

## Nota de Clase 2

### Sistemas de Ecuaciones: Estimación

#### 1. Introducción

La estimación de un sistema de ecuaciones es el proceso mediante el cual es posible obtener “estimadores” consistentes y eficientes de los parámetros estructurales de las ecuaciones que conforman el sistema. Puesto de esta manera, el problema de estimación conlleva a que el investigador realice una serie de “preguntas” previas a los procesos generadores de datos que caracterizan al sistema.

La primera pregunta que se debe hacer el investigador está referida a la **identificación** del sistema. Solo en el caso de sistemas exactamente identificados o sobre-identificados será posible recuperar los parámetros de interés. Caso, contrario el investigador solo podrá estimar las formas reducidas del sistema, procedimiento que si bien puede ser útil para predecir las variables endógenas o establecer relaciones de estas con las exógenas del sistema resulta infructuoso para obtener una adecuada caracterización de la estructura económica bajo análisis. Como resolver esta pregunta ya fue discutido en la nota de clase 1.

La segunda pregunta surge una vez que se ha logrado identificar al sistema y tiene que ver con resolver los potenciales problemas de **estimación**. Así, el investigador estará interesado en evaluar la existencia o no de variables endógenas en el sistema ( $\text{cov}(u_i, u_j) \neq 0$ ) y la correlación entre los errores de las diferentes ecuaciones ( $E(u_i, u_j) = \sum_{g,xg}$ ). La presencia de tales condiciones en el sistema son las que guiarán la elección del investigador entre los métodos de estimación a ser discutidos en esta nota clase.

Hay que notar de la discusión previa que un procedimiento que resultaría atractivo para el investigador no familiarizado con las técnicas de estimación de los sistemas de ecuaciones es estimar las formas reducidas en sistema identificados mediante MCO y luego resolver las funciones  $f_i$  para hallar los parámetros estructurales. Sin embargo, este procedimiento puede resultar engorroso por lo que las técnicas de estimación que se presentan en esta nota de clase permiten evitar el uso de “papel y lápiz” al estimar directamente los parámetros estructurales. Más aún, si es que se pretende hacer inferencia sobre los parámetros usar “papel y lápiz” no sería el procedimiento más adecuado toda vez que las funciones  $f_i$  no solo deberán utilizarse para encontrar los parámetros estructurales sino además las desviaciones estándar de cada uno de ellos, procedimiento doblemente engorroso. Asimismo, como se mencionó previamente, cuando los errores de las ecuaciones estructurales se encuentran correlacionados emplear MCO en las formas reducidas puede resultar en un procedimiento menos eficiente que utilizar técnicas de estimación simultánea.

En pocas palabras, conocer las técnicas de estimación asociadas a los sistemas de ecuaciones permite ampliar el abanico de posibilidades de estimación del que se dispone con el objetivo de mejorar (posiblemente) la consistencia y eficiencia de los estimados. Al respecto, existen múltiples métodos de estimación que han sido desarrollados por la

literatura (ver Greene, 2003 y Wooldridge, 2002 para una discusión amplia de todas ellas). En esta nota de clase nos centraremos en cuatro métodos que son los más usados en las aplicaciones empíricas: MCO ecuación por ecuación, MC2E (mínimos cuadrados en dos etapas), MC3E (mínimos cuadrados en tres etapas) y MCG/SUR (mínimos cuadrados generalizados para ecuaciones aparentemente no relacionadas). Asimismo, se provee una discusión de cómo elegir entre cada uno de ellos.

## 2. Definiciones previas

Considérese un sistema de ecuaciones como<sup>1</sup>

$$\begin{aligned}
 \gamma_{11}y_1 + \gamma_{21}y_2 + \dots + \gamma_{g1}y_g + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{k1}x_k + u_1 &= 0 \\
 \gamma_{12}y_1 + \gamma_{22}y_2 + \dots + \gamma_{g2}y_g + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{k2}x_k + u_2 &= 0 \\
 \cdot & \\
 \cdot & \\
 \cdot & \\
 \gamma_{1g}y_1 + \gamma_{2g}y_2 + \dots + \gamma_{gg}y_g + \beta_{1g}x_1 + \beta_{2g}x_2 + \dots + \beta_{kg}x_k + u_g &= 0
 \end{aligned} \tag{1}$$

donde los parámetros  $\gamma_{ii}$  se normalizan a -1. Es decir, los parámetros de la variable  $y_i$  en la ecuación  $i$ . Asimismo, se define  $X_i$  como el vector de variables explicativas exógenas en la ecuación  $i$  e  $Y_i$  como el vector de variables explicativas endógenas en la ecuación  $i$ . Con estas dos condiciones finales se estaría especificando un sistema (más general que los anteriores) donde los regresores incluidos en cada una de las ecuaciones puede diferir. De este modo, el sistema puede ser expresado como<sup>2</sup>

$$\begin{aligned}
 y_1 &= \Gamma_1' Y_1 + B_1 X_1 + u_1 \\
 y_2 &= \Gamma_2' Y_2 + B_2 X_2 + u_2 \\
 \cdot & \\
 \cdot & \\
 y_g &= \Gamma_g' Y_g + B_g X_g + u_g
 \end{aligned} \tag{2}$$

que en términos matriciales toma la forma

---

<sup>1</sup> Nótese el cambio en el orden de las matrices para presentarlo en forma vectorial. Estos cambios no alteran las conclusiones a las que se llegaron en la nota de clase 1 y son realizados simplemente para facilitar las operaciones de álgebra matricial.

<sup>2</sup> Nótese que se han introducido las matrices  $\Gamma_i'$  que hacen referencia a aquellas matrices que acompañan a las endógenas incluidas en cada ecuación. Dicho de otro modo, son matrices de parámetros estructurales resultantes luego de normalizar los parámetros originales y despejar las variables endógenas por ser explicadas en cada ecuación.

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_g \end{bmatrix} = \begin{bmatrix} \Gamma_1' & 0 & \dots & 0 \\ 0 & \Gamma_2' & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Gamma_g' \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_g \end{bmatrix} + \begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & B_g \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_g \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_g \end{bmatrix}$$

o del mismo modo<sup>3</sup>,

$$y_{g \times 1} = \Gamma'_{g \times g} Y_{g \times 1} + B_{g \times g} X_{g \times 1} + U_{g \times 1} \quad (3)$$

Si se realiza el análisis sólo para la ecuación 1 (o para la ecuación  $i$  de modo general), se tiene que

$$[y_1] = [y_1 \ y_2 \ \dots \ y_g] \begin{bmatrix} 0 \\ \gamma_{21} \\ \dots \\ \gamma_{g1} \end{bmatrix} + [x_1 \ x_2 \ \dots \ x_k] \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \dots \\ \beta_{k1} \end{bmatrix} + [u_1]$$

Asumiendo que en esta ecuación particular del sistema se han incluido  $g-1$  regresores endógenos y  $k$  regresores exógenos. Nuevamente en forma matricial se tiene que

$$y_{1(n \times 1)} = Y_{1(n \times g)} \Gamma'_{1(g \times 1)} + X_{1(n \times k)} B_{1(k \times 1)} + u_{1(n \times 1)} \quad (4)$$

donde  $n$  representa al número de observaciones utilizadas en la estimación. Luego, se definen las matrices

$$W_{1[n \times (g+k)]} = [Y_1 \ X_1]$$

$$\Psi_{1[(g+k) \times 1]} = \begin{bmatrix} \Gamma_1' \\ B_1 \end{bmatrix}$$

y la ecuación queda definida de la siguiente manera

$$y_1 = W_1 \Psi_1 + u_1 \quad (5)$$

donde la matriz  $\Psi_1$  de parámetros estructurales es la que se intenta estimar, la cual contiene  $(g+k) - 1$  parámetros estructurales<sup>4</sup>. Realizar la estimación para la matriz  $\Psi_1$

<sup>3</sup> Nótese que cuando las matrices tienen el subíndice 1, los argumentos de cada una de ellas son escalares. Mientras tanto, cuando no está este subíndice los argumentos son sub-matrices. En otras palabras el sistema general (sin subíndice) está expresado en términos de matrices particionadas y el sistema individual (con subíndice) en términos de matrices "normales".

<sup>4</sup> Nótese que si es que el número de regresores endógenos y exógenos incluidos en cada una de las ecuaciones es el mismo, el número de parámetros estructurales por estimar es  $g(g+k-1)$ . Es decir, el número al que se llegó en la nota de clase 1

permitirá obtener los parámetros estructurales de forma directa sin recurrir a estimaciones de la forma reducida y utilizar luego “papel y lápiz”. Nótese además que la estimación de esta matriz (independientemente de la técnica utilizada) tendrá una matriz de varianza – covarianza asociada por lo que las desviaciones estándares de los estimadores podrán también ser identificados.

Si es que se define la relación (5) para cada una de las ecuaciones del sistema se obtendrá

$$y_{gxl} = W_{g \times g} \Psi_{g \times l} + U_{g \times l} \quad (6)$$

o lo que es lo mismo

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_g \end{bmatrix} = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & W_g \end{bmatrix} \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \dots \\ \Psi_g \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_g \end{bmatrix}$$

Paralelamente, definimos que cada error de cada una de las ecuaciones tiene su propia matriz de varianza-covarianza. Luego puede definirse

$$Var(U) = \Omega = \begin{bmatrix} Var(u_1) & cov(u_1, u_2) & \dots & cov(u_1, u_g) \\ cov(u_2, u_1) & Var(u_2) & \dots & cov(u_2, u_g) \\ \dots & & & \\ cov(u_g, u_1) & cov(u_g, u_2) & \dots & Var(u_g, u_g) \end{bmatrix} \quad (5)$$

donde en la diagonal principal se observan las matrices varianza-covarianza de los errores de cada ecuación y en las triangulares las covarianzas entre los errores de distintas ecuaciones.

### 3. Estimadores típicos de sistemas de ecuaciones

En base a las definiciones de la sección anterior, surgen los siguientes casos de estimación

a) Caso 1: MCO ecuación por ecuación

Sea  $cov(u_i, u_j) = 0$ ;  $Var(u_i) = \sigma_i^2 I$ ;  $W_i = [X_i]$ ; por lo que en un sistema como este no hay correlación entre los errores de diferentes ecuaciones y no hay variables endógenas en el sistema. De este modo, no existe ninguna fuente de simultaneidad entre las ecuaciones del sistema por ser estimado. Además, la varianza del error de cada ecