango para la identificación de la ecuación establece que $rango(\Pi^{+'}_{2j}) = G_j$. La condición necesaria es que $G_j \geq K_j^*$. Por lo tanto, las restricciones de identificación relevantes a tener en cuenta a la hora de la estimación por MVIL en la ecuación j-ésima son (3.2). Hay tres casos a considerar:

1) La ecuación no está identificada. No hay estimación posible. No podemos recuperar los parámetros estructurales a partir de los de la Forma Reducida.

2) La ecuación está exactamente identificada. No hay restricciones sobre Π^+_{j2}. Dada la normalización

$$\pi_j^* - \Pi^{*'}_{1j}\beta_j = \mathbf{0} \Longrightarrow \beta_j = (\Pi^{*'}_{1j})^{-1}\pi_j^*$$

ya que la matriz $\Pi^{*'}_{1j}$ es cuadrada y de rango completo por lo que existe su inversa.

En este caso, en ausencia de toda restricción sobre los coeficientes de la forma reducida, el estimador Maximo Verosimil de β_j se obtiene de

$\hat{\beta}_j^{MVIL} = (\hat{\Pi}^{*'}_{1j})^{-1}\hat{\pi}_j^*$ y el de γ_j como $\hat{\gamma}_j = \hat{\pi}_j - \hat{\Pi}'_{1j}\hat{\beta}_j^{MVIL}$

donde el estimador de Π_j^+ que maximiza el logaritmo de la verosimilitud coincide con el estimador MCO. Por lo tanto, en este caso el estimador de β_j y γ_j por MVIL coincide con MCI y entonces también con MC2E.

3) La ecuación esta sobreidentificada. La sobreidentificación impone restricciones sobre Π_j^+. Las restricciones (3.2) son importantes. En este caso la solución analítica al problema de maximizar lnL_j sujeto a $\Pi_{j2}^{+'}\beta_j^+ = \mathbf{0}$ con respecto a Π_j^+, β_j^+ y Ω_j^+ es extremadamente larga y complicada.

Se ha demostrado que el estimador MVIL de β_j^+ coincide con el estimador que minimiza el siguiente ratio de varianzas residuales:

$$l = \frac{\beta_j^{+\prime} A \beta_j^+}{\beta_j^+ S \beta_j^+}$$

donde

$$\beta_j^{+\prime} A \beta_j^+ = \beta_j^{+\prime} Y_j^{+\prime} M_j Y_j^+ \beta_j^+$$

siendo $M_j = I_T - X_j(X_j'X_j)^{-1}X_j'$

y

$$\beta_j^{+\prime} S \beta_j^+ = \beta_j^{+\prime} Y_j^{+\prime} M Y_j^+ \beta_j^+$$

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

El numerador del cociente es la suma de cuadrados residual de la regresión de $\tilde{y}_j \equiv Y_j^+\beta_j^+$, dado β_j^+, sobre las variables exógenas X_j que son las incluidas en la ecuación estructural a estimar

$$\underbrace{Y_j^+\beta_j^+}_{\tilde{y}_j} = X_j\gamma_j + u_j$$

El denominador es la suma de cuadrados residual de la regresión de $\tilde{y}_j \equiv Y_j^+\beta_j^+$, dado β_j^+, sobre todas las variables exógenas X. Dado que en la ecuación estructural solamente las variables exógenas X_j son importantes para explicar $\tilde{y}_j \equiv Y_j^+\beta_j^+$, entonces al incluir el resto de variables exógenas en la regresión la disminución de la suma de cuadrados residual debe de ser mínima.

Una vez obtenido el estimador de β_j^+, impuesta la normalización, que minimiza ese ratio de varianzas residuales, $\hat{\beta}_j^{+\prime} = (1, \ -\hat{\beta}_j^{MVIL})$, el estimador MVIL de γ_j se puede obtener de resolver para $\hat{\gamma}_j$

$$X_j'(\mathbf{y}_j - Y_j\hat{\beta}_j^{MVIL}) = X_j'X_j\hat{\gamma}_j \quad \Rightarrow \quad \hat{\gamma}_j^{MVIL} = (X_j'X_j)^{-1}X_j'(\mathbf{y}_j - Y_j\hat{\beta}_j^{MVIL})$$

que es simplemente el resultado de la regresión de $(\mathbf{y}_j - Y_j\hat{\beta}_j^{MVIL})$ sobre X_j.

El estimador $\hat{\beta}_j^{MVIL}$ se puede obtener como sigue:

a) Calcular las matrices

$$S_j^0 = Y_j^{+\prime} M_j Y_j^+ = Y_j^{+\prime}(I_T - X_j(X_j'X_j)^{-1}X_j')Y_j^+ = E_j^{0\prime}E_j^0$$

$$S_j^1 = Y_j^{+\prime} M Y_j^+ = Y_j^{+\prime}(I_T - X(X'X)^{-1}X')Y_j^+ = E_j^{1\prime}E_j^1$$

Cada columna de $E_j^0 = M_j Y_j^+$ es el vector de residuos minimocuadráticos obtenidos de la regresión de la columna correspondiente de Y_j^+ sobre X_j. Para $E_j^1 = MY_j^+$ es lo mismo pero incluyendo en las regresiones también a X_j^*, esto es todas las variables exógenas X.

b) Se obtienen las raíces características de la matriz $(S_j^1)^{-1}S_j^0$ o equivalentemente de la matriz $D = (S_j^1)^{-1/2}S_j^0(S_j^1)^{-1/2}$. Todas las raíces son reales y mayores o iguales a la unidad.

Se elige la menor raíz característica λ_1

c) Particionamos la matriz S_j^0 y S_j^1

$$S_j^0 = \begin{pmatrix} s_{jj}^0 & \mathbf{s}_j^{0\prime} \\ \underline{\mathbf{s}}_j^0 & S_{jj}^0 \end{pmatrix} \qquad S_j^1 = \begin{pmatrix} s_{jj}^1 & \mathbf{s}_j^{1\prime} \\ \underline{\mathbf{s}}_j^1 & S_{jj}^1 \end{pmatrix}$$

Con estos elementos disponibles se puede obtener

$$\hat{\beta}_j^{MVIL} = \left[S_{jj}^0 - \lambda_1 S_{jj}^1\right]^{-1}(\mathbf{s}_j^0 - \lambda_1 \mathbf{s}_j^1)$$

La matriz asintótica de varianzas y covarianzas para el estimador de MVIL es idéntica a la del estimador MC2E. La implicación es que, con perturbaciones normalmente distribuidas, MC2E es eficiente asintóticamente dentro de los estimadores de Información Limitada.

Si la ecuación está identificada, puede demostrarse que $\lambda_1 = 1$, lo que conduce al estimador MCI.

Capítulo 4

Métodos de Información Completa

4.1. Mínimos Cuadrados en tres Etapas

Un método de información completa considera estimar conjuntamente el sistema de ecuaciones de la forma estructural. Por lo tanto, requiere la especificación concreta de cada una de las ecuaciones de la forma estructural del sistema y todas ellas tienen que estar identificadas. La motivación para considerar un método de información completa es que en general son más eficientes asintóticamente ya que incorporan toda la información del sistema: todas las restricciones y la matriz de varianzas y covarianzas de los términos de error estructurales.

Veremos que hay casos en los estimar una ecuación por información completa no mejora en eficiencia frente a estimarla de forma aislada con información limitada. Otra cuestión es si alguna o algunas de las ecuaciones del sistema está mal especificada porque esto puede producir estimadores no consistentes de otra ecuación del sistema.

Vamos a proceder a explicar uno de estos métodos de información completa en la estimación de Sistemas de Ecuaciones Simultáneas, el método de Mínimos Cuadrados en tres Etapas (MC3E).

Supongamos que normalizamos las G ecuaciones cada una con respecto a una de las variables endógenas y consideramos el siguiente sistema de las G ecuaciones:

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad j = 1, \dots, G$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad j = 1, \dots, G$$

donde $Z_j = [Y_j \;\; X_j]$, $\quad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$, $\quad \mathbf{u}_j \sim (\mathbf{0}, \sigma_j^2 I_T)$ y $E(\mathbf{u}_i\mathbf{u}_j') = \sigma_{ij} I_T$.

Matricialmente,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{bmatrix} = \begin{bmatrix} Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & Z_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_G \end{bmatrix} \qquad \Leftrightarrow \underbrace{\mathbf{y}}_{GTx1} = \underbrace{Z}_{GTxJ} \underbrace{\delta}_{Jx1} + \underbrace{\mathbf{u}}_{GTx1}$$

donde $J = \sum_{j=1}^{G}(G_j + K_j)$ y $\mathbf{u} \sim (\mathbf{0}, (\Sigma \bigotimes I_T))$.

$$E(\mathbf{u}\mathbf{u}') = (\Sigma \bigotimes I_T) = \begin{bmatrix} \sigma_1^2 I_T & \sigma_{12} I_T & \cdots & \sigma_{1G} I_T \\ \sigma_{12} I_T & \sigma_2^2 I_T & \cdots & \sigma_{2G} I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{G1} I_T & \sigma_{G2} I_T & \cdots & \sigma_G^2 I_T \end{bmatrix}$$

Este es un sistema de ecuaciones aparentemente no relacionadas donde en la matriz Z hay variables endógenas que están correlacionadas con los términos de error. Por esta razón, en principio si aplicamos el método de MCG(F) directamente a este sistema, el estimador no sería consistente. Para tener en cuenta esto, premultiplicamos cada una de las ecuaciones por la matriz de variables exógenas X que son ortogonales a los errores estructurales y posteriormente, aplicamos MCG(F) al sistema de ecuaciones resultante:

$$X'\mathbf{y}_j = X'Z_j\delta_j + X'\mathbf{u}_j \quad j = 1, \ldots, G$$

$$\begin{bmatrix} X'\mathbf{y}_1 \\ X'\mathbf{y}_2 \\ \vdots \\ X'\mathbf{y}_G \end{bmatrix} = \begin{bmatrix} X'Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & X'Z_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & X'Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} X'\mathbf{u}_1 \\ X'\mathbf{u}_2 \\ \vdots \\ X'\mathbf{u}_G \end{bmatrix} \qquad \Leftrightarrow \underbrace{\mathbf{y}^*}_{KGx1} = \underbrace{Z^*}_{KGxJ} \underbrace{\delta}_{Jx1} + \underbrace{\mathbf{u}^*}_{KGx1}$$

teniendo en cuenta la estructura de varianzas y covarianzas que tiene $\mathbf{u}^*$. Si consideramos a X fija entonces $\mathbf{u}^* \sim (\mathbf{0}, \Sigma \bigotimes (X'X))$[1]. El estimador MCG, si se conociera Σ, vendría definido como

$$\hat{\delta} = \left(Z^{*\prime}(\Sigma^{-1} \bigotimes (X'X)^{-1})Z^*\right)^{-1} \left(Z^{*\prime}(\Sigma^{-1} \bigotimes (X'X)^{-1})\mathbf{y}^*\right)$$

Si no se conoce Σ al menos se debería de sustituir por un estimador consistente, $\hat{\Sigma}$. Notar que

$$Z^* = \begin{bmatrix} X'Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & X'Z_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & X'Z_G \end{bmatrix} = (I \bigotimes X')Z \qquad \mathbf{y}^* = \begin{bmatrix} X'\mathbf{y}_1 \\ X'\mathbf{y}_2 \\ \vdots \\ X'\mathbf{y}_G \end{bmatrix} = (I \bigotimes X')\mathbf{y}$$

Por lo tanto, podemos desarrollar los siguientes términos:

$$Z^{*\prime}(\Sigma^{-1} \bigotimes (X'X)^{-1}) = Z'(I \bigotimes X')' \left(\Sigma^{-1} \bigotimes (X'X)^{-1}\right) = Z'(I \bigotimes X) \left(\Sigma^{-1} \bigotimes (X'X)^{-1}\right) = Z' \left(\Sigma^{-1} \bigotimes X(X'X)^{-1}\right)$$

[1]Si consideramos X estocástica pero independiente de $\mathbf{u}_j$ $\forall j$ entonces $\mathbf{u}^* \sim (\mathbf{0}, \Sigma \bigotimes E(X'X))$.

Entonces,

$$Z^{*\prime}(\Sigma^{-1}\bigotimes(X'X)^{-1})Z^* = Z'\left(\Sigma^{-1}\bigotimes X(X'X)^{-1}\right)(I\bigotimes X')Z = Z'\left(\Sigma^{-1}\bigotimes X(X'X)^{-1}X'\right)Z = Z'\left(\Sigma^{-1}\bigotimes P_X\right)Z = W'Z$$

Por otro lado,

$$Z^{*\prime}(\Sigma^{-1}\bigotimes(X'X)^{-1})\mathbf{y}^* = Z'\left(\Sigma^{-1}\bigotimes X(X'X)^{-1}\right)(I\bigotimes X')\mathbf{y} = Z'\left(\Sigma^{-1}\bigotimes X(X'X)^{-1}X'\right)\mathbf{y} = Z'\left(\Sigma^{-1}\bigotimes P_X\right)\mathbf{y} = W'\mathbf{y}$$

Sustituyendo en el estimador obtenemos la expresión del estimador de MC3E, utilizando un estimador consistente de Σ:

$$\hat{\delta}_{MC3E} = \left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)Z\right)^{-1}\left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)\mathbf{y}\right) = (\hat{W}'Z)^{-1}\hat{W}'\mathbf{y}$$

De esta forma, el estimador MC3E se puede interpretar como un estimador de VI donde la matriz de instrumentos es $\hat{W}$.

Si previamente se ha estimado cada ecuación de la Forma estructural por MC2E, los elementos de Σ se pueden estimar consistentemente con

$$\hat{\sigma}_{ij} = \frac{1}{T}\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j \qquad i = 1,\ldots,G \quad j = 1,\ldots,G \quad \texttt{donde} \quad \hat{\mathbf{u}}_i = \mathbf{y}_i - Z_i\hat{\delta}_i^{MC2E}$$

donde las estimaciones de las varianzas corresponden a $i = j$.

Los estimadores de MC2E ecuación por ecuación y MC3E coinciden en los siguientes casos:

1) Si $\sigma_{ij} = 0 \quad \forall i \neq j$. Como ocurre en un SURE, si no existe correlación contemporánea entre las perturbaciones de las distintas ecuaciones de la forma estructural (Σ diagonal), entonces no se gana en eficiencia de estimar el sistema conjunto por MCG(F) que estimar separadamente ecuación por ecuación.

   En este caso tenemos que:

$$Z'\left(\Sigma^{-1}\bigotimes P_X\right)Z = \begin{bmatrix} Z_1' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & Z_2' & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & Z_G' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2}P_X & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_2^2}P_X & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_G^2}P_X \end{bmatrix} \begin{bmatrix} Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & Z_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & Z_G \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma_1^2}Z_1'P_XZ_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_2^2}Z_2'P_XZ_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_G^2}Z_G'P_XZ_G \end{bmatrix}$$

$$\left(Z'\left(\Sigma^{-1}\bigotimes P_X\right)Z\right)^{-1}=\begin{bmatrix}\sigma_1^2(Z_1'P_XZ_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \sigma_2^2(Z_2'P_XZ_2)^{-1} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_G^2(Z_G'P_XZ_G)^{-1}\end{bmatrix}$$

$$Z'\left(\Sigma^{-1}\bigotimes P_X\right)\mathbf{y}=\begin{bmatrix}Z_1' & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & Z_2' & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & Z_G'\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1^2}P_X & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma_2^2}P_X & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \frac{1}{\sigma_G^2}P_X\end{bmatrix}\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \vdots\\ \mathbf{y}_G\end{bmatrix}$$

$$=\begin{bmatrix}\frac{1}{\sigma_1^2}Z_1'P_X\mathbf{y}_1\\ \frac{1}{\sigma_2^2}Z_2'P_X\mathbf{y}_2\\ \vdots\\ \frac{1}{\sigma_G^2}Z_G'P_X\mathbf{y}_G\end{bmatrix}$$

Por lo tanto

$$\hat{\delta}_{MC3E}=\left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)Z\right)^{-1}\left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)\mathbf{y}\right)=$$

$$=\begin{bmatrix}\sigma_1^2(Z_1'P_XZ_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \sigma_2^2(Z_2'P_XZ_2)^{-1} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_G^2(Z_G'P_XZ_G')^{-1}\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma_1^2}Z_1'P_X\mathbf{y}_1\\ \frac{1}{\sigma_2^2}Z_2'P_X\mathbf{y}_2\\ \vdots\\ \frac{1}{\sigma_G^2}Z_G'P_X\mathbf{y}_G\end{bmatrix}$$

$$=\begin{bmatrix}\sigma_1^2\frac{1}{\sigma_1^2}(Z_1'P_XZ_1)^{-1}Z_1'P_X\mathbf{y}_1\\ \sigma_2^2\frac{1}{\sigma_2^2}(Z_2'P_XZ_2)^{-1}Z_2'P_X\mathbf{y}_2\\ \vdots\\ \sigma_G^2\frac{1}{\sigma_G^2}Z_G'P_XZ_G')^{-1}Z_G'P_X\mathbf{y}_G\end{bmatrix}=\begin{bmatrix}(Z_1'P_XZ_1)^{-1}Z_1'P_X\mathbf{y}_1\\ (Z_2'P_XZ_2)^{-1}Z_2'P_X\mathbf{y}_2\\ \vdots\\ Z_G'P_XZ_G')^{-1}Z_G'P_X\mathbf{y}_G\end{bmatrix}=\begin{bmatrix}\hat{\delta}_1^{MC2E}\\ \hat{\delta}_2^{MC2E}\\ \vdots\\ \hat{\delta}_G^{MC2E}\end{bmatrix}$$

2) Si **todas** las ecuaciones del sistema están exactamente identificadas.

   En este caso la condición de orden en cada ecuación se satisface con igualdad, esto es, $G_i+K_i=K \quad i=1,\ldots,G$. Por lo tanto las matrices $X'Z_i$ son todas ellas cuadradas (KxK).

   Dado que $(X'X)^{-1}$ es una matriz simétrica y definida positiva, existe una matriz L de orden (KxK) tal que $L'L=(X'X)^{-1}\Rightarrow(X'X)=L^{-1}(L')^{-1}\Rightarrow L(X'X)L'=I_K$.

Premultiplicando por L,

$$LX'\mathbf{y}_j = LX'Z_j\delta_j + LX'\mathbf{u}_j \quad j = 1, \dots, G$$

donde $LX'\mathbf{u}_j \sim (0, \sigma_j^2 I_K)$ y $E(LX'\mathbf{u}_i\mathbf{u}_j XL') = \sigma_{ij} I_K$.

Si denotamos con $R_j = (LX'Z_j)$, $\mathbf{r}_j = LX'\mathbf{y}_j$ y $\eta_j = LX'\mathbf{u}_j$ podemos escribir una ecuación de este sistema transformado como $\mathbf{r}_j = R_j\delta_j + \eta_j$ y al sistema

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_G \end{bmatrix} = \begin{bmatrix} R_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & R_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & R_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_G \end{bmatrix} \quad \Longleftrightarrow \quad \mathbf{r} = R\delta + \eta$$

donde $\eta \sim (\mathbf{0},\ \Sigma \bigotimes I_K)$

Por lo tanto, el estimador de δ_j para cada ecuación $j = 1, \dots, G$ por MC2E será el resultante de aplicar MCO a la ecuación j-ésima de este sistema transformado,

$$\hat{\delta}_j^{MC2E} = (R_j'R_j)^{-1}R_j'\mathbf{r}_j$$

y el estimador de δ por MC2E lo podemos escribir como

$$\hat{\delta}_{MC2E} = \begin{bmatrix} \hat{\delta}_1^{MC2E} \\ \hat{\delta}_2^{MC2E} \\ \vdots \\ \hat{\delta}_G^{MC2E} \end{bmatrix} = \begin{bmatrix} (R_1'R_1)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (R_2'R_2)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (R_G'R_G)^{-1} \end{bmatrix} \begin{bmatrix} R_1'\mathbf{r}_1 \\ R_2'\mathbf{r}_2 \\ \vdots \\ R_G'\mathbf{r}_G \end{bmatrix} = (R'R)^{-1}R'\mathbf{r}$$

El estimador de δ por MC3E será el resultante de aplicar MCG al sistema transformado anterior teniendo en cuenta $\eta \sim (\mathbf{0},\ \Sigma \bigotimes I_K)$, esto es

$$\hat{\delta}_{MC3E} = (R'\ (\Sigma \bigotimes I_K)^{-1}R)^{-1}R'(\Sigma \bigotimes I_K)^{-1}\mathbf{r}$$

Si todas las ecuaciones del sistema anterior están exactamente identificadas entonces,

$X'Z_j$ es (KxK) $j = 1, \cdots, G \ \Rightarrow\ R_j$ es (KxK) $\Rightarrow R$ es una matriz cuadrada (GKxGK). Entonces tenemos que

$\hat{\delta}_{MC2E} = (R'R)^{-1}R'\mathbf{r} = R^{-1}(R')^{-1}R'\mathbf{r} = R^{-1}\mathbf{r}$

y

$\hat{\delta}_{MC3E} = (R'\ (\Sigma \bigotimes I_K)^{-1}R)^{-1}R'(\Sigma \bigotimes I_K)^{-1}\mathbf{r} = R^{-1}(\Sigma \bigotimes I_K)(R')^{-1}R'(\Sigma \bigotimes I_K)^{-1}\mathbf{r} = R^{-1}\mathbf{r}$

Como queríamos demostrar.

4.1.1. Propiedades del estimador MC3E

Las propiedades para muestras finitas de este estimador no son conocidas y en general también puede presentar sesgos. Las propiedades asintóticas o para muestras grandes se pueden analizar considerando que este estimador es un estimador de VI. Para analizar la consistencia del estimador,

$$plim\,\hat{\delta}_{MC3E} = \delta + plim\left(\frac{1}{T}\hat{W}'Z\right)^{-1} plim\,\frac{1}{T}\hat{W}'\mathbf{u}$$

se debe satisfacer que

$$plim\frac{1}{T}\,\hat{W}'\mathbf{u} = plim\,\frac{1}{T}\underbrace{\begin{bmatrix} \hat{\sigma}^{11}Z_1'P_X & \hat{\sigma}^{12}Z_1'P_X & \cdots & \sigma^{1G}Z_1'P_X \\ \hat{\sigma}^{21}Z_2'P_X & \hat{\sigma}^{22}Z_2'P_X & \cdots & \hat{\sigma}^{2G}Z_2'P_X \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}^{G1}Z_G'P_X & \hat{\sigma}^{G2}Z_G'P_X & \cdots & \hat{\sigma}^{GG}Z_G'P_X \end{bmatrix}}_{Z'\left(\widehat{\Sigma}^{-1}\otimes P_X\right)}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \vdots\\ \mathbf{u}_G\end{bmatrix} =$$

$$= \begin{bmatrix} plim\frac{1}{T}\sum_{j=1}^{G}\hat{\sigma}^{1j}Z_1'P_X\mathbf{u}_j \\ plim\frac{1}{T}\sum_{j=1}^{G}\hat{\sigma}^{2j}Z_2'P_X\mathbf{u}_j \\ \vdots \\ plim\frac{1}{T}\sum_{j=1}^{G}\hat{\sigma}^{Gj}Z_G'P_X\mathbf{u}_j \end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix} \quad\Leftrightarrow\quad plim\,\frac{1}{T}\sum_{j=1}^{G}\hat{\sigma}^{ij}Z_i'P_X\mathbf{u}_j = \mathbf{0} \quad \forall i = 1,\ldots,G$$

Esto implica instrumentar y especificar correctamente todas y cada una de las ecuaciones.

Además $plim\left(\frac{1}{T}\hat{W}'Z\right)$ tiene que ser una matriz finita y definida positiva, esta matriz es[2]

$$\begin{bmatrix} plim\,\hat{\sigma}^{11}plim\frac{1}{T}Z_1'P_XZ_1 & plim\,\hat{\sigma}^{12}plim\frac{1}{T}Z_1'P_XZ_2 & \cdots & plim\,\hat{\sigma}^{1G}plim\frac{1}{T}Z_1'P_XZ_G \\ plim\hat{\sigma}^{21}plim\frac{1}{T}Z_2'P_XZ_1 & plim\,\hat{\sigma}^{22}plim\frac{1}{T}Z_2'P_XZ_2 & \cdots & plim\,\hat{\sigma}^{2G}plim\frac{1}{T}Z_2'P_XZ_G \\ \vdots & \vdots & \ddots & \vdots \\ plim\hat{\sigma}^{G1}plim\,\frac{1}{T}Z_G'P_XZ_1 & plim\,\hat{\sigma}^{G2}plim\,\frac{1}{T}Z_G'P_XZ_2 & \cdots & plim\,\hat{\sigma}^{GG}plim\,\frac{1}{T}Z_G'P_XZ_G \end{bmatrix}$$

Notar que

$$\hat{\delta}_{MC3E} = \left(Z'\left(\widehat{\Sigma}^{-1}\otimes P_X\right)Z\right)^{-1}\left(Z'\left(\widehat{\Sigma}^{-1}\otimes P_X\right)\mathbf{y}\right) =$$

$$= \left(Z'\left(I\otimes P_X\right)\left(\widehat{\Sigma}^{-1}\otimes I\right)Z\right)^{-1}\left(Z'\left(I\otimes P_X\right)\left(\widehat{\Sigma}^{-1}\otimes I\right)\mathbf{y}\right) =$$

$$= \left(\underbrace{Z'\left(I\bigotimes P_X\right)}_{\hat{Z}'}\left(\widehat{\Sigma}^{-1}\otimes I\right)\underbrace{\left(I\bigotimes P_X\right)Z}_{\hat{Z}}\right)^{-1}\left(\underbrace{Z'\left(I\bigotimes P_X\right)}_{\hat{Z}}\left(\widehat{\Sigma}^{-1}\otimes I\right)\mathbf{y}\right)$$

De esta forma el estimador MC3E se puede ver como un estimador de MCGF en el sistema de ecuaciones $\mathbf{y} = \hat{Z}\delta + \mathbf{w}$.

[2]La identificación de todas las ecuaciones por la condición de rango es suficiente.Una prueba formal se da en Schmidt (1976, pp.205-207)

Bajo las condiciones de identificación de todas las ecuaciones del sistema y suponiendo que las perturbaciones del sistema satisfacen $\mathbf{u} \sim (\mathbf{0}, (\Sigma \bigotimes I_T))$, la distribución asintótica del estimador MC3E es

$$\sqrt{T}\left(\hat{\delta}^{MC3E} - \delta\right) \xrightarrow{d} N\left(\mathbf{0},\ plim\left(\frac{1}{T}\hat{Z}'\left(\widehat{\Sigma}^{-1} \bigotimes I\right)\hat{Z}\right)^{-1}\right)$$

Un estimador consistente de la matriz de varianzas y covarianzas asintótica es

$$\widehat{Va}(\sqrt{T}\hat{\delta}^{MC3E}) = T\left(\hat{Z}'\left(\widehat{\Sigma}^{-1} \bigotimes I\right)\hat{Z}\right)^{-1}$$

En cuestión de eficiencia asintótica, se puede demostrar que, entre todos los estimadores de VI que sólo utilizan la información muestral incorporada en el sistema, MC3E es el más eficiente asintóticamente. Para perturbaciones normalmente distribuidas, puede mostrarse también que MC3E tiene la misma distribución asintótica que el estimador de máxima verosimilitud con información completa, el cual es asintóticamente eficiente entre todos los estimadores.

4.1.2. Inferencia

Utilizando los resultados anteriores sobre distribución asintótica y estimación de las matrices de covarianzas asintóticas de estos estimadores, se puede derivar el estadístico de Wald para contrastar restricciones lineales sobre δ. En general, podemos escribir la hipótesis nula y la alternativa como:

$$H_0: \underset{(q \times \sum_{j=1}^{G}(G_j+K_j))}{R} \cdot \underset{(\sum_{j=1}^{G}(G_j+K_j)\times 1)}{\delta} = \underset{(q\times 1)}{r}$$

$$H_A: R\delta \neq r$$

El estadístico de Wald de contraste y su distribución asintótica bajo H_0 son:

$$T(R\,\hat{\delta}_{MC3E} - r)'\,[R\,\widehat{Va}(\sqrt{T}\hat{\delta}_{MC3E})\,R']^{-1}\,(R\,\hat{\delta}_{MC3E} - r) \xrightarrow{d,H_0} \chi^2_q$$

4.2. Máxima Verosimilitud Información Completa

El método de estimación por Máxima Verosimilitud Información Completa (MVIC), considera la estimación de los parámetros estructurales de todo el sistema conjuntamente. Se considera la función de verosimilitud de todas las variables endógenas del sistema dadas las exógenas, teniendo en cuenta todas las restricciones que la sobreidentificación de alguna o algunas de las ecuaciones estructurales impone sobre los parámetros de la Forma Reducida. Esto es lo que se considera como Información Completa.

Las ecuaciones de la Forma Reducida para todas las variables endógenas del sistema Y son las siguientes

$$\underbrace{Y}_{TxG} = \underbrace{X}_{TxK}\underbrace{\Pi'}_{KxG} + \underbrace{V}_{TxG}$$

Denotemos por $\mathbf{v}_t'$ al vector $(1xG)$ correspondiente a la fila t-ésima de la matriz V. Dado el supuesto sobre la distribución de los errores estructurales $\mathbf{u}_t \sim NID(\mathbf{0}, \Sigma)$, el vector de errores de la forma reducida correspondiente a las ecuaciones de las variables Y en el momento t, se distribuye $\mathbf{v}_t \sim NID(\mathbf{0}, \Omega)$ donde $\Omega = (B^{-1})'\Sigma B^{-1}$.

Si denotamos por $\mathbf{y}_t'$ y $\mathbf{x}_t'$ a la fila t-ésima de Y y de X respectivamente

$$\mathbf{y}_t' = \mathbf{x}_t'\Pi' + \mathbf{v}_t' \qquad \Longrightarrow \qquad \mathbf{y}_t = \Pi\mathbf{x}_t + \mathbf{v}_t \qquad \Longleftrightarrow \qquad \mathbf{y}_t - \Pi\mathbf{x}_t = \mathbf{v}_t$$

La función de verosimilitud de Y es:

$$\begin{aligned}
L &= L(\Pi, \Omega; Y, X) = \textstyle\prod_{t=1}^{T} L(\Pi, \Omega; \mathbf{y}_t, \mathbf{x}_t) = \\
&= \prod_{t=1}^{T} \left((2\pi)^{-G/2} |\Omega|^{-1/2} \exp\left[(-1/2)\mathbf{v}_t'\Omega^{-1}\mathbf{v}_t\right] \right) = \\
&= \prod_{t=1}^{T} \left((2\pi)^{-G/2} |\Omega|^{-1/2} \exp\left[(-1/2)(\mathbf{y}_t - \Pi\mathbf{x}_t)'\Omega^{-1}(\mathbf{y}_t - \Pi\mathbf{x}_t)\right] \right)
\end{aligned}$$

Tomando logaritmos, obtenemos

$$\begin{aligned}
lnL &= -((TG/2)\, ln\, 2\pi) - (T/2)\, ln\, |\Omega| \, - 1/2 \left[\sum_{t=1}^{T} (\mathbf{y}_t - \Pi\mathbf{x}_t)'\Omega^{-1}(\mathbf{y}_t - \Pi\mathbf{x}_t) \right] = \\
&= -((TG/2)\, ln\, 2\pi) - (T/2)\, ln\, |\Omega| \, - 1/2 \left[\sum_{t=1}^{T} \mathbf{v}_t'\Omega^{-1}\mathbf{v}_t \right] = \\
&= -(T/2) \left[G\, ln\, 2\pi + \, ln\, |\Omega| \, + \, tr\left(\Omega^{-1}\, W\right) \right]
\end{aligned}$$

donde[3] $W = \frac{1}{T}\sum_{t=1}^{T} \mathbf{v}_t\, \mathbf{v}_t' = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{y}_t - \Pi\mathbf{x}_t)(\mathbf{y}_t - \Pi\mathbf{x}_t)'$.

Sustituyendo

$\Pi = -B^{-1}\Gamma$ y $\Omega = B^{-1}\Sigma(B^{-1})'$ por lo que $\Omega^{-1} = B'\Sigma^{-1}B$ obtenemos

$$lnL(B,\, \Gamma,\, \Sigma; Y, X) = -(1/2)\, TG\, ln\, (2\pi) - 1/2\, T\, ln\, |B^{-1}\Sigma(B^{-1})'| - (1/2)\, T\, tr\left(B'\Sigma^{-1}B\, W\right)$$

donde $W = \frac{1}{T}\sum_{t=1}^{T} (\mathbf{y}_t + B^{-1}\Gamma\mathbf{x}_t)(\mathbf{y}_t + B^{-1}\Gamma\mathbf{x}_t)'$.

- Utilizando propiedades de los determinantes,

$$ln\, |B^{-1}\Sigma(B^{-1})'| = ln\left(|B^{-1}||\Sigma||(B^{-1})'|\right) = ln(\tfrac{1}{|B|^2}\, |\Sigma|) = -2\, ln\, |B| + ln\, |\Sigma|$$

[3]Ver apéndice.

- Utilizando propiedades de la traza,

$$tr\left(B'\Sigma^{-1}B\,W\right) = tr\left(\Sigma^{-1}B\,WB'\right) = \frac{1}{T}tr\left(\Sigma^{-1}B\,\sum_{t=1}^{T}(\mathbf{y}_t + B^{-1}\Gamma\mathbf{x}_t)(\mathbf{y}_t + B^{-1}\Gamma\mathbf{x}_t)'B'\right)$$
$$= \frac{1}{T}tr\left(\Sigma^{-1}\,\sum_{t=1}^{T}(B\mathbf{y}_t + \Gamma\mathbf{x}_t)(B\mathbf{y}_t + \Gamma\mathbf{x}_t)'\right) = \frac{1}{T}tr\left(\Sigma^{-1}\,(BY + \Gamma X)'(BY + \Gamma X)\right)$$

Sustituyendo se obtiene

$$lnL(B,\,\Gamma,\,\Sigma;Y,X) = -(1/2)\,TG\,ln\,(2\pi) - 1/2Tln\,|\Sigma| + T\,ln\,|B| - (1/2)tr\left(\Sigma^{-1}\,(BY+\Gamma X)'(BY+\Gamma X)\right)$$

El estimador MVIC maximiza esta función sujeto a todas las restricciones sobre B, Γ y sobre la matriz de varianzas y covarianzas Σ si las hubiera. Esto se puede llevar a cabo calculando las derivadas parciales de lnL con respecto a los parámetros que quedan sin restringir e igualar estos gradientes a cero. En general, salvo algunas excepciones, la resolución analítica de este sistema para todos los parámetros no es factible ya que es un problema altamente no lineal, y por lo tanto se requiere de la utilización de métodos de optimización numérica.

Dos excepciones son las siguientes:

1. El modelo recursivo. En ese caso B es triangular con unos en la diagonal principal y Σ diagonal. Se puede demostrar que el estimador MVIC en este caso coincide con el estimador MCO de cada ecuación estructural.

2. Si todas las ecuaciones del sistema están exactamente identificadas. En este caso, hay una correspondencia uno a uno entre (B, Γ) y Π. Al no haber restricciones sobre Π debidas a la sobreidentificación, MVIC se obtiene estimando Π por MCO y resolviendo para B, Γ de la relación $B\Pi + \Gamma = 0$. Podemos apelar simplemente al principio de invarianza de los estimadores máximo-verosímiles. En un modelo exactamente identificado bajo Normalidad de los errores, $MCI = MC2E = MVIL = MC3E = MVIC$

Si el supuesto de Normalidad[4] es cierto y todas las restricciones a priori impuestas de identificación son correctas, el estimador MVIC es consistente y asintóticamente eficiente.

La matriz asintótica de varianzas y covarianzas para el estimador de MVIC bajo Normalidad es idéntica a la del estimador MC3E. La implicación es que, con perturbaciones normalmente distribuidas, MC3E es eficiente asintóticamente dentro de los estimadores de Información Completa.

[4] Si el supuesto de Normalidad no es correcto, el estimador MVIC es consistente pero ya no será asintóticamente eficiente.

Apéndice

Denotemos por σ^{ij} el ij-ésimo elemento de Ω^{-1}. Entonces

$$\mathbf{v}_t'\Omega^{-1}\mathbf{v}_t = \begin{pmatrix} v_{1t} & v_{2t} & \cdots & v_{Gt} \end{pmatrix} \begin{pmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1G} \\ \sigma^{21} & \sigma^{22} & \cdots & \sigma^{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{G1} & \sigma^{G2} & \cdots & \sigma^{GG} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Gt} \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{i=1}^G v_{it}\sigma^{i1} & \sum_{i=1}^G v_{it}\sigma^{i2} & \cdots & \sum_{i=1}^G v_{it}\sigma^{iG} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Gt} \end{pmatrix} = \sum_{i=1}^G \sum_{j=1}^G v_{it}\, v_{jt}\, \sigma^{ij}$$

Luego,

$$\sum_{t=1}^T \mathbf{v}_t'\Omega^{-1}\mathbf{v}_t = \sum_{t=1}^T \sum_{i=1}^G \sum_{j=1}^G v_{it}\, v_{jt}\, \sigma^{ij}$$

Ahora, invertimos el orden de la suma, sumando primero en t y luego sobre $i, j = 1, \ldots, G$

$$\sum_{t=1}^T \mathbf{v}_t'\Omega^{-1}\mathbf{v}_t = \sum_{i=1}^G \sum_{j=1}^G \sigma^{ij} \sum_{t=1}^T v_{jt}\, v_{it} = T \sum_{i=1}^G \sum_{j=1}^G \sigma^{ij} W_{ij}$$

donde denotamos por $W_{ij} = \frac{1}{T}\sum_{t=1}^T v_{it}\, v_{jt}$ al elemento ij-ésimo de la matriz (GxG),

$W = \frac{1}{T}\sum_{t=1}^T \mathbf{v}_t\, \mathbf{v}_t'$ siendo $v_{it} = y_{it} - \pi_{i1}x_{1t} - \pi_{i2}x_{2t} - \cdots - \pi_{iK}x_{Kt} = y_{it} - \pi_i'\mathbf{x}_t$.
El elemento diagonal i-ésimo de la matriz $\Omega^{-1}\, W$ es $\sum_j \sigma^{ij} W_{ij}$, por lo tanto, la suma anterior es T veces la traza de esa matriz, esto es

$$\sum_{t=1}^T \mathbf{v}_t'\Omega^{-1}\mathbf{v}_t = T\, tr(\Omega^{-1}\, W)$$

Capítulo 5

Contrastes de Especificación

5.1. Introducción

Partimos de la especificación de un modelo estructural lineal

$$\underbrace{B}_{GxG}\underbrace{\mathbf{y}_t}_{Gx1}+\underbrace{\Gamma}_{GxK}\underbrace{\mathbf{x}_t}_{Kx1}=\underbrace{\mathbf{u}_t}_{Gx1} \qquad t=1,\ldots,T$$

Si denotamos

$$\underbrace{Y}_{TxG}=\begin{bmatrix}\mathbf{y}_1'\\ \mathbf{y}_2'\\ \vdots\\ \mathbf{y}_T'\end{bmatrix} \quad \underbrace{X}_{TxK}=\begin{bmatrix}\mathbf{x}_1'\\ \mathbf{x}_2'\\ \vdots\\ \mathbf{x}_T'\end{bmatrix} \quad \underbrace{U}_{TxG}=\begin{bmatrix}\mathbf{u}_1'\\ \mathbf{u}_2'\\ \vdots\\ \mathbf{u}_T'\end{bmatrix}$$

podemos escribir la Forma Estructural como

$$YB'+X\Gamma'=U \tag{5.1}$$

donde Y es la matriz (TxG) de observaciones de las variables endógenas, X es la matriz (TxK) de observaciones de las variables predeterminadas y/o exógenas, U es la matriz (TxG) de perturbaciones estructurales, B' es la matriz (GxG) de coeficientes estructurales que acompañan a las variables endógenas y Γ' la matriz (KxG) de coeficientes estructurales que acompañan a las variables exógenas y/o predeterminadas[1].

En el modelo (5.1) las matrices de coeficientes B y Γ están sujetas a restricciones *a priori* las cuales permiten la identificación de la ecuación o ecuaciones a examinar. Por simplicidad, suponemos que estas restricciones son de exclusión, por lo que ciertos elementos de B y Γ serán ceros. Supondremos además que $\mathbf{u}_t \sim NID(0,\Sigma)$ donde Σ es no-singular[2].

[1]Se supone no singular y sujeta a la regla de normalización $\beta_{jj}=1 \quad j=1,\cdots,G$. En el caso de que B fuera una matriz diagonal, sería simplemente un sistema de ecuaciones aparentemente no relacionadas SURE, sin simultaneidad.

[2]Para muchos de los resultados basados en IV no será necesario el supuesto de Normalidad. Se considera que las identidades se han eliminado una vez han sido sustituidas en las ecuaciones y que la matriz Σ está sin restringir.

Si las variables con coeficientes igual a cero son excluidas, entonces cada una de las ecuaciones del sistema (5.1) se puede escribir como

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \tag{5.2}$$

o de forma más compacta

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Z_j}_{Tx(G_j+K_j)}\underbrace{\delta_j}_{(G_j+K_j)x1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim N(\mathbf{0}, \sigma_j^2 I_T)$$

donde $Z_j = \begin{bmatrix} \underbrace{Y_j}_{TxG_j} & \underbrace{X_j}_{TxK_j} \end{bmatrix} \qquad \delta_j = \begin{bmatrix} \beta_j \\ \gamma_j \end{bmatrix}$.

La condición necesaria para que los elementos de δ_j estén identificados en cada ecuación j es que el número de variables exógenas y/o predeterminadas en el sistema $K = K_j + K_j^*$ sea tal que $K \geq G_j + K_j$, o lo que es lo mismo, el número de variables endógenas incluidas en la derecha de la ecuación j-ésima G_j no exceda el número de variables exógenas y/o predeterminadas excluidas de esa ecuación K_j^*, esto es, $K_j^* \geq G_j$. Además se supone que la condición suficiente o de rango se satisface.

Este modelo de partida será el modelo que consideraremos bajo la hipótesis nula a contrastar. Una decisión importante que ha de hacerse previamente, es si se contrastará todo el sistema (5.1) en su conjunto, utilizando los resultados obtenidos con algún método de estimación de información completa, o si se va a considerar cada ecuación (5.2) separadamente, después de la aplicación de alguna técnica de información limitada. La primera aproximación puede requerir en ocasiones un importante coste computacional y tamaños muestrales relativamente grandes para que la teoría asintótica sea una aproximación adecuada. Más aún, si una o dos ecuaciones estructurales no están bien especificadas, esto puede llevar a rechazar la especificación de todo el sistema debido a que estas contaminen la adecuada estimación de otras ecuaciones. Ahora bien, esto último también puede utilizarse a la hora de desarrollar un contraste de especificación del sistema en su conjunto y utilizarse para, si se rechaza la adecuada especificación del sistema, considerar estimar cada ecuación por separado.

Vamos a considerar los siguientes contrastes de especificación, que están relacionados con la adecuación de los supuestos de ortogonalidad entre instrumentos y perturbaciones estructurales:

- **Contrastes de restricciones de sobreidentificación.** Si consideramos una ecuación del sistema y esta se considera sobreidentificada, por lo que $K_j^* - G_j > 0$, esto permite llevar a cabo un contraste de la validez de los instrumentos o de las condiciones de ortogonalidad de estos con las perturbaciones y a su vez de la adecuación de la especificación de la ecuación propuesta en concreto[3]. Podemos considerar que en la alternativa la ecuación estaría exactamente identificada ya que hubiéramos incluido ese exceso de instrumentos o

[3]Es decir las variables consideradas exógenas y/o predeterminadas que se han excluido de la ecuación K_j^* exceden las endógenas incluidas a la derecha de la ecuación G_j.

de variables exógenas $d_j = K_j^* - G_j$. Bajo la hipótesis nula, los coeficientes que acompañan a esas d_j variables se igualan a cero.

Si no hay sobreidentificación entonces esto no se puede llevar a cabo, ya que tenemos exactamente los instrumentos que necesitamos para poder identificar los parámetros de la ecuación estructural.

Este tipo de contraste también se puede realizar considerando el sistema en su conjunto si alguna o algunas de las ecuaciones está sobreidentificada. El número de restricciones a contrastar bajo la hipótesis nula de correcta especificación del sistema serían $\Sigma_{j=1}^{G} d_j$.

- **Contrastes de "exogeneidad".**

  En el modelo (5.1) se especifica en cada ecuación cuales son las variables endógenas incluidas y cuales son las exógenas o predeterminadas incluidas y excluidas que se suponen incorreladas con los términos de perturbación estructurales.

  Estos contrastes tratan de diagnosticar si existe evidencia de la existencia de variables consideradas como exógenas que puedan no serlo[4].

Otros contrastes de especificación relevantes son sobre los supuestos del comportamiento de las perturbaciones (heterocedasticidad, autocorrelación, normalidad) o sobre la estabilidad de los coeficientes estrucutrales aunque no se tratarán en estas notas, refiriendo al lector al libro de Godfrey (1988).

5.2. Contrastes de restricciones de sobreidentificación

- **Contrastes en una ecuación o basados en información limitada**:

  a) Contraste de razón de verosimilitudes (LR) de Anderson y Rubin[5].

$$LR_j = T(\hat{\lambda}_j - 1) \overset{d,H_0}{\longrightarrow} \chi^2_{(K_j^* - G_j)} \tag{5.3}$$

donde $\hat{\lambda}_j$ es el menor valor propio asociado a la estimación LIML de la ecuación j-ésima,

$$\hat{\lambda}_j = \frac{\hat{\beta}^{+\prime}_{j,MVIL} S_j^0 \hat{\beta}^{+}_{j,MVIL}}{\hat{\beta}^{+\prime}_{j,MVIL} S_j^1 \hat{\beta}^{+}_{j,MVIL}} \tag{5.4}$$

donde $\hat{\beta}^{+\prime}_{j,MVIL} = (1, -\hat{\beta}^{\prime}_{j,MVIL})$, $\quad S_j^0 = Y_j^{+\prime} M_j Y_j^{+}$, $\quad M_j = I_T - X_j(X_j^{\prime} X_j)^{-1} X_j^{\prime}$,

$S_j^1 = Y_j^{+\prime} M Y_j^{+}$, y $\quad M = I_T - X(X^{\prime}X)^{-1}X^{\prime}$

[4]En el sentido de ortogonalidad entre esas variables y los términos de perturbación estructurales o, asintóticamente, $plim \frac{1}{T}(X^{\prime}U) = 0$.

[5]En Gretl al estimar por MVIL, el ouput del programa computa una variante de este estadístico, $T\, ln\, \hat{\lambda}_j$, que es una buena aproximación a LR_j para valores de $\hat{\lambda}_j$ cercanos a 1.

b) Estadístico de Sargan (1958).

$$TR_u^2 = T\,\frac{\hat{\mathbf{u}}_j' X(X'X)^{-1}X'\hat{\mathbf{u}}_j}{\hat{\mathbf{u}}_j'\hat{\mathbf{u}}_j}$$

donde $\hat{\mathbf{u}}_j = y_j - Z_j\hat{\delta}_j$ siendo $\hat{\delta}_j$ bien el estimador MC2E o el de MVIL. Por lo tanto el R_u^2 es el coeficiente de determinación no centrado de la regresión por MCO de $\hat{\mathbf{u}}_j$ sobre todas las variables exógenas o predeterminadas X[6]. Bajo la hipótesis nula de correcta especificación de la ecuación j-ésima (5.2) o de que todos los instrumentos X son ortogonales al error $\mathbf{u}_j$, el estadístico

$$TR_u^2 \xrightarrow{d,H_0} \chi^2_{(K_j^*-G_j)} \tag{5.5}$$

c) Estadístico de Basmann (1960).

$$F = \frac{\hat{\mathbf{u}}_j' X(X'X)^{-1}X'\hat{\mathbf{u}}_j/(K_j^*-G_j)}{(\hat{\mathbf{u}}_j'\hat{\mathbf{u}}_j - \hat{\mathbf{u}}_j' X(X'X)^{-1}X'\hat{\mathbf{u}}_j)/(T-K)} = \frac{\hat{\mathbf{u}}_j' P_X\hat{\mathbf{u}}_j/(K_j^*-G_j)}{\hat{\mathbf{u}}_j'\, M_X\,\hat{\mathbf{u}}_j/(T-K)} = \frac{T-K}{K_j^*-G_j}(\hat{\lambda}_j'-1) \tag{5.6}$$

donde $\hat{\lambda}_j'$ se obtiene como en (5.4) pero utilizando $\hat{\beta}^{+\prime}_{j,MC2E} = (1, -\hat{\beta}'_{j,MC2E})$. Se puede utilizar la misma regresión auxiliar de los residuos MC2E sobre X para obtener ambos estadísticos, Sargan y Basmann. La diferencia entre los dos estadísticos está en el denominador y en la corrección por grados de libertad. El estadístico de Basmann se distribuye bajo la hipótesis nula de correcta especificación de la ecuación j-ésima como una $F[(K_j^*-G_j),(T-K)]$.

- **Contrastes del sistema en conjunto basados en Información Completa**:

a) Contraste de Razón de Verosimilitudes (LR).

El contraste se construye comparando el valor del logaritmo de la función de verosimilitud una vez estimado el modelo no restringido sin imponer las restricciones de sobreidentificación (modelo exactamente identificado) con el modelo restringido (se imponen restricciones no lineales en los parámetros de la forma reducida). En el primer caso no se impone ninguna restricción sobre los coeficientes de la forma reducida, $\hat{\Pi}_U = \hat{\Pi}_{MCO}$ y en el segundo $\hat{\Pi}_R = -\hat{B}^{-1}\hat{\Gamma}$ donde $\hat{B}$ y $\hat{\Gamma}$ son estimadores MVIC de los parámetros estructurales bajo la hipótesis nula de sobreidentificación[7].

$$LR = T\,(ln \mid W_R \mid - ln \mid W_U \mid) \xrightarrow{d,H_0} \chi^2_{(\Sigma_{j=1}^{G}(K_j^*-G_j))} \tag{5.7}$$

donde $W = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{v}}_t\,\hat{\mathbf{v}}_t'$ siendo $\hat{v}_{it} = y_{it} - \hat{\pi}_{i1}x_{1t} - \hat{\pi}_{i2}x_{2t} - \cdots - \hat{\pi}_{iK}x_{Kt} = y_{it} - \hat{\pi}_i'\mathbf{x}_t$ utilizando el estimador de $\pi_i, i = 1, \cdots, G$ en cada caso.

[6]En el denominador aparece $\hat{\mathbf{u}}_j'\hat{\mathbf{u}}_j$ sin centrar en la media de los residuos del ajuste por MC2E o MVIL que no tiene porqué ser igual a cero.

[7]También pueden ser utilizados los de MC3E.

b) Contraste de Hansen-Sargan.

En este caso se contrasta la especificación de todo el sistema en su conjunto, o todas las restricciones de sobreidentificación. Para ello se utiliza un estimador de información completa, bien 3SLS o bien MVIC.

Dado el sistema

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_G \end{bmatrix} = \begin{bmatrix} Z_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & Z_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & Z_G \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_G \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_G \end{bmatrix} \Leftrightarrow \underbrace{\mathbf{y}}_{GTx1} = \underbrace{Z}_{GTxJ} \underbrace{\delta}_{Jx1} + \underbrace{\mathbf{u}}_{GTx1} \tag{5.8}$$

donde $J = \sum_{j=1}^{G}(G_j + K_j)$ y $\mathbf{u} \sim (\mathbf{0}, (\Sigma \bigotimes I_T))$ y todas las ecuaciones están identificadas tal que $\Sigma_{j=1}^{G}(K_j^* - G_j) > 0$.

Una vez obtenidos el estimador de δ por MC3E se evalúa la función que minimiza este estimador, utilizando un estimador consistente de Σ, por ejemplo, el basado en los residuos de estimar cada ecuación por MC2E.

Esta variable aleatoria, resultante de evaluar el criterio que minimiza $\hat{\delta}_{MC3E}$, es

$$(\mathbf{y} - Z\,\hat{\delta}_{MC3E})'(\hat{\Sigma}^{-1} \bigotimes X(X'X)^{-1}X')(\mathbf{y} - Z\,\hat{\delta}_{MC3E}) = \hat{\mathbf{u}}'(\hat{\Sigma}^{-1} \bigotimes P_X)\hat{\mathbf{u}} \tag{5.9}$$

que bajo la hipótesis nula de correcta especificación del sistema (5.8) sigue una distribución asintótica $\chi^2_{(\Sigma_{j=1}^{G}(K_j^* - G_j))}$. Este contraste es asintóticamente equivalente al anterior LR test[8].

5.3. Contrastes de exogeneidad

Hausman (1978) en este artículo se preocupa de la forma de contrastar una incorrecta especificación del modelo tal que no se satisfaga el supuesto de ortogonalidad entre regresores y el término de perturbación. En el contexto de un modelo de regresión estandar esto se puede deber a errores de medida en las variables explicativas, omisión de variables o factores no observables correlacionados con las variables incluidas o endogeneidad de ciertos regresores que se determinan simultáneamente con la variable dependiente.

Básicamente, la idea que tiene Hausman para derivar un estadístico de contraste es la de considerar la diferencia entre dos estimadores alternativos del mismo conjunto de parámetros, tal que bajo la hipótesis nula de incorrelación entre regresores y error, los dos estimadores a comparar fueran consistentes y uno de ellos asintóticamente eficiente[9]. Bajo la hipótesis alternativa de no-ortogonalidad entre regresores y error, el que era asintóticamente eficiente bajo la nula se vuelve inconsistente siendo aún consistente el estimador alternativo.

[8] Bajo la hipótesis alternativa, todas las ecuaciones estrían exactamente identificadas. En ese caso la función criterio tomaría el valor cero evaluada en $\hat{\delta}_{MC3E} = \hat{\delta}_{MC2E}$.

[9] Esto último es importante para que la matriz de varianzas y covarianzas asintótica de la diferencia de los dos estimadores sea simplemente la diferencia de las respectivas matrices de varianzas y covarianzas.

En el contexto de una ecuación,

$$\underbrace{\mathbf{y}_j}_{Tx1} = \underbrace{Y_j}_{TxG_j}\underbrace{\beta_j}_{G_jx1} + \underbrace{X_j}_{TxK_j}\underbrace{\gamma_j}_{K_jx1} + \underbrace{\mathbf{u}_j}_{Tx1} \qquad \mathbf{u}_j \sim N(\mathbf{0}, \sigma_j^2 I_T) \tag{5.10}$$

Se puede querer contrastar si las variables en Y_j son o no endógenas. En este caso bajo la hipótesis nula todas las variables explicativas estarían incorrelacionadas con $\mathbf{u}_j$ y bajo la alternativa las variables en Y_j serían endógenas y estarían correlacionadas con el término de error. Las variables en X_j se las considera en ambos casos ortogonales al término de perturbación.

Los dos estimadores a comparar en el contexto de información limitada serían el de MCO (bajo H_0 consistente y asintóticamente eficiente y bajo H_a inconsistente) y el de $MC2E$ (bajo H_0 y H_a consistente) donde se utilizan todos los posibles instrumentos $X = (X_j, X_j^*)$ siendo $K = K_j + K_j^*$. Para que se pueda llevar a cabo el contraste, la ecuación tiene que estar bajo la hipótesis nula sobreidentificada, esto es $K_j^* > G_j = 0$[10]. Dado que bajo H_0 Y_j está incorrelacionada con $\mathbf{u}_j$, tenemos en exceso K_j^* condiciones de ortogonalidad o instrumentos ya que tenemos que estimar por MC2E y para ello necesitamos K_j^* variables X_j^* que no estén correlacionadas con $\mathbf{u}_j$ y no estén en la ecuación. Además tienen que estar correlacionadas con las variables Y_j. De esta forma se satisfacen las condiciones para que el estimador MC2E esté bien definido y sea consistente.

En este caso, el estadístico de Hausman sería

$$H = (\hat{\delta}_{j,MC2E} - \hat{\delta}_{j,MCO})' D^+ (\hat{\delta}_{j,MC2E} - \hat{\delta}_{j,MCO}) \xrightarrow{d,H_0} \chi^2_{(G_j)} \tag{5.11}$$

donde D^+ denota la inversa generalizada de la matriz[11]

$$D = \hat{\sigma}_j^2 \left((Z_j' P_X Z_j)^{-1} - (Z_j' Z_j)^{-1} \right)$$

donde $\hat{\sigma}_j^2$ es un estimador consistente de σ_j^2 y $Z_j = (Y_j \; X_j)$. Se puede utilizar:

- Usando los residuos de estimar por MCO

$$\hat{\sigma}^2_{j,MCO} = \frac{\hat{\mathbf{u}}'_{j,MCO}\hat{\mathbf{u}}_{j,MCO}}{T}$$

  Esta versión del estadístico fué también propuesta por Durbin(1954) y separadamente por Wu(1973) y Hausman (1978).

- Usando los residuos de estimar por VI o MC2E

$$\hat{\sigma}^2_{j,MC2E} = \frac{\hat{\mathbf{u}}'_{j,MC2E}\hat{\mathbf{u}}_{j,MC2E}}{T}$$

  Esta versión del estadístico fué también propuesta por Wu (1973) y Hausman (1978).

[10] Bajo la hipótesis alternativa estará exactamente identificada ya que necesitamos instrumentalizar a Y_j por lo que $K_j^* = G_j$. Con $G_j = 0$ se quiere decir que no hay variables endógenas en la derecha de la ecuación, no que omitamos a las variables Y_j bajo la hipótesis nula

[11] Si no aparecieran variables exógenas adicionales en la ecuación, esto es no apareciera X_j, la matriz D no sería singular ya que su rango sería completo. Pero si aparece X_j aunque simplemente sea el término constante, la matriz D es de rango reducido G_j, por lo que es necesario obtener la inversa generalizada.

Sea A una matriz (mxn). Cualquier matriz B que satisface ABA = A es una inversa generalizada de A. La matriz inversa generalizada no es única en general. Una excepción es la llamada de *Moore-Penrose*. Esta matriz B además satisface que $BAB = B$, $(AB)' = AB$ y $(BA)' = BA$.

Aunque ambas versiones del contraste son asintóticamente equivalentes, ambas diferirán numéricamente y pueden tener un comportamiento distinto en muestras finitas. Puestos a elegir entre ambas versiones, parece que es más recomendable la primera basada en $\hat{\sigma}^2_{j,MCO}$, porque el estimador con el que se calcula la suma de cuadrados residual bajo la hipótesis nula es además de consistente más eficiente. También parece ser superior cuando los instrumentos utilizados no están muy correlacionados con Y_j. Lo que sí parece mejor es utilizar un sólo estimador de σ^2_j, ya que si para computar D se utiliza $\hat{\sigma}^2_{j,MC2E}(Z_j'P_XZ_j)^{-1} - \hat{\sigma}^2_{j,MCO}(Z_j'Z_j)^{-1}$, no hay garantías de que la inversa generalizada de D exista y pueden obtenerse valores negativos del estadístico H.

Dado que el obtener el estadístico H de esta forma utilizando la inversa generalizada de D puede ser complicada, hay una forma más sencilla y equivalente de llevar a cabo el contraste.

Sea la forma reducida para Y_j

$$\underbrace{Y_j}_{TxG_j} = \underbrace{X}_{TxK}\underbrace{\Pi_j'}_{KxG_j} + \underbrace{V_j}_{TxG_j}$$

$$\hat{\Pi}'_{j\,MCO} = (X'X)^{-1}X'Y_j$$

$\hat{\Pi}'_{j\,MCO} = \begin{bmatrix}\hat{\pi}_1^{MCO} & \cdots & \hat{\pi}_{G_j}^{MCO}\end{bmatrix}$, donde las columnas de $\hat{\Pi}'_{j\,MCO}$ son respectivamente los coeficientes de la regresión de cada variable en Y_j sobre X.

Se obtienen los valores ajustados de Y_j dada la estimación de la forma reducida

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

y lo residuos

$$\hat{V}_j = Y_j - \hat{Y}_j$$

Ahora estimar por MCO la ecuación

$$\mathbf{y}_j = Y_j\beta_j + X_j\gamma_j + \hat{V}_j\alpha_j + \epsilon_j \tag{5.12}$$

y contrastar mediante un estadístico F usual la significatividad de $\hat{V}_j$ donde se contrastan bajo $H_0 : \alpha_j = 0$, G_j restricciones de exclusión.

Finalmente, Hausman (1978) considera la aplicación de su contraste como un contraste de especificación del sistema de ecuaciones simultáneas en su conjunto (5.8) comparando los estimadores MC3E y MC2E del conjunto de parámetros del sistema[12].

Se considera que bajo la hipótesis nula de correcta especificación de todas las ecuaciones del sistema el estimador de δ en (5.8) por MC3E es consistente y asintóticamente eficiente pero bajo la hipótesis alternativa de mala especificación no es consistente[13] . Por otro lado, el estimador de δ obtenido por $MC2E$ en cada una de las ecuaciones será consistente bajo la hipótesis nula pero no eficiente asintóticamente. Bajo la hipótesis alternativa, solamente se estimarán inconsistentemente por MC2E los parámetros de la ecuación o ecuaciones mal-especificadas.

[12]Notar que al menos alguna ecuación ha de estar sobreidentificada por que sino ambos estimadores son iguales y su diferencia sería cero.

[13]Esto ocurrirá si al menos una de las ecuaciones del sistema no está correctamente especificada.

El estimador MC3E de δ en (5.8) es

$$\hat{\delta}_{MC3E} = \left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)Z\right)^{-1}\left(Z'\left(\widehat{\Sigma}^{-1}\bigotimes P_X\right)\mathbf{y}\right)$$

y el estimador MC2E de δ obtenido ecuación por ecuación es

$$\hat{\delta}_{MC2E} = \left(Z'\left(I_G\bigotimes P_X\right)Z\right)^{-1}\left(Z'\left(I_G\bigotimes P_X\right)\mathbf{y}\right)$$

El estadístico de contraste es

$$H = (\hat{\delta}_{MC2E} - \hat{\delta}_{MC3E})'[V(\hat{\delta}_{MC2E}) - V(\hat{\delta}_{MC3E})]^{-1}(\hat{\delta}_{MC2E} - \hat{\delta}_{MC3E})$$

el cual bajo la hipótesis nula de correcta especificación se distribuye asintóticamente como una χ^2 con J grados de libertad.

Hausman (1978) comenta en su artículo que bajo la alternativa, la distribución del estadístico es una χ^2 no centrada cuyo parámetro de no-centralidad es proporcional a $plim\frac{1}{T}\hat{Z}_j'\mathbf{u}_j \neq 0$ para cualquier ecuación j mal especificada en el sistema y también de la magnitud de $\hat{\sigma}^{ij}$. Si los elementos de la inversa de $\hat{\Sigma}$ son grandes el contraste será potente para una tamaño dado de inconsistencia. A medida que esos elementos tienden a cero, entonces MC3E se aproxima a MC2E y el contraste tendrá poca potencia.

Otra cuestión que comenta Hausman es que rechazar la hipótesis nula no nos indica qué ecuación o ecuaciones están mal especificadas. Solamente se sabe que existe una mala especificación en algún lugar del sistema. Si se tiene confianza en que alguna de las ecuaciones está correctamente especificada, entonces se podría utilizar para contrastar la especificación en el resto del sistema, por ejemplo de forma secuencial. Por ejemplo, si la ecuación 2 está correctamente especificada y se quiere contrastar la especificación de la tercera ecuación, entonces se debería comparar la estimación de los coeficientes de la segunda ecuación por MC2E y MC3E, considerando $\hat{\sigma}_{ij} = 0$ para $i \neq j$ excepto para $i = 2, j = 3$ y viceversa en la obtención del estimador MC3E. Utilizando este método se podría ir identificando donde se encuentra el problema de mala especificación en el resto del sistema. La cuestión es que es un contraste secuencial por lo que el tamaño final del contraste puede ser muy complicado de obtener.

Capítulo 6

Ejemplos Empíricos utilizando Gretl

6.1. Ejemplo 1: Inflación y grado de apertura.

Este ejemplo ilustra la estimación de un modelo de ecuaciones simultáneas por el método de Mínimos Cuadrados en 2 etapas (MC2E) utilizando el software libre Gretl[1]. Los datos se encuentran en el fichero de muestra **openness**.

Archivo → Abrir datos → Archivo de muestra

Seleccionar la pestaña Wooldridge. La información que nos muestra Gretl si pinchamos en

Datos → Ver descripción

es la siguiente:

```
Tipo de datos: sin fecha. Rango: 1 - 114 (n = 114)

Lista de variables:
           open  imports as % GDP, '73-
            inf  avg. annual inflation, '73-
          pcinc  1980 per capita inc., U.S. $
           land  land area, square miles
            oil  =1 if major oil producer
           good  =1 if 'good' data
         lpcinc  log(pcinc)
          lland  log(land)
          lopen  log(open)
           linf  log(inf)
        opendec  open/100
        linfdec  log(inf/100)
```

[1]El programa econométrico GRETL, junto con el manual y diversas bases de datos, se puede descargar de http://gretl.sourceforge.net

Son datos de sección cruzada para 114 países utilizados en el libro de Wooldridge, ejemplo 16.6, Inflación y grado de apertura. Este ejemplo utiliza el análisis realizado en el artículo de D. Romer, "Openess and Inflation: Theory and Evidence" publicado en 1993 en el Quarterly Journal of Economics.

6.1.1. Forma estructural y análisis de la identificación

Romer (1993) desarrolla diversos modelos teóricos sobre la inflación que implican que los países más abiertos deberían tener tasas de inflación más bajas. Su análisis empírico explica las tasas anuales medias de inflación (*Inf*), en términos de la proporción media de las importaciones en el PIB (o PNB) desde 1973, lo cual es su medida del grado de apertura (*open*). Si bien Romer no especifica ambas ecuaciones en un sistema simultáneo, tiene presente un sistema de dos ecuaciones, donde *Inf* y posiblemente *open* son variables endógenas: Forma estructural:

$$Inf = \gamma_{11} + \beta_{12} open + \gamma_{12} lpcinc + u_1 \tag{6.1}$$

$$Open = \gamma_{21} + \beta_{21} inf + \gamma_{22} lpcinc + \gamma_{23} lland + u_2 \tag{6.2}$$

Las variables exógenas se suponen que son *lpcinc* que es el logaritmo de la renta per cápita en dólares U.S.A de 1980 y *lland*, el logaritmo de la superficie del país en millas cuadradas (también supuesta exógena).

La primera ecuación estructural es la ecuación de interés, con la hipótesis de que las economías con mayor grado de apertura al exterior, manteniendo el resto de factores constantes, tienen unas tasas de inflación menores, esto es $\beta_{12} < 0$. La segunda ecuación refleja el hecho de que el grado de apertura podría depender de la tasa de inflación media, así como de otros factores. La variable *lpcinc* aparece en ambas ecuaciones, pero se supone que *lland* sólo aparece en la segunda ecuación. Se espera que $\gamma_{23} < 0$, esto es, manteniendo el resto de factores constantes, es probable que un país más pequeño tenga un grado de apertura mayor.

La primera ecuación está identificada, siempre y cuando se satisfaga la condición de rango, esto es $\gamma_{23} \neq 0$. La segunda ecuación no está identificada porque contiene ambas exógenas, por lo que no se satisface la condición de orden. Pero estamos interesados solamente en la primera, por lo que podemos estimar esta ecuación aislada por un método de información limitada como es Mínimos Cuadrados en 2 Etapas (MC2E), si queremos tener en cuenta la posible endogeneidad del regresor *open*.

6.1.2. Estimación por Mínimos cuadrados ordinarios

Primeramente procedemos a estimar la Forma Estructural por Mínimos Cuadrados Ordinarios (MCO), sin tener en cuenta esa posibilidad. El resultado obtenido es, una vez que hemos elegido en Gretl

Modelo → Mínimos Cuadrados Ordinarios

Modelo 1: estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	25,1040	15,2052	1,6510	0,1016
open	−0,215070	0,0946289	−2,2728	0,0250
lpcinc	0,0175673	1,97527	0,0089	0,9929

Media de la var. dependiente	17,2640
D.T. de la variable dependiente	23,9973
Suma de cuadrados de los residuos	62127,5
Desviación típica de los residuos ($\hat{\sigma}$)	23,6581
R^2	0,0452708
$\bar{R}^2$ corregido	0,0280685
$F(2, 111)$	2,63167
valor p para $F()$	0,0764453
Log-verosimilitud	−520,90
Criterio de información de Akaike	1047,80
Criterio de información Bayesiano de Schwarz	1056,01
Criterio de Hannan–Quinn	1051,13

El signo obtenido para el coeficiente que acompaña a *open* es el esperado, tal que parece haber evidencia de que las economías con un mayor grado de apertura tienen unas tasas de inflación menores. Esta conclusión puede no ser adecuada si la variable *open* es endógena, por lo que el estimador MCO, además de ser sesgado no sería consistente. Estimamos ahora esta ecuación por un método de variables instrumentales, Mínimos cuadrados en dos etapas.

6.1.3. Estimación por Mínimos Cuadrados en dos etapas

Estimación MC2E de la primera ecuación estructural:

Modelo → Otros modelos lineales → Mínimos cuadrados en dos etapas

Modelo 1: estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: inf
Instrumentos: const lland lpcinc

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	26,8993	15,4012	1,7466	0,0807
open	−0,337487	0,144121	−2,3417	0,0192
lpcinc	0,375823	2,01508	0,1865	0,8520

Media de la var. dependiente	17,2640
D.T. de la variable dependiente	23,9973
Suma de cuadrados de los residuos	63064,2
Desviación típica de los residuos ($\hat{\sigma}$)	23,8358
$F(2, 111)$	2,62498
valor p para $F()$	0,0769352
Criterio de información de Akaike	1049,51
Criterio de información Bayesiano de Schwarz	1057,72
Criterio de Hannan–Quinn	1052,84

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 1{,}35333$ con valor p = 0,244697

First-stage $F(1, 111) = 86{,}3734$

El valor First-stage $F(1, 111) = 86, 3734$ corresponde al cuadrado del estadístico t para contrastar la significatividad individual de la variable *lland* en la ecuación para *open* de la Forma reducida. Por lo tanto, un valor significativo indica que este instrumento presenta una correlación significativa para la variable a la que hace de instrumento, en este caso *open*.

El resultado del contraste de Hausman parece señalar que no hay evidencia de que la variable open sea endógena, ya que no se rechaza la hipótesis nula. Eso indicaría que la diferencia entre las estimaciones obtenidas entre el estimador MCO y MC2E no es estadísticamente significativa.

Veamos a continuación que el resultado obtenido por MC2E es equivalente a realizar las siguientes dos etapas:

1ª etapa: Estimación MCO de la Forma Reducida para open:

Modelo 2: estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	117,085	15,8483	7,3878	0,0000
lpcinc	0,546479	1,49324	0,3660	0,7151
lland	−7,5671	0,814216	−9,2937	0,0000

Media de la var. dependiente	37,0789
D.T. de la variable dependiente	23,7535
Suma de cuadrados de los residuos	35151,8
Desviación típica de los residuos ($\hat{\sigma}$)	17,7956
R^2	0,448668
$\bar{R}^2$ corregido	0,438734
$F(2, 111)$	45,1654

En la ventana de resultados de la estimación por MCO añadimos al conjunto de datos los valores ajustados de *open* en la estimación de la forma reducida.

Guardar → valores estimados

En la ventana que emerge `gretl:atributos de variable` le asignamos el nombre de *Openhat*.

2ª etapa: Estimamos la ecuación estructural por MCO pero añadiendo como regresor a *openhat* en lugar de *open*.

Modelo → Mínimos cuadrados ordinarios

Elegir como variable dependiente *inf* y como variables explicativas *openhat* y *lpcinc*. Obtenemos los siguientes resultados:

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	26,8993	15,2588	1,7629	0,0807
openhat	−0,337487	0,142788	−2,3635	0,0198
lpcinc	0,375823	1,99645	0,1882	0,8510

Media de la var. dependiente	17,2640
D.T. de la variable dependiente	23,9973
Suma de cuadrados de los residuos	61903,2
Desviación típica de los residuos ($\hat{\sigma}$)	23,6154
R^2	0,0487174
$\bar{R}^2$ corregido	0,0315772
$F(2, 111)$	2,84229
valor p para $F()$	0,0625432

Si comparamos estos resultados con los obtenidos anteriormente utilizando la opción MC2E, vemos que obtenemos las mismas estimaciones de los coeficientes, como era de esperar. Pero las desviaciones típicas han cambiado y estas últimas no son las adecuadas. La razón está en el cálculo de la suma de cuadrados residual en esta segunda etapa. Al considerar como regresor a *openhat* también se utiliza esta variable en el cómputo de los residuos, cuando se debería de utilizar la variable *open*.

Dado que *lpcinc* no es significativo ni en la estimación de la ecuación estructural ni en la forma reducida lo eliminamos del análisis. Los resultados y conclusiones no cambian, tanto en las estimaciones por MC2E como MCO.

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	25,2342	4,10212	6,1515	0,0000
open	−0,214952	0,0932751	−2,3045	0,0230

Suma de cuadrados de los residuos	62127,5
R^2	0,0452702
$\bar{R}^2$ corregido	0,0367458

Estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: inf
Instrumentos: const lland

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	29,6066	5,65827	5,2325	0,0000
open	−0,332874	0,140347	−2,3718	0,0177

Suma de cuadrados de los residuos 63014,1

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_1 = 1{,}33583$ con valor p = 0,247771
First-stage $F(1, 112) = 90{,}8997$

6.1.4. Análisis de la bondad de los instrumentos

Manteniendo *lpcinc* fuera del análisis, ¿Cuál de las siguientes variables *land* o *lland* es mejor instrumento para *open*? Realizamos la regresión de *open* sobre cada una de estas variables por separado y sobre las dos a la vez, y obtenemos los siguientes resultados:

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	40,4499	2,34209	17,2709	0,0000
land	−1,12797e-05	3,28925e-06	−3,4293	0,0008

Suma de cuadrados de los residuos	57699,6
R^2	0,0950218
$\bar{R}^2$ corregido	0,0869416

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	121,838	9,04381	13,4720	0,0000
lland	−7,6182	0,799045	−9,5341	0,0000

Suma de cuadrados de los residuos	35194,2
R^2	0,448003
$\bar{R}^2$ corregido	0,443075

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	129,217	10,4709	12,3406	0,0000
lland	−8,3978	0,975494	−8,6088	0,0000
land	4,33406e-06	3,13616e-06	1,3820	0,1698

Suma de cuadrados de los residuos	34598,9
R^2	0,457340
$\bar{R}^2$ corregido	0,447562
$F(2,111)$	46,7740

Cuando consideramos cada una por separado, ambas variables son significativas para explicar *open*, pero al considerarlas conjuntamente, la variable *lland* es la variable que se mantiene significativa. Esto indica que *lland* es mejor instrumento para *open*.

6.1.5. Extensiones del modelo original

Podemos tratar de identificar la segunda ecuación estructural añadiendo la variable ficticia *oil* a la primera ecuación estructural como una variable exógena. Esta variable indica si el país en cuestión es un gran productor de petróleo. Si no la incluimos en la segunda ecuación entonces ahora ambas ecuaciones están exactamente identificadas: la primera si $\gamma_{23} \neq 0$ y la segunda si $\gamma_{12} \neq 0$.

$$Inf = \gamma_{11} + \beta_{12}\, open + \gamma_{12}\, oil + u1 \tag{6.3}$$
$$Open = \gamma_{21} + \beta_{21}\, inf + \gamma_{23}\, lland + u2 \tag{6.4}$$

Estimación MC2E de la primera ecuación estructural utilizando las 114 observaciones:

Estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: inf
Instrumentos: const lland oil

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	29,7630	5,67387	5,2456	0,0000
open	−0,328101	0,141060	−2,3260	0,0200
oil	−5,4289	9,30221	−0,5836	0,5595

Suma de cuadrados de los residuos	62750,2
$F(2, 111)$	2,81668
valor p para $F()$	0,0640855

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: χ^2_1 = 1,29195 con valor p = 0,255689

First-stage $F(1, 111)$ = 89,6277

Dado el valor del estadístico t asociado al coeficiente de la variable *oil*, el ser productor de petróleo no parece tener un efecto estadísticamente significativo sobre la inflación. El valor $p = 0,560658$ es mayor que cualquier nivel de significación razonable por lo que no se rechaza la hipótesis nula de que el coeficiente sea cero. ¿Qué implicación tiene esto para la identificación de la segunda ecuación? Como hemos dicho antes, la segunda ecuación estará identificada si ese coeficiente es distinto de cero. La condición de rango para la identificación de la segunda ecuación no parece satisfacerse.

Procedemos a la estimación por MCO de la Forma Reducida:

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: inf

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−10,161	12,0922	−0,8403	0,4025
lland	2,49704	1,06442	2,3459	0,0208
oil	−5,8015	9,21258	−0,6297	0,5302

Suma de cuadrados de los residuos	61688,0
R^2	0,0520243
$\bar{R}^2$ corregido	0,0349437
$F(2, 111)$	3,04581
valor p para $F()$	0,0515518

Estimaciones MCO utilizando las 114 observaciones 1–114
Variable dependiente: open

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	121,684	9,13247	13,3243	0,0000
lland	−7,6106	0,803892	−9,4672	0,0000
oil	1,13547	6,95768	0,1632	0,8707

Suma de cuadrados de los residuos	35185,8
R^2	0,448136
$\bar{R}^2$ corregido	0,438192
$F(2, 111)$	45,0682

Vemos que tanto en la primera ecuación como en la segunda, la variable exógena *oil* no es significativa. Esto también parece indicar que esta variable no es un buen instrumento para identificar la segunda ecuación estructural. ¿Qué ocurrirá entonces si la estimamos por MC2E?

Estimaciones MC2E utilizando las 114 observaciones 1–114
Variable dependiente: open
Instrumentos: const lland oil

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	119,695	16,1341	7,4188	0,0000
inf	−0,195718	1,21178	−0,1615	0,8717
lland	−7,1218	3,17814	−2,2409	0,0250

Suma de cuadrados de los residuos	35922,9
$F(2, 111)$	43,2180

Contraste de Hausman –
  Hipótesis nula: Los estimadores de MCO son consistentes
  Estadístico de contraste asintótico: $\chi^2_1 = 0{,}0118219$ con valor p = 0,913418
First-stage $F(1, 111) = 0{,}396574$

El coeficiente que acompaña a *inf* no es estadísticamente significativo y el contraste de Hausman no parece rechazar la hipótesis nula de que esta variable no es endógena. Esto último puede ser consecuencia de que el coeficiente que acompaña a *inf* en la segunda ecuación estructural no está bien identificado utilizando como instrumento la variable *oil*.

6.2. Ejemplo 2: Oferta y Demanda de trabajo.

Para este ejemplo utilizamos los datos que están disponibles en Gretl como archivo de muestra **Mroz** en la pestaña asociada a Wooldridge. La descripción que nos muestra Gretl de los datos es la siguiente:

```
Source:  T.A. Mroz (1987), "The Sensitivity of an Empirical Model
of Married Women's Hours of Work to Economic and Statistical
Assumptions," Econometrica 55, 765-799.

Professor Ernst R. Berndt, of MIT, kindly provided the data, which
he obtained from Professor Mroz.

Used in Text:  pages 241-243, 251-252, 280-281, 490-491, 501, 506,
508,536-537, 550 561-563, 569-572, 590-591, 596-597

Tipo de datos: sin fecha Rango: 1 - 753 (n = 753)

Lista de variables:

        inlf  =1 if in labor force, 1975
       hours  hours worked, 1975
     kidslt6  # kids < 6 years
     kidsge6  # kids 6-18
         age  woman's age in yrs
        educ  years of schooling
        wage  estimated wage from earns., hours
     repwage  reported wage at interview in 1976
      hushrs  hours worked by husband, 1975
      husage  husband's age
     huseduc  husband's years of schooling
     huswage  husband's hourly wage, 1975
      faminc  family income, 1975
         mtr  fed. marginal tax rate facing woman
    motheduc  mother's years of schooling
    fatheduc  father's years of schooling
        unem  unem. rate in county of resid.
        city  =1 if live in SMSA
       exper  actual labor mkt exper
    nwifeinc  (faminc - wage*hours)/1000
       lwage  log(wage)
     expersq  exper^2
```

Las variables que vamos a utilizar de este fichero son las siguientes:

- hours: horas trabajadas, 1975.
- kidslt6: número de hijos en la familia menores de 6 años.
- kidsge6: número de hijos en la familia entre 6 y 18 años.
- age: edad de la mujer en años.

- educ: años de educación.
- lwage: logaritmo del salario por hora.
- exper: experencia laboral previa.
- nwifeinc: ingreso no salarial de la mujer en la familia, incluye el salario del marido.
- expersq: experiencia al cuadrado.

Son datos de sección cruzada y aunque el fichero tiene un total de 753 individuos, todos ellos mujeres, solamente consideramos las que se encontraban trabajando en 1975 que son un total de 428. Las variables consideradas endógenas son *hours* y *lwage*, siendo el resto variables exógenas.

Partimos de la siguiente forma estructural del modelo donde la primera ecuación es la función de oferta de mujeres casadas trabajadoras y la segunda ecuación es una función de oferta salarial ya que hemos normalizado con respecto a *lwage*. Esto implica que *lwage* es la variable dependiente de esa ecuación.

$$\begin{aligned} hours &= \gamma_{11} + \beta_{12} lwage + \gamma_{12} educ + \gamma_{13} age + \gamma_{14} kidslt6 + \gamma_{15} kidsge6 + \gamma_{16} nwifeinc + u_1 \\ lwage &= \gamma_{21} + \beta_{21} hours + \gamma_{22} educ + \gamma_{23} exper + \gamma_{24} expersq + u_2 \end{aligned} \tag{6.5}$$

6.2.1. Identificación

La primera ecuación satisface la condición de orden, necesaria para la identificación, con exceso. El número de variables exógenas excluidas de esa ecuación, $K_1^* = 2$, es mayor que el número de endógenas incluidas en la ecuación como explicativas, $G_1 = 1$. Por lo tanto, la ecuación estará sobreidentificada si se satisface la condición de rango. Est condición, necesaria y suficiente para la identificación, se satisface si $\gamma_{23} \neq 0$ y \ o $\gamma_{24} \neq 0$.

La segunda ecuación satisface la condición de orden también con exceso. El número de variables exógenas excluidas de esa ecuación, $K_2^* = 4$, es mayor que el número de endógenas incluidas en la ecuación como explicativas, $G_2 = 1$. Por lo tanto, la ecuación estará sobreidentificada si se satisface la condición de rango. Esta condición, necesaria y suficiente para la identificación, se satisface si $\gamma_{13} \neq 0$ y, o $\gamma_{14} \neq 0$ y, o $\gamma_{15} \neq 0$ y, o $\gamma_{16} \neq 0$.

Por lo tanto, si se satisfacen las condiciones de rango en cada ecuación, el sistema estará identificado, es decir, solamente existirá una estructura o valores de los parámetros estructurales consistente con la Forma reducida que es lo que observamos en los datos. Es decir, podemos recuperar de forma única los parámetros de la forma estructural a partir de los parámetros de la forma reducida del modelo. La forma reducida del modelo es la siguiente:

$$\begin{aligned} hours &= \pi_{11} + \pi_{12} educ + \pi_{13} age + \pi_{14} kidslt6 + \pi_{15} kidsge6 + \pi_{16} nwifeinc + \pi_{17} exper + \pi_{18} expersq + v_1 \\ lwage &= \pi_{21} + \pi_{22} educ + \pi_{23} age + \pi_{24} kidslt6 + \pi_{25} kidsge6 + \pi_{26} nwifeinc + \pi_{27} exper + \pi_{28} expersq + v_2 \end{aligned} \tag{6.6}$$

6.2.2. Estimación de la Forma Estructural

Estimaciones MCO

Primera ecuación: Función de oferta de mujeres casadas trabajadoras

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: hours

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2114,70	340,131	6,2173	0,0000
lwage	−17,407	54,2154	−0,3211	0,7483
educ	−14,444	17,9679	−0,8039	0,4219
age	−7,7299	5,52945	−1,3980	0,1629
kidslt6	−342,50	100,006	−3,4248	0,0007
kidsge6	−115,02	30,8293	−3,7309	0,0002
nwifeinc	−4,2458	3,65581	−1,1614	0,2461

Media de la var. dependiente	1302,93
D.T. de la variable dependiente	776,274
Suma de cuadrados de los residuos	2,40083e+08
Desviación típica de los residuos ($\hat{\sigma}$)	755,161
R^2	0,0669555
$\bar{R}^2$ corregido	0,0536579
$F(6, 421)$	5,03518
valor p para $F()$	5,30546e-05
Log-verosimilitud	−3440,1
Criterio de información de Akaike	6894,21
Criterio de información Bayesiano de Schwarz	6922,62
Criterio de Hannan–Quinn	6905,43

La estimación MCO del coeficiente que afecta a *lwage* indica una función de oferta con pendiente negativa, signo no esperado, aunque el coeficiente no es significativamente distinto de cero.

Las únicas variables significativas, además de la constante, son las que recogen el efecto marginal de tener un hijo más en la familia, tanto si es menor de 6 como si está entre 6 y 18 años.

Manteniendo el resto de factores constante, se estima que un hijo más en la familia menor de 6 años disminuye en media el número de horas trabajadas al año en 342 horas. Y adicionalmente, manteniendo el resto de variables constante, un hijo más en la familia de entre 6 y 18 años disminuye en media el número de horas trabajadas al año en 115 horas.

Segunda ecuación: Función de oferta salarial.

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: lwage

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,461995	0,203848	−2,2664	0,0239
hours	−5,65486e-05	4,37823e-05	−1,2916	0,1972
educ	0,106214	0,0141698	7,4958	0,0000
exper	0,0447035	0,0133870	3,3393	0,0009
expersq	−0,000858544	0,000394639	−2,1755	0,0301

Media de la var. dependiente	1,19017
D.T. de la variable dependiente	0,723198
Suma de cuadrados de los residuos	187,565
Desviación típica de los residuos ($\hat{\sigma}$)	0,665896
R^2	0,160133
$\bar{R}^2$ corregido	0,152191
$F(4, 423)$	20,1627
Log-verosimilitud	−430,75
Criterio de información de Akaike	871,513
Criterio de información Bayesiano de Schwarz	891,809
Criterio de Hannan–Quinn	879,529

El coeficiente estimado que acompaña a *hours* tampoco tiene el signo esperado, aunque tampoco es estadísticamente significativo, por lo que no hay evidencia de que la oferta salarial varíe con las horas trabajadas. El resto de variables *educ*, *exper* y *expersq* sí son significativas con los signos esperados, recogiendo esta última variable una relación cuadrática del logaritmo del salario con el nivel de experiencia.

Puede que estos efectos no esperados sean consecuencia de los sesgos en la estimación por MCO si *hours* y *lwage* son ambas variables endógenas. Consideremos estimar por métodos que tienen en cuenta esa posible endogeneidad. Estos métodos pueden ser de información limitada, Mínimos cuadrados en dos etapas (MC2E) y Máxima Verosimilitud información limitada (MVIL), o de información completa como son el de Mínimos cuadrados en tres etapas (MC3E) y el de Máxima Verosimilitud información completa (MVIC).

Para estimar un sistema de ecuaciones en Gretl por todos estos métodos de estimación elegimos:

Modelo → Ecuaciones simultáneas

En la ventana que aparece se han de escribir las ecuaciones del modelo de la siguiente forma:

```
equation hours const lwage educ age    kidslt6 kidsge6 nwifeinc
equation lwage const hours educ exper  expersq
endog    hours lwage
```

Dado que en la tercera línea se han identificado cuales son las variables endógenas del sistema, Gretl asigna al resto la categoría de variables exógenas.

En la parte inferior de la ventana elegimos el estimador que se quiere utilizar, dentro de un listado que nos muestra el programa. Los resultados de estimar cada una de estas ecuaciones por los distintos métodos se muestran en el siguiente cuadro resumen:

Modelo MROZ - Estimación Forma Estructural.

Comparación de resultados:

- Primera ecuación estructural:

Función de oferta de mujeres casadas trabajadoras
Variable dependiente: hours

Explicativas	MCO	MC2E	MVIL	MC3E	MVIC
const	$2114,7^*$	$2432,20^*$	$2449,33^*$	$2504,80^*$	$2435,10^*$
	$(340,131)$	$(594,172)$	$(616,070)$	$(535,892)$	$(579,001)$
lwage	$-17,4078$	$1544,82^*$	$1629,13^*$	$1676,93^*$	$1773,93^*$
	$(54,2154)$	$(480,739)$	$(510,876)$	$(431,169)$	$(497,304)$
educ	$-14,4449$	$-177,449^*$	$-186,247^*$	$-205,027^*$	$-216,729^*$
	$(17,9679)$	$(58,1426)$	$(61,3963)$	$(51,8473)$	$(61,8412)$
age	$-7,72998$	$-10,7841$	$-10,9489$	$-12,2812$	$-10,5961$
	$(5,52945)$	$(9,57735)$	$(9,92583)$	$(8,26153)$	$(8,84614)$
kidslt6	$-342,505^*$	$-210,834$	$-203,727$	$-200,567$	$-167,984$
	$(100,006)$	$(176,934)$	$(183,576)$	$(134,268)$	$(143,024)$
kidsge6	$-115,021^*$	$-47,5571$	$-43,9160$	$-48,6399$	$-40,8436$
	$(30,8293)$	$(56,9179)$	$(59,1775)$	$(35,9514)$	$(36,5103)$
nwifeinc	$-4,24581$	$-9,24912$	$-9,51916$	$0,367895$	$1,24342$
	$(3,65581)$	$(6,48112)$	$(6,72509)$	$(3,45152)$	$(2,13017)$

$*$ significativo al 5 % $N = 428$

Menor valor propio estimado = 1,00194

Contraste LR: Chi-cuadrado(1)= 0,829301 v.p. = 0,3625
Contraste de Sargan : TR2 = 0,858169 valor p = 0,354252
Contraste de Hausman - Chi-cuadrado(1) = 35,9481

a) La primera ecuación estructural o curva de oferta estimada tanto por MC2E, MVIL, MC3E y MVIC tiene una pendiente positiva y significativa[2]. A efectos de comparación cuando esta ecuación de oferta de trabajo se estima por MCO, el coeficiente estimado asociado a *lwage* es negativo y no significativo, lo que implica la ausencia de efectos de variaciones en el salario sobre las horas trabajadas.

[2]Esta pendiente la recoge el coeficiente asociado a *lwage*.

b) El coeficiente estimado para *lwage* se interpreta como la variación estimada en el número de horas ofrecidas por una mujer trabajadora a una variación unitaria en el salario, manteniendo constantes los demás factores. Por ejemplo, utilizando la estimación MC2E, sería de un 15, 4 % de la variación en el salario. Utilizando MVIL, MC3E y MVIC es algo mayor.

c) La elasticidad de la oferta de trabajo respecto del salario no es constante en este modelo, porque es *hours* y no *log(hours)* la variable dependiente en la ecuación de oferta, primera ecuación del modelo. Por ejemplo, utilizando el método MC2E, al nivel medio de horas trabajadas, que es de 1303, la elasticidad estimada es de 1540/1303 que es algo menor que 1, 18. Esto implica que en respuesta a un incremento salarial del 1 %, se da un incremento de más de un 1 % en las horas trabajadas. La elasticidad será menor para niveles más altos de horas ofrecidas. A niveles más bajos, como por ejemplo $hours = 700$, la elasticidad es superior a 2.

d) También hay diferencias importantes para los efectos de la variable *educ* entre MCO y el resto de métodos que tienen en cuenta la posible endogeneidad de *lwage*. El coeficiente estimado que acompaña a esta variable es negativo en todos los casos pero no así su significatividad. Utilizando MCO no es significativo pasando a serlo con el resto de métodos, así como aumentando considerablemente su magnitud en valor absoluto.

e) En cuanto al resto de variables, también cambia la significatividad de los coeficientes asociados a la existencia de hijos menores de 6 años, *kidslt6*, e hijos entre 6 y 18 años, *kidsge6*, pasando de ser ambas variables significativas con MCO a no serlo con el resto de métodos, aunque el signo negativo se mantiene y es el esperado. Por otro lado, las variables *age* y *nwifeinc* no son significativas en todos los casos.

- Segunda ecuación estructural:

Función de oferta salarial

Variable dependiente: lwage

Explicativas	MCO	MC2E	MVIL	MC3E	MVIC
const	$-0,461995^{*}$	$-0,69279^{*}$	$-0,735315^{*}$	$-0,705110^{*}$	$-0,740600^{*}$
	$(0,203848)$	$(0,3066)$	$(0,324821)$	$(0,304590)$	$(0,314122)$
hours	$-0,000056$	$0,000161$	$0,000201$	$0,000201$	$0,0002456$
	$(0,0000438)$	$(0,000215)$	$(0,0002362)$	$(0,0002108)$	$(0,0002232)$
educ	$0,106214^{*}$	$0,111118^{*}$	$0,112021^{*}$	$0,112970^{*}$	$0,113986^{*}$
	$(0,0141698)$	$(0,0153319)$	$(0,0156374)$	$(0,0151452)$	$(0,01562)$
exper	$0,0447035^{*}$	$0,032646$	$0,0304243$	$0,0208906$	$0,0171624$
	$(0,013387)$	$(0,018061)$	$(0,0189511)$	$(0,0142782)$	$(0,0142774)$
expersq	$-0,000858^{*}$	$-0,0006765$	$-0,000643$	$0,00029433$	$-0,0002381$
	$(0,000395)$	$(0,0004426)$	$(0,000454)$	$(0,0002614)$	$(0,0002261)$

$*$ significativo al 5 % $N = 428$

Menor valor propio estimado = 1,00685

Contraste LR: Chi-cuadrado(3) = 2,92124 con valor p = 0,4039
Contraste de Sargan : TR2 = 2,94083 con valor p = 0,4008
Contraste de Hausman - Chi-cuadrado(1) = 1,14311
Contraste de sobreidentificación en todo el sistema de Hansen-Sargan:
Chi-cuadrado(4) = 4,10677 con valor p = 0,3917

a) El coeficiente que acompaña a la variable *hours* no es significativo tanto utilizando MCO como el resto de métodos, aunque el signo cambia siendo negativo en el primer caso y positivo en el resto.

b) En cuanto al resto de variables, el coeficiente que acompaña a *educ* se mantiene con signo positivo y significativo en todos los casos, no así para las variables que recogen el efecto sobre la oferta salarial del nivel de experiencia. Los signos de estos coeficientes se mantienen pero pasan de ser significativos con MCO a no serlo en el resto de métodos.

- Contrastes de especificación y exogeneidad.

a) Contrastes de Sargan y Hausman.
Al estimar por MC2E en Gretl utilizando

Modelo → Otros modelos lineales → Mínimos Cuadrados en 2 etapas

además de mostrar los resultados de la estimación por MC2E también nos muestra los resultados del Contraste de sobreidentificación de Sargan y el contraste de Hausman comparando MC2E y MCO. En cambio, si se utiliza la opción

Modelo → Ecuaciones simultáneas → Mínimos Cuadrados en 2 etapas

solamente se muestra el resultado del contraste de Sargan.
Los resultados de estos contrastes en cada una de las ecuaciones indica lo siguiente:

- Tanto en el caso de la primera ecuación como en la segunda, el contraste de sobreidentificación de Sargan no rechaza la hipótesis nula de que todos los instrumentos son válidos.
- En cuanto a los resultados del contraste de Hausman, para la primera ecuación estructural, se rechaza la hipótesis nula de que la variable *lwage* no esté correlacionado con el término de perturbación de esa ecuación. Esto indica evidencia de endogeneidad de *lwage*. Considerando en cambio la segunda ecuación estructural, el contraste de Hausman no rechaza la hipótesis nula de incorrelación entre la variable *hours* y el término de preturbación de esa ecuación.

b) Los contrastes de sobreidentificación LR en ambas ecuaciones no rechazan la hipótesis nula, lo que indica que se aceptan como buenas las restricciones de identificación en cada ecuación.

c) Por último el contraste de sobreidentificación de Hansen-Sargan sobre la bondad de las restricciones de sobreidentificación en el sistema no rechaza la hipótesis nula. Los instrumentos parecen adecuados.

6.2.3. Estimación de la Forma Reducida

La estimación de la forma reducida nos da la posibilidad de analizar la bondad de los instrumentos para cada ecuación. Podemos contrastar el grado de correlación de los instrumentos, o variables exógenas, con las variables endógenas *hours* y *lwage*.

Forma Reducida: Ecuación para hours.

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: hours

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2056,64	346,484	5,9357	0,0000
educ	−22,788	16,4345	−1,3866	0,1663
age	−19,663	5,89403	−3,3362	0,0009
kidslt6	−305,72	96,4501	−3,1697	0,0016
kidsge6	−72,366	30,3610	−2,3835	0,0176
nwifeinc	0,443852	3,61350	0,1228	0,9023
exper	47,0051	14,5565	3,2291	0,0013
expersq	−0,513644	0,437358	−1,1744	0,2409

Media de la var. dependiente	1302,93
D.T. de la variable dependiente	776,274
Suma de cuadrados de los residuos	2,21208e+08
Desviación típica de los residuos ($\hat{\sigma}$)	725,731
R^2	0,140308
$\bar{R}^2$ corregido	0,125980
$F(7, 420)$	9,79245
Log-verosimilitud	−3422,5
Criterio de información de Akaike	6861,16
Criterio de información Bayesiano de Schwarz	6893,63
Criterio de Hannan–Quinn	6873,99

```
Contraste de omisión de variables -
 Hipótesis nula: los parámetros son cero para las variables
    age
    kidslt6
    kidsge6
    nwifeinc
 Estadístico de contraste: F(4, 420) = 4,80035
 con valor p  = P(F(4, 420) > 4,80035) = 0,00085263
```

Se rechaza la hipótesis nula, por lo que alguno o algunos de los instrumentos *age*, *kidslt6*, *kidsge6*, *nwifeinc* disponibles para instrumentalizar *hours* en la segunda ecuación (ecuación de oferta salarial) están suficientemente correlacionados con la variable a instrumentalizar que es la variable *hours*.

Forma Reducida: Ecuación para lwage

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: lwage

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,357997	0,318296	−1,1247	0,2613
educ	0,0998844	0,0150975	6,6160	0,0000
age	−0,00352039	0,00541452	−0,6502	0,5159
kidslt6	−0,0558725	0,0886034	−0,6306	0,5287
kidsge6	−0,0176485	0,0278910	−0,6328	0,5272
nwifeinc	0,00569422	0,00331952	1,7154	0,0870
exper	0,0407097	0,0133723	3,0443	0,0025
expersq	−0,000747326	0,000401777	−1,8601	0,0636

Media de la var. dependiente	1,19017
D.T. de la variable dependiente	0,723198
Suma de cuadrados de los residuos	186,680
Desviación típica de los residuos ($\hat{\sigma}$)	0,666690
R^2	0,164098
$\bar{R}^2$ corregido	0,150167
$F(7, 420)$	11,7788
Log-verosimilitud	−429,74
Criterio de información de Akaike	875,488
Criterio de información Bayesiano de Schwarz	907,961
Criterio de Hannan–Quinn	888,313

```
Contraste de omisión de variables -
 Hipótesis nula: los parámetros son cero para las variables
    exper
    expersq
 Estadístico de contraste: F(2, 420) = 8,25023
 con valor p  = P(F(2, 420) > 8,25023) = 0,000305887
```

Se rechaza la hipótesis nula, por lo que para instrumentalizar a *lwage* en la primera ecuación estructural o función de oferta de trabajo, los instrumentos *exper* y *expersq* parecen estar correlacionadas con esta variable.

¿Qué ocurrirá en cambio, si en lugar de normalizar en la segunda ecuación estructural para *lwage* se hace en función de *hours* ?

Los instrumentos que se tienen para instrumentalizar a *lwage* en la segunda ecuación estructural son *kidslt6*, *kidsge6*, *age*, *nwifeinc*. Si realizamos el contraste de omisión de estas variables en la ecuación de la forma reducida para *lwage* se obtiene:

```
Contraste de omisión de variables -
 Hipótesis nula: los parámetros son cero para las variables
    kidslt6
    kidsge6
    age
    nwifeinc
 Estadístico de contraste: F(4, 420) = 0,914214 con valor p = 0,455487
```

Estos instrumentos no parecen estar correlacionadas con *lwage*, dado que no se rechaza la hipótesis nula. ¿Qué consecuencias tiene este hecho en la estimación de la forma estructural? Esto lo tratamos en la siguiente sección.

6.2.4. Cambio de normalización de la segunda ecuación estructural.

¿Qué ocurre si en la segunda ecuación de la Forma Estructural normalizamos con respecto a *hours* y tenemos a *lwage* como endógena explicativa? En este caso en la segunda ecuación tenemos que utilizar como instrumentos para *lwage* a las variables exógenas que no aparecen en esa ecuación, esto es, *kidslt6*, *kidsge6*, *age* y *nwifeinc*.

Si estimamos el sistema de ecuaciones por MC2E, que es un método de información limitada, obtenemos en este caso los siguientes resultados:

Sistema de ecuaciones, Mínimos cuadrados en dos etapas

Ecuación 1, Forma Estructural: Estimaciones MC2E utilizando las 428 observaciones
Variable dependiente: hours
Instrumentos: const educ age kidslt6 kidsge6 nwifeinc exper expersq

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2432,20	594,172	4,0934	0,0000
educ	−177,44	58,1426	−3,0520	0,0023
age	−10,784	9,57735	−1,1260	0,2602
kidslt6	−210,83	176,934	−1,1916	0,2334
kidsge6	−47,557	56,9179	−0,8355	0,4034
nwifeinc	−9,2491	6,48112	−1,4271	0,1536
lwage	1544,82	480,739	3,2134	0,0013

Media de la var. dependiente	1302,93
D.T. de la variable dependiente	776,274
Suma de cuadrados de los residuos	7,13583e+08
Desviación típica de los residuos ($\hat{\sigma}$)	1301,91
$F(6, 421)$	0,0812210
valor p para $F()$	0,997966
Criterio de información de Akaike	7360,43
Criterio de información Bayesiano de Schwarz	7388,85
Criterio de Hannan–Quinn	7371,66

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi_1^2 = 35{,}9481$
con valor p = 2,0264e-009

Contraste de sobreidentificación de Sargan –
Hipótesis nula: todos los instrumentos son válidos
Estadístico de contraste: $TR^2 = 0{,}858169$ con valor p $= P(\chi_1^2 > 0{,}858169) = 0{,}354252$

First-stage $F(2, 420) = 8{,}25023$

Ecuación 2, Forma Estructural: estimaciones MC2E utilizando las 428 observaciones
Variable dependiente: hours
Instrumentos: const educ age kidslt6 kidsge6 nwifeinc exper expersq

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	1584,15	520,055	3,0461	0,0023
educ	−130,10	89,2759	−1,4574	0,1450
lwage	1000,53	805,418	1,2423	0,2141
exper	13,8850	39,1525	0,3546	0,7229
expersq	−0,0257319	0,891065	−0,0289	0,9770

Media de la var. dependiente	1302,93
D.T. de la variable dependiente	776,274
Suma de cuadrados de los residuos	4,46004e+08
Desviación típica de los residuos ($\hat{\sigma}$)	1026,83
$F(4, 423)$	0,319902
valor p para $F()$	0,864643
Criterio de información de Akaike	7155,29
Criterio de información Bayesiano de Schwarz	7175,58
Criterio de Hannan–Quinn	7163,30

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi_1^2 = 3{,}51544$
con valor p = 0,0607996

Contraste de sobreidentificación de Sargan –
Hipótesis nula: todos los instrumentos son válidos
Estadístico de contraste: TR^2 = 8,14343
con valor p = $P(\chi^2_3 >$ 8,14343) = 0,0431385

First-stage $F(4, 420)$ = 0,914214
Un valor < 10 puede indicar que los instrumentos son débiles.

Como era de esperar los resultados para la primera ecuación no cambian porque es un método de información limitada y no influye el cambio en la normalización de la segunda ecuación. Pero en el caso de los coeficientes de la segunda ecuación, ahora ninguno es significativo excepto la constante. Esto se debe a que los instrumentos *age*, *kidslt6*, *kidsge6* y *nwifeinc* están muy poco correlacionados con la variable a la que están instrumentalizando ahora que es *lwage*. Esto lo vimos en el análisis usando la forma reducida.

Lo mismo ocurre si estimamos por MVIL. Para ello elegimos

Modelo → Ecuaciones simultáneas

En la ventana que aparece se han de escribir las ecuaciones del modelo de la siguiente forma:

equation hours const lwage educ age kidslt6 kidsge6 nwifeinc
equation hours const lwage educ exper expersq
endog hours lwage

Una vez elegido el método de máxima verosimilitud información limitada, el resultado que muestra Gretl es el siguiente:

```
Endogenous variables: hours lwage

Exogenous variables: const educ age kidslt6 kidsge6 nwifeinc exper
expersq

Sistema de ecuaciones, Máxima Verosimilitud con Información
Limitada

Ecuación 1: estimaciones MVIL utilizando las 428 observaciones
1-428 Variable dependiente: hours

 VARIABLE        COEFICIENTE       DESV.TÍP.         ESTAD T   VALOR P
  const             2449,33           616,070          3,976   0,00007 ***
  lwage             1629,13           510,876          3,189   0,00143 ***
  educ              -186,247           61,3963        -3,034   0,00242 ***
  age                -10,9489           9,92583       -1,103   0,26999
  kidslt6           -203,727          183,576         -1,110   0,26710
  kidsge6            -43,9160          59,1775        -0,742   0,45802
  nwifeinc            -9,51916          6,72509       -1,415   0,15693
```

```
  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 7,66074e+008
  Desviación típica de los residuos = 1348,95
  Log-verosimilitud = -6016,92
  Smallest eigenvalue = 1,00194
  Contraste de sobreidentificación LR:
    Chi-cuadrado(1) = 0,829301 con valor p  0,3625

Ecuación 2: estimaciones MVIL utilizando las 428 observaciones
1-428 Variable dependiente: hours

 VARIABLE        COEFICIENTE       DESV.TÍP.         ESTAD T   VALOR P
  const              3660,92          3224,49          1,135   0,25623
  lwage              4978,71          5854,57          0,850   0,39510
  educ               -557,721          633,548        -0,880   0,37869
  exper              -151,474          252,721        -0,599   0,54893
  expersq               3,20134          5,16668       0,620   0,53551

  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 5,0292e+009
  Desviación típica de los residuos = 3448,1
  Log-verosimilitud = -6017,97
  Smallest eigenvalue = 1,00685
  Contraste de sobreidentificación LR:
    Chi-cuadrado(3) = 2,92124 con valor p  0,4039

Matriz de covarianzas cruzada residual (correlaciones por encima
de la diagonal principal)

  1,7899e+006      (0,934)
  4,2812e+006  1,1750e+007

logaritmo del determinante = 28,6254
```

Los resultados para la primera ecuacion no cambian porque es un método de informacion limitada y no influye el cambio en la normalización de la segunda ecuación. Pero en el caso de los coeficientes de la segunda ecuación ninguno es significativo como ocurría con MC2E. ¿Y si estimamos por un método de Información Completa, por ejemplo MC3E o MVIC?

Sistema de ecuaciones, Mínimos cuadrados en tres etapas

```
Ecuación 1: estimaciones MC3E utilizando las 428 observaciones
1-428 Variable dependiente: hours Instrumentos: const educ age
kidslt6 kidsge6 nwifeinc exper expersq

 VARIABLE        COEFICIENTE       DESV.TÍP.         ESTAD T   VALOR P
  const             1575,53          406,597           3,875   0,00011 ***
  lwage             1649,56          425,531           3,876   0,00011 ***
  educ             -197,057           52,2255         -3,773   0,00016 ***
  age                 6,36431          4,29410         1,482   0,13831
  kidslt6            97,7946          70,6825          1,384   0,16649
  kidsge6            25,1756          22,5259          1,118   0,26373
  nwifeinc           -2,96603          5,98872        -0,495   0,62041

  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 7,91079e+008
  Desviación típica de los residuos = 1359,53

Ecuación 2: estimaciones MC3E utilizando las 428 observaciones
1-428 Variable dependiente: hours Instrumentos: const educ age
kidslt6 kidsge6 nwifeinc exper expersq

 VARIABLE          COEFICIENTE        DESV.TÍP.        ESTAD T    VALOR P
  const             1525,46          506,819           3,010   0,00261 ***
  lwage              938,526         793,394           1,183   0,23684
  educ              -124,260          88,1676         -1,409   0,15873
  exper               24,7100         34,0601          0,725   0,46816
  expersq             -0,377879        0,639631       -0,591   0,55467

  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 4,22055e+008
  Desviación típica de los residuos = 993,03

Contraste de sobreidentificación de Hansen-Sargan:
  Chi-cuadrado(4) = 8,4736 con valor p  0,0757
```

El valor del estadístico de Hansen-Sargan es un valor bastante mayor que el que se obtenía con la normalización para *lwage*. Se rechaza la hipótesis nula al 1 %. En este caso el contraste de sobre identificación señala que los instrumentos no son adecuados.

Sistema de ecuaciones, Máxima Verosimilitud con Información Completa

El resultado mostrado por Gretl es el siguiente:

```
Endogenous variables: hours lwage

Exogenous variables:

const educ age kidslt6 kidsge6 nwifeinc exper expersq

Se alcanzó la convergencia después de 20 iteraciones
Log-verosimilitud = -3853,14

Ecuación 1: estimaciones MVIC utilizando las 428 observaciones
1-428 Variable dependiente: hours

 VARIABLE        COEFICIENTE       DESV.TÍP.          ESTAD T   VALOR P
  const            2435,10          579,001           4,206   0,00003 ***
  lwage            1773,93          497,304           3,567   0,00036 ***
  educ            -216,729           61,8412         -3,505   0,00046 ***
  age              -10,5961           8,84614        -1,198   0,23099
  kidslt6         -167,984          143,024          -1,175   0,24019
  kidsge6          -40,8436          36,5103         -1,119   0,26327
  nwifeinc           1,24342          2,13017         0,584   0,55941

  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 8,68212e+008
  Desviación típica de los residuos = 1424,27

Ecuación 2: estimaciones MVIC utilizando las 428 observaciones
1-428 Variable dependiente: hours

 VARIABLE        COEFICIENTE       DESV.TÍP.          ESTAD T   VALOR P
  const            3015,81         1951,40            1,545   0,12224
  lwage            4072,11         3700,34            1,100   0,27113
  educ            -464,165          405,994          -1,143   0,25292
  exper            -69,8873         116,278          -0,601   0,54781
  expersq            0,969511         1,68077         0,577   0,56406

  Media de la var. dependiente = 1302,93
  Desviación típica de la var. dependiente. = 776,274
  Suma de cuadrados de los residuos = 3,46719e+009
  Desviación típica de los residuos = 2846,21
```

Observamos lo mismo que ocurría con los métodos de información limitada. La especificación de la segunda ecuación normalizando por *hours* en lugar de por *lwage* no es la más adecuada ya que no se pueden identificar bien los coeficientes con los instrumentos disponibles.

En el caso del método de estimación MC3E, a diferencia de lo que ocurría en los métodos de información limitada, la normalización de la segunda ecuación afecta tanto a la estimación de la primera ecuación como de la segunda. En cambio, en el caso de MVIC, la estimación de la primera ecuación no se ve afectada por la normalización elegida para la segunda ecuación.

6.2.5. Contraste de restricciones de sobreidentificación uniecuacional: Estadístico de Sargan

Aunque Gretl muestra el valor de este estadístico cuando se estima por MC2E, en esta sección vamos a explicar como se calcula. Consideremos primero la primera ecuación estructural (Función de oferta de trabajo). El procedimiento es el siguiente:

- Estimar por MC2E y guardar los residuos que denotamos por *uhat1*.
- Realizamos la regresión de los residuos MC2E, *uhat1*, sobre todas las variables exógenas. Obtenemos los siguientes resultados:

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: uhat1

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	177,486	621,684	0,2855	0,7754
educ	0,357340	29,4878	0,0121	0,9903
age	−3,4410	10,5754	−0,3254	0,7451
kidslt6	−8,5740	173,057	−0,0495	0,9605
kidsge6	2,45400	54,4756	0,0450	0,9641
nwifeinc	0,896436	6,48356	0,1383	0,8901
exper	−15,884	26,1182	−0,6082	0,5434
expersq	0,640840	0,784734	0,8166	0,4146

Media de la var. dependiente	5,69865e-13
D.T. de la variable dependiente	1292,73
Suma de cuadrados de los residuos	7,12153e+08
Desviación típica de los residuos ($\hat{\sigma}$)	1302,15
R^2	0,00200507
$\bar{R}^2$ corregido	−0,0146282
$F(7, 420)$	0,120546
valor p para $F()$	0,996923
Log-verosimilitud	−3672,7
Criterio de información de Akaike	7361,57
Criterio de información Bayesiano de Schwarz	7394,05
Criterio de Hannan–Quinn	7374,40

El valor del estadístico de Sargan se calcula con el tamaño muestral $T = 428$ y el coeficiente de determinación $R^2 = 0,00200507$ de esta última regresión. El valor muestral obtenido $TR^2 = 2,94$ es menor que el valor crítico $\chi^2(3)_{0,05} = 7,81473$. Por lo tanto, no se rechaza las restricciones de sobreidentificación, que son 3 en este caso. Los instrumentos y la especificación de la segunda ecuación parece adecuada dada esta normalización.

¿Qué nos dirá este contraste si utilizamos la normalización sobre *hours* en la segunda ecuación y consideramos estimar la demanda de trabajo en lugar de la oferta salarial?

En este caso procedemos del siguiente modo:

1. Estimamos por MC2E y guardamos los residuos en *uhat2*.
2. Realizamos la regresión de estos residuos MC2E, *uhat2*, sobre todas las variables exógenas. El resultado es el siguiente:

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: uhat2

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	0,00407139	0,327084	0,0124	0,9901
kidslt6	−0,00671065	0,0910496	−0,0737	0,9413
kidsge6	−0,00601142	0,0286610	−0,2097	0,8340
age	−0,000358372	0,00556401	−0,0644	0,9487
educ	−0,00756862	0,0155143	−0,4878	0,6259
exper	0,000505048	0,0137414	0,0368	0,9707
nwifeinc	0,00562285	0,00341117	1,6484	0,1000
expersq	1,18111e-05	0,000412869	0,0286	0,9772

Media de la var. dependiente	0,000000
D.T. de la variable dependiente	0,681804
Suma de cuadrados de los residuos	197,130
Desviación típica de los residuos ($\hat{\sigma}$)	0,685096
R^2	0,00687110
$\bar{R}^2$ corregido	−0,00968105
$F(7, 420)$	0,415118
valor p para $F()$	0,892930
Log-verosimilitud	−441,40
Criterio de información de Akaike	898,800
Criterio de información Bayesiano de Schwarz	931,273
Criterio de Hannan–Quinn	911,625

El valor del estadístico de Sargan en este caso es $TR^2 = 8,14343$ que es mayor que el valor crítico $\chi^2(3)_{0,05} = 7,81473$. En este caso se rechaza la hipótesis nula, confirmando lo que indicaba el contraste de razón de verosimilitudes LR de información limitada también. Dada esta normalización, los instrumentos no son adecuados para *lwage*. No se pueden identificar adecuadamente los parámetros de la ecuación de demanda de trabajo.

6.2.6. Contraste de Hausman como un contraste basado en residuos. (Residual based test)

Ecuación de Oferta de Trabajo (primera ecuación estructural):

1. Se estima por MCO la ecuación de la Forma Reducida para lwage (posible endógena que está como explicativa en la ecuación estructural)

   Estimaciones MCO. Forma Reducida para lwage utilizando las 428 observaciones
   Variable dependiente: lwage

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,357997	0,318296	−1,1247	0,2613
age	−0,00352039	0,00541452	−0,6502	0,5159
educ	0,0998844	0,0150975	6,6160	0,0000
nwifeinc	0,00569422	0,00331952	1,7154	0,0870
kidslt6	−0,0558725	0,0886034	−0,6306	0,5287
kidsge6	−0,0176485	0,0278910	−0,6328	0,5272
exper	0,0407097	0,0133723	3,0443	0,0025
expersq	−0,000747326	0,000401777	−1,8601	0,0636

   y se guardan los residuos en esa ventana con la elección de Guardar → Residuos.

2. Estos residuos (denotamos por vhat1) se incluyen como regresor en la primera ecuación estructural junto con el resto de variables explicativas y se estima por MCO.

   Estimaciones MCO utilizando las 428 observaciones 1–428
   Variable dependiente: hours

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2432,20	331,416	7,3388	0,0000
lwage	1544,82	268,146	5,7611	0,0000
educ	−177,44	32,4307	−5,4716	0,0000
age	−10,784	5,34204	−2,0187	0,0442
kidslt6	−210,83	98,6899	−2,1363	0,0332
kidsge6	−47,557	31,7475	−1,4980	0,1349
nwifeinc	−9,2491	3,61503	−2,5585	0,0109
vhat1	−1623,6	273,362	−5,9394	0,0000

3. El contraste de Hausman es equivalente a contrastar si el coeficiente de ese residuo *vhat1* es o no significativamente distinto de cero. Si lo es se rechaza la hipótesis nula de que la variable *lwage* no estuviera correlacionada con el error de la ecuación estructural indicando evidencia de endogeneidad de esa variable.

   En este caso se rechaza la hipótesis nula ya que el coeficiente que acompaña al residuo *vhat1* es significativamente distinto de cero. Hay evidencia de endogeneidad de *lwage* en esta ecuación. Hay que notar que los coeficientes estimados del resto de variables coinciden con los de MC2E, no así las desviaciones típicas estimadas. Estas últimas no son adecuadas.

Ecuación de la Oferta Salarial (segunda ecuación estructural):

1. Se estima por MCO la ecuación de la Forma Reducida para *hours* (posible endógena que está como explicativa en la ecuación estructural) y se guardan los residuos *vhat2*.

Estimaciones MCO utilizando las 428 observaciones
Variable dependiente: hours

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	2056,64	346,484	5,9357	0,0000
kidslt6	−305,72	96,4501	−3,1697	0,0016
kidsge6	−72,366	30,3610	−2,3835	0,0176
age	−19,663	5,89403	−3,3362	0,0009
educ	−22,788	16,4345	−1,3866	0,1663
exper	47,0051	14,5565	3,2291	0,0013
nwifeinc	0,443852	3,61350	0,1228	0,9023
expersq	−0,513644	0,437358	−1,1744	0,2409

2. Estos residuos se incluyen como regresor en la ecuación estructural junto con el resto de variables explicativas y se estima por MCO. Los resultados son:

Estimaciones MCO utilizando las 428 observaciones
Variable dependiente: lwage

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,692790	0,297996	−2,3248	0,0206
hours	0,000160806	0,000209362	0,7681	0,4429
educ	0,111118	0,0149016	7,4568	0,0000
exper	0,0326460	0,0175542	1,8597	0,0636
expersq	−0,000676540	0,000430214	−1,5726	0,1166
vhat2	−0,000227292	0,000214095	−1,0616	0,2890

El coeficiente que acompaña al residuo *vhat2* no es significativamente distinto de cero. Luego no se rechazaría la hipótesis nula del contraste de Hausman. Parece que no hay evidencia de que hours sea endógena. De todas formas, también es cierto que, cuando se ha estimado la segunda ecuación estructural por cualquiera de los métodos considerados, el coeficiente que acompaña a *hours* no era significativamente distinto de cero. Esta puede ser la razón por la que el contraste de Hausman no rechaza la hipótesis nula. Simplemente esa variable no es significativa en la ecuación de oferta salarial.

Si consideramos la normalización sobre *hours* de la segunda ecuación y realizamos el contraste de Hausman:

1. Se estima por MCO la ecuación de la Forma Reducida para *lwage* y se guardan los residuos, que hemos denotado antes como *vhat1*:

2. Se estima la ecuación estructural para *hours* en función de *lwage* , el resto de variables explicativas y los residuos de la forma reducida para *lwage*:

   Los resultados que se obtienen son:

Estimaciones MCO utilizando las 428 observaciones 1–428
Variable dependiente: hours

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	1584,15	372,710	4,2504	0,0000
lwage	1000,53	577,222	1,7334	0,0838
educ	−130,10	63,9817	−2,0335	0,0426
exper	13,8850	28,0596	0,4948	0,6210
expersq	−0,0257319	0,638603	−0,0403	0,9679
vhat1	−1079,3	579,729	−1,8618	0,0633

En este caso el residuo si es significativo aunque no al 5 % sino al 1 %. Hay evidencia de que en la función de demanda de trabajo la variable *lwage* sea endógena y estuviera correlacionada con el error de esa ecuación estructural. Habría que instrumentalizar, pero vimos que los instrumentos *age*, *kidslt6*, *kidsge6* y *nwifeinc*, no son buenos para instrumentalizar a *lwage*. El contraste de Sargan ponía en entredicho la identificación de la ecuación.

6.3. Ejemplo 3: Modelo de Klein.

El Modelo I de KLEIN[3] es un modelo pionero y fué el primero de los elaborados por el profesor L. R. Klein, premio Nobel de Economía. Su objetivo fué estudiar la economía de los Estados Unidos en el periodo comprendido entre las dos guerras mundiales (1921-1941), y examinar las consecuencias de las distintas políticas económicas.

El modelo econométrico de Klein consta de seis ecuaciones: tres ecuaciones de comportamiento, sobre el consumo, la inversión y los salarios privados, y tres identidades:

$$
\begin{aligned}
C_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) + u_{1t} && (6.7)\\
I_t &= \delta_0 + \delta_1 P_t + \delta_2 P_{t-1} + \delta_3 K_{t-1} + u_{2t} && (6.8)\\
W_t^p &= \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + u_{3t} && (6.9)\\
X_t &= C_t + I_t + G_t \quad \texttt{Condición de equilibrio} && (6.10)\\
P_t &= X_t - T_t - W_t^p \quad \texttt{Identidad} && (6.11)\\
K_t &= K_{t-1} + I_t \quad \texttt{Identidad} && (6.12)
\end{aligned}
$$

[3]Klein, L., Economic Fluctuations in the United States 1921-1941. New York: Jhon Wiley and Sons, 1950.

Variables Endógenas	Variables Exógenas y Predeterminadas
C_t = Consumo privado en t	G_t = Gasto del gobierno en t (sin incluir salarios)
I_t = Inversión en t	T_t = Impuestos indirectos sobre las empresas más exportaciones netas
W_t^p = Salarios pagados pro el sector privado en t	W_t^g = Salarios pagados por el sector público
X_t = Demanda Total en t	X_{t-1} = Demanda Total en $t-1$
P_t = Beneficios Totales en t	P_{t-1} = Beneficios Totales en $t-1$
K_t = Stock de capital privado al final del año t	K_{t-1} = Stock de capital al final del año $t-1$
	A_t = Tendencia temporal medida en años a partir de 1931
	1 $\forall t$ Término constante

La ecuación del consumo descompone las causas del consumo en función de las diferentes fuentes de ingresos: salariales ($W^p + W^g$) y no salariales (P). Los hábitos de consumo serían diferentes según que los ingresos fueran vía salarios o vía beneficios. La segunda ecuación considera la inversión como función de los beneficios presentes (P_t) y pasados (P_{t-1}) a los que se añade el stock de capital con un retardo (K_{t-1}). Los parámetros de esta ecuación pueden interpretarse como coeficientes del beneficio o propensión a invertir.

La tercera ecuación de comportamiento es la de salarios. Esta explica la formación de los salarios privados W^p en función de la demanda total, tanto del periodo actual (X_t) como del año anterior (X_{t-1}). Se incluye, además, una tendencia temporal $A_t = t - 1931$ que está centrada en 1931, fecha que recogería el periodo de la Gran Depresión que comenzó en 1929 en Estados Unidos y que su final en este país se asocia con el final de la segunda Guerra Mundial a comienzos de 1939. Los salarios en este periodo se mantuvieron estables hasta 1930, pero comenzaron a bajar en 1931.

El modelo también se puede especificar con un número menor de ecuaciones, eliminando las tres variables endógenas, que se obtienen a partir de las identidades, y expresándolas en función de las endógenas restantes (desplazadas o no) y de todas las exógenas. Nosotros consideraremos las seis ecuaciones y estimaremos los parámetros desconocidos de las ecuaciones de comportamiento. Las identidades completan el sistema, ya que hay seis variables endógenas. Es un modelo dinámico ya que aparecen retardos de variables, por lo que también incorpora nuevos elementos al análisis de ecuaciones simultáneas.

En la versión 1.7.4 de Gretl , el archivo de muestra que está en la pestaña gretl, klein.gdt, incluye los datos de la variable stock de capital corriente K_t. Esto ha cambiado relativamente a versiones anteriores en las que este fichero incluía K_{t-1} pero no K_t. Por otro lado, al añadir retardos de las variables seleccionadas o definiendo los retardos con definir variable, estas variables quedan ocultas debajo de la variable corriente. Además al estimar por MCO o MC2E no aparecen en el listado de variables para poder elegirlas. Hay que proceder de otra forma, definiendo los retardos directamente en la ventana de estimación.

Veamos por ejemplo como estimar la primera ecuación del modelo de Klein 1, la función de Consumo. Primeramente definimos la variable $W = W^p + W^g$. Esta será una variable endógena ya que es la suma de una endógena (W^p) y una exógena (W^g). También definimos la variable

$A_t = t - 1931$ que aparece en el modelo como variable exógena por lo que será un instrumento. Para ambos cálculos utilizamos la opción *Variable → Definir nueva variable* y en la ventana que aparece escribimos en cada caso $W = Wp + Wg$ y $A = t - 1931$, respectivamente.

Por el momento no incorporamos el primer retardo de P, de K ni de X. Veremos como incluirlos cuando estimemos por MCO o por MC2E en sus ventanas correspondientes.

6.3.1. Estimación con Información Limitada

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) + u_{1t}$$

MCO: Elegir *Modelo → Mínimos Cuadrados Ordinarios* y en la ventana que aparece primeramente elegimos la variable dependiente C y las variables explicativas corrientes que aparecen en t , P y W. Posteriormente pinchamos la opción retardos que ahora aparecerá en negrita.

En la ventana que surge **orden de retardos** elegimos para la variable P retardos de 0 a 1. De esta forma incluye como regresores a P_t y a P_{t-1}. Si hubiéramos necesitado incluir retardos de alguna otra variable aunque no estuviera incluida también en t o corriente, deberíamos de haberla incluído antes en el listado de variables independientes pero no poner el 0 en el número de retardos.

Estimaciones MCO utilizando las 21 observaciones 1921–1941
Variable dependiente: C

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	16,2366	1,30270	12,4638	0,0000
P	0,192934	0,0912102	2,1153	0,0495
P_1	0,0898849	0,0906479	0,9916	0,3353
W	0,796219	0,0399439	19,9334	0,0000

Media de la var. dependiente	53,9952
D.T. de la variable dependiente	6,86087
Suma de cuadrados de los residuos	17,8794
Desviación típica de los residuos ($\hat{\sigma}$)	1,02554
R^2	0,981008
$\bar{R}^2$ corregido	0,977657
$F(3, 17)$	292,708
Estadístico de Durbin–Watson	1,36747
Coef. de autocorr. de primer orden.	0,246300
Log-verosimilitud	−28,108
Criterio de información de Akaike	64,2171
Criterio de información Bayesiano de Schwarz	68,3952
Criterio de Hannan–Quinn	65,1239

MC2E: Elegir *Modelo* → *Otros modelos lineales*→ *Mínimos cuadrados en dos etapas.* En la ventana **gretl: especificar modelo** elegir la variable dependiente C, las variables explicativas ya nos salen de antes. Sino tendríamos que hacer los mismo que antes incluir primero P y W y luego seleccionar en **retardos**, aquellos retardos que aparecen de estas variables. Si aparecen además en t incluir el 0 en los retardos.

Lo mismo ocurre con los instrumentos. En el listado incluímos todas las exógenas y para las predeterminadas incluímos las corrientes y luego en la ventanilla **retardos** seleccionamos los retardos que usaremos como instrumentos.

Por ejemplo en este caso, primero añadimos a la lista de instrumentos todas aquellas variables exógenas corrientes, (G, T, Wg, A) y también K P y X ya que tendremos que definir su primer retardo K_{t-1} P_{t-1} y X_{t-1} como instrumentos. Posteriormente, pinchamos en retardos y dejamos el 0 para las variables que como instrumentos están sin retardar (G, T, Wg, A) y elegimos en retardos específicos para P, X y K un retardo solamente. Esto nos incluye en la ventana **gretl: especificar modelo** los instrumentos P_{t-1}, X_{t-1}, K_{t-1} junto con G, T, Wg, A. Obtenemos los siguientes resultados:

Estimaciones MC2E utilizando las 21 observaciones 1921–1941
Variable dependiente: C
Instrumentos: const G T Wg A K_1 X_1 P_1

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	16,5548	1,46798	11,2772	0,0000
P	0,0173022	0,131205	0,1319	0,8951
P_1	0,216234	0,119222	1,8137	0,0697
W	0,810183	0,0447351	18,1107	0,0000

Media de la var. dependiente	53,9952
D.T. de la variable dependiente	6,86087
Suma de cuadrados de los residuos	21,9252
Desviación típica de los residuos ($\hat{\sigma}$)	1,13566
$F(3, 17)$	238,638
Estadístico de Durbin–Watson	1,48507
Coef. de autocorr. de primer orden.	0,204234
Criterio de información de Akaike	68,5009
Criterio de información Bayesiano de Schwarz	72,6790
Criterio de Hannan–Quinn	69,4076

Contraste de Hausman –
Hipótesis nula: Los estimadores de MCO son consistentes
Estadístico de contraste asintótico: $\chi^2_2 = 15{,}6891$
con valor p = 0,000391872

Contraste de sobreidentificación de Sargan –
Hipótesis nula: todos los instrumentos son válidos

Estadístico de contraste: $TR^2 = 8{,}77151$
con valor p = $P(\chi^2_4 > 8{,}77151) = 0{,}0670715$

Otra forma de estimar por los distintos métodos ya vistos era utilizar la opción

Modelo → Ecuaciones Simultáneas

Previamente hemos tenido que añadir los retardos de las variables que van a estar en el modelo tanto de las endógenas como de las exógenas.

Luego en la ventana que surge hay que especificar el sistema de ecuaciones. Con **equation** especificamos las ecuaciones de comportamiento y con **identity** las identidades. Además hay que especificar cuales son las endógenas, de esta forma gretl entiende que el resto de variables son los instrumentos. En este caso tenemos el siguiente sistema de instrucciones:

```
equation C 0 P P_1 W
equation I 0 P P_1 K_1
equation Wp 0 X X_1 A
identity X = C + I + G
identity P= X - T - Wp
identity K = K_1 + I
identity W=Wp+Wg
endog C I Wp P W X K
```

La tercera identidad sirve para identificar que, aunque W es endógena, también está en el modelo W^g que es exógena. Si no lo ponemos no utiliza W^g como instrumento.

Seleccionando el método de estimación se puede estimar por otros métodos de información limitada y completa (MCO, MC2E, MVIL, MC3E y MVIC). Además de poder estimar un sistema de ecuaciones aparentemente no relacionadas (SURE) pero esto sería indicado cuando no tenemos variables endógenas como explicativas.

Si se especifica todo el sistema y la lista de variables endógenas, no hace falta especificar la lista de instrumentos. La variable asociada al término constante se denota, además de con su nombre const, con su ID, que es el 0.

Otra forma equivalente de estimar el sistema sería

```
equation C 0 P P_1 W
equation I 0 P P_1 K_1
equation Wp 0 X X_1 A
identity X = C + I + G
identity P= X - T - Wp
identity K = K_1 + I
endog C I Wp P W X K
instr G T Wg X_1 P_1 K_1 A 0
```

Así no incluimos la identidad $W = W^p + W^g$ porque ya hemos introducido como instrumento en la lista a W^g.

Los resultados de la estimación de todas las ecuaciones del sistema por MC2E obtenidos de esta forma son los siguientes:

```
Sistema de ecuaciones, Mínimos cuadrados en dos etapas, periodo
1921-1941

Ecuación 1: Variable dependiente: C. Instrumentos: G T Wg X_1 P_1
K_1 A const

 VARIABLE        COEFICIENTE       DESV.TÍP.         ESTAD T   VALOR P
  const           16,5548          1,46798      11,277   <0,00001 ***
  P                0,0173022       0,131205      0,132   0,89509
  P_1              0,216234        0,119222      1,814   0,06972 *
  W                0,810183        0,0447351    18,111  <0,00001 ***

  Media de la var. dependiente = 53,9952
  Desviación típica de la var. dependiente. = 6,86087
  Suma de cuadrados de los residuos = 21,9252
  Desviación típica de los residuos = 1,13566

Ecuación 2: Variable dependiente: I. Instrumentos: G T Wg X_1 P_1
K_1 A const

 VARIABLE        COEFICIENTE       DESV.TÍP.     ESTAD T   VALOR P
  const           20,2782           8,38325       2,419     0,01557 **
  P                0,150222         0,192534      0,780     0,43525
  P_1              0,615944         0,180926      3,404     0,00066 ***
  K_1             -0,157788         0,040152     -3,930     0,00009 ***

  Media de la var. dependiente = 1,26667
  Desviación típica de la var. dependiente. = 3,55195
  Suma de cuadrados de los residuos = 29,0469
  Desviación típica de los residuos = 1,30715

Ecuación 3: Variable dependiente: Wp. Instrumentos: G T Wg X_1 P_1
K_1 A const

 VARIABLE        COEFICIENTE       DESV.TÍP.     ESTAD T   VALOR P
  const            1,50030          1,27569       1,176   0,23957
  X                0,438859         0,0396027    11,082  <0,00001 ***
  X_1              0,146674         0,0431639     3,398   0,00068 ***
  A                0,130396         0,0323884     4,026   0,00006 ***
```

```
Media de la var. dependiente = 36,3619
Desviación típica de la var. dependiente. = 6,3044
Suma de cuadrados de los residuos = 10,005
Desviación típica de los residuos = 0,767155
```

Esta forma de obtener las estimaciones por MC2E no muestra los contrastes de Hausman ni de Sargan que sí aparecen en el ouput al hacerlo con la opción *Modelo* → *Otros modelos lineales*.

A continuación mostramos los resultados de la estimación del Modelo de Klein por el resto de métodos: MVIL, MC3E y MVIC.

Sistema de ecuaciones, Máxima Verosimilitud con Información Limitada (MVIL)

```
Ecuación 1: Variable dependiente C.

 VARIABLE        COEFICIENTE     DESV.TÍP.     ESTAD T   VALOR P
  const           17,1477         2,04537        8,384  <0,00001 ***
  P               -0,222513       0,224230      -0,992   0,32103
  P_1              0,396027       0,192943       2,053   0,04012 **
  W                0,822559       0,0615494     13,364  <0,00001 ***

  Suma de cuadrados de los residuos = 40,8842
  Desviación típica de los residuos = 1,55079
  Log-verosimilitud = -132,419
  Smallest eigenvalue = 1,49875
  Contraste de sobreidentificación LR:
    Chi-cuadrado(4) = 8,4972 con valor p  0,0750

Ecuación 2: Variable dependiente I.

 VARIABLE        COEFICIENTE       DESV.TÍP.     ESTAD T   VALOR P
  const           22,5908           9,49815        2,378   0,01739 **
  P                0,0751848        0,224712       0,335   0,73794
  P_1              0,680386         0,209145       3,253   0,00114 ***
  K_1             -0,168264         0,0453445     -3,711   0,00021 ***

  Suma de cuadrados de los residuos = 34,9965
  Desviación típica de los residuos = 1,43479
  Log-verosimilitud = -121,054
  Smallest eigenvalue = 1,08595
  Contraste de sobreidentificación LR:
    Chi-cuadrado(4) = 1,73161 con valor p  0,7850
```

```
Ecuación 3: Variable dependiente Wp.

 VARIABLE        COEFICIENTE       DESV.TÍP.    ESTAD T   VALOR P
  const            1,52619          1,32084       1,155   0,24790
  X                0,433941         0,0755074     5,747  <0,00001 ***
  X_1              0,151321         0,0745268     2,030   0,04231 **
  A                0,131593         0,0359955     3,656   0,00026 ***

  Suma de cuadrados de los residuos = 10,0219
  Desviación típica de los residuos = 0,767805
  Log-verosimilitud = -136,891
  Smallest eigenvalue = 2,46858
  Contraste de sobreidentificación LR:
    Chi-cuadrado(4) = 18,9765 con valor p  0,0008
```

6.3.2. Estimación con Información Completa

Sistema de ecuaciones, Mínimos cuadrados en tres etapas, (MC3E)

```
Ecuación 1: Variable dependiente: C. Instrumentos: const P_1 K_1
X_1 A G T Wg

 VARIABLE        COEFICIENTE       DESV.TÍP.    ESTAD T   VALOR P
  const           16,4408          1,30455      12,603  <0,00001 ***
  P                0,124890        0,108129      1,155   0,24809
  P_1              0,163144        0,100438      1,624   0,10431
  W                0,790081        0,0379379    20,826  <0,00001 ***

  Suma de cuadrados de los residuos = 18,727
  Desviación típica de los residuos = 0,94433

Ecuación 2: Variable dependiente: I. Instrumentos: const P_1 K_1
X_1 A G T Wg

 VARIABLE        COEFICIENTE       DESV.TÍP.    ESTAD T   VALOR P
  const           28,1778           6,79377       4,148   0,00003 ***
  P               -0,0130792        0,161896     -0,081   0,93561
  P_1              0,755724         0,152933      4,942  <0,00001 ***
  K_1             -0,194848         0,0325307    -5,990  <0,00001 ***

  Suma de cuadrados de los residuos = 43,954
  Desviación típica de los residuos = 1,44674
```

```
Ecuación 3: Variable dependiente: Wp. Instrumentos: const P_1 K_1
X_1 A G T Wg

 VARIABLE       COEFICIENTE      DESV.TÍP.   ESTAD T   VALOR P
  const          1,79722         1,11585       1,611   0,10726
  X              0,400492        0,0318134    12,589  <0,00001 ***
  X_1            0,181291        0,0341588     5,307  <0,00001 ***
  A              0,149674        0,0279352     5,358  <0,00001 ***

  Suma de cuadrados de los residuos = 10,9206
  Desviación típica de los residuos = 0,721129

Contraste de sobreidentificación de Hansen-Sargan:
  Chi-cuadrado(12) = 24,291 con valor p  0,0186
```

Sistema de ecuaciones, Máxima Verosimilitud con Información Completa, (MVIC)

```
Se alcanzó la convergencia después de 35 iteraciones
Log-verosimilitud = -83,3238

Ecuación 1: estimaciones MVIC utilizando las 21 observaciones
1921-1941 Variable dependiente: C

 VARIABLE       COEFICIENTE      DESV.TÍP.    ESTAD T   VALOR P
  const          18,3433          2,48502       7,382  <0,00001 ***
  P              -0,232387        0,311955     -0,745   0,45631
  P_1             0,385672        0,217357      1,774   0,07600 *
  W               0,801844        0,0358931    22,340  <0,00001 ***

  Suma de cuadrados de los residuos = 44,1869
  Desviación típica de los residuos = 1,45057

Ecuación 2: estimaciones MVIC utilizando las 21 observaciones
1921-1941 Variable dependiente: I

 VARIABLE       COEFICIENTE      DESV.TÍP.     ESTAD T   VALOR P
  const          27,2638          7,93770       3,435   0,00059 ***
  P              -0,801003        0,491420     -1,630   0,10311
  P_1             1,05185         0,352459      2,984   0,00284 ***
  K_1            -0,148099        0,0298547    -4,961  <0,00001 ***

  Suma de cuadrados de los residuos = 268,201
  Desviación típica de los residuos = 3,57372
```

```
Ecuación 3: estimaciones MVIC utilizando las 21 observaciones
1921-1941 Variable dependiente: Wp

 VARIABLE        COEFICIENTE     DESV.TÍP.     ESTAD T   VALOR P
  const           5,79428          1,80442       3,211    0,00132 ***
  X               0,234118         0,0488180     4,796   <0,00001 ***
  X_1             0,284677         0,0452086     6,297   <0,00001 ***
  A               0,234835         0,0345002     6,807   <0,00001 ***

  Suma de cuadrados de los residuos = 37,8234
  Desviación típica de los residuos = 1,34206
```

En la siguiente sección compararemos los resultados obtenidos en la estimación de las ecuaciones de comportamiento estructurales, Conumo, Inversión y Salarios privados. También comentaremos las diferencias obtenidas y los contrastes de especificación.

6.3.3. Comparación de los resultados

Comenzamos por presentar los resultados de la estimación en varias tablas que nos ayudarán a compararlos más fácilmente. La siguiente tabla muestra los resultados obtenidos en la estimación de la función de consumo por los distintos métodos de estimación tanto de información limitada como de información completa:

Función de Consumo

Variable dependiente: Consumo C

Explicativas	MCO	MC2E	MVIL	MC3E	MVIC
const	$16,2^*$	$16,6^*$	$17,5^*$	$16,4^*$	$18,34^*$
	$(1,3)$	$(1,32)$	$(1,93)$	$(1,3)$	$(2,48)$
P_t	$0,193^*$	$0,017$	$-0,222$	$0,125$	$-0,232$
	$(0,091)$	$(0,118)$	$(0,212)$	$(0,108)$	$(0,096)$
P_{t-1}	$0,090$	$0,216$	$0,396^*$	$0,163$	$0,385$
	$(0,091)$	$(0,107)$	$(0,182)$	$(0,100)$	$(0,217)$
W_t	$0,796^*$	$0,810^*$	$0,823^*$	$0,790^*$	$0,801^*$
	$(0,040)$	$(0,040)$	$(0,058)$	$(0,033)$	$(0,035)$

$*$ significativo al 5 % $T = 21$ observaciones

Menor valor propio estimado = 1,49875

Contraste LR: Chi-cuadrado(4)= 8,4972 valor p = 0,075
Contraste de Sargan : Chi-cuadrado(4)= 8,77 valor p = 0,06707
Contraste de Hausman: Chi-cuadrado(2) = 15,6891 valor p = 0,000391

Se puede observara que el coeficiente que acompaña a P_t es positivo y significativo cuando se estima por MCO, pasando a no ser significativo en los métodos de MC2E,MVIL,MC3E y MVIC. Para el resto de variables se mantienen en general los signos y la significatividad. La variable más significativa parece ser los salarios W_t.

Los contrastes de sobreidentificación y exogeneidad rechazan en cada caso la hipótesis nula al 5 %, indicando que hay problemas de endogeneidad y que los instrumentos utilizados en la estimación de la ecuación pueden no ser muy adecuados.

En cuanto a los resultados de estimación de la función de inversión, la siguiente tabla muestra los resultados obtenidos:

Función de Inversión

Variable dependiente: Inversión I

Explicativas	MCO	MC2E	MVIL	MC3E	MVIC
const	$10,13$	$20,3^*$	$22,6^*$	$28,2^*$	$27,26^*$
	$(5,47)$	$(7,54)$	$(9,498)$	$(6,79)$	$(7,937)$
P_t	$0,48^*$	$0,15$	$0,075$	$-0,013$	$-0,801$
	$(0,097)$	$(0,193)$	$(0,224)$	$(0,162)$	$(0,491)$
P_{t-1}	$0,333^*$	$0,616^*$	$0,680^*$	$0,756^*$	$1,052^*$
	$(0,101)$	$(0,181)$	$(0,209)$	$(0,153)$	$(0,352)$
K_{t-1}	$-0,112^*$	$-0,158^*$	$-0,168^*$	$-0,195^*$	$-0,148^*$
	$(0,027)$	$(0,040)$	$(0,045)$	$(0,032)$	$(0,029)$

$*$ significativo al 5 % $T = 21$ observaciones

Menor valor propio estimado = 1,08595

Contraste LR: Chi-cuadrado(4)= 1,7316 valor p = 0,785
Contraste de Sargan : Chi-cuadrado(4)= 1,81497 valor p = 0,76973
Contraste de Hausman: Chi-cuadrado(1) = 21,3022 valor p = 0,0000039

Los resultados cambian en relación a la significatividad para la constante y P_t entre MCO y el resto de métodos. El resto de coeficientes se mantiene estable entre métodos. La variable no significativa es P_t siendo el resto significativas.

En cuanto a los contrastes de sobreidentificación, no se rechaza la hipótesis nula por lo que los instrumentos parecen adecuados para esta ecuación. El contraste de Hausman rechaza la hipótesis nula, lo que indica que si hay evidencia de endogeneidad de la variable P_t.

Por último, en la siguiente tabla se presentan los resultados obtenidos en la estimación de la ecuación de salarios por los distintos métodos:

Función de Salarios

Variable dependiente: Salarios Privados W^p

Explicativas	MCO	MC2E	MVIL	MC3E	MVIC
const	$1,497$	$1,5$	$1,53^*$	$1,797$	$5,79^*$
	$(1,27)$	$(1,27)$	$(1,32)$	$(1,115)$	$(1,804)$
X_t	$0,439^*$	$0,439^*$	$0,434^*$	$0,400^*$	$0,234^*$
	$(0,032)$	$(0,039)$	$(0,075)$	$(0,032)$	$(0,048)$
X_{t-1}	$0,146^*$	$0,147^*$	$0,151^*$	$0,181^*$	$0,284^*$
	$(0,037)$	$(0,043)$	$(0,074)$	$(0,034)$	$(0,045)$
A_t	$0,13^*$	$0,13^*$	$0,132^*$	$0,149^*$	$0,234^*$
	$(0,032)$	$(0,032)$	$(0,035)$	$(0,028)$	$(0,034)$

$*$ significativo al 5 % $\quad T = 21$ observaciones

Menor valor propio estimado = 2,46858

Contraste LR: Chi-cuadrado(4)= 18,97 valor p = 0,0008
Contraste de Sargan : Chi-cuadrado(4) = 12,495 valor p = 0,01402
Contraste de Hausman: Chi-cuadrado(1) = 0,0009103 valor = 0,97593

Contraste de Hansen-Sargan de restricciones de identificación en el sistema:
Chi-cuadrado(12) = 24,291 valor p = 0,0186

En este caso, los resultados no cambian a través de los diferentes métodos, excepto para el término constante. Todas las variables son significativas.

Los contrastes de sobreidentificación para esta ecuación rechazan la hipótesis nula indicando que los instrumentos no son adecuados, aunque el contraste de Hausman no rechaza la hipótesis nula. Parece no existir evidencia de endogeneidad de X_t en esa ecuación.

El contraste de sobreidentificación de todo el sistema, de Hansen-Sargan, rechaza las resticciones de sobreidentificación en el sistema. Esto indica que en conjunto, no parecen ser adecuados los instrumentos para estimar el sistema, aunque utilizando información limitada sí parecían ser adecuados para la segunda ecuación. En cambio la primera ecuación y la tercera no parecen estar bien especificadas, no siendo los instrumentos adecuados en ese caso.

BIBLIOGRAFÍA

[1] JHONSTON, J.(1984): Métodos de Econometría, Vicens-Vivens.

[2] JHONSTON, J. y DINARDO, J. (2001): Métodos de Econometría, Vicens-Vivens.

[3] GREENE, W.H. (1999): Análisis Econométrico, Prentice Hall, 3ª edición.

[4] SCHMIDT, P. (1976): Econometrics, Marcel Dekker, INC.

[5] WOOLDRIDGE, J.M. (2006): Introducción a la Econometría: Un enfoque moderno, Paraninfo-Thomson.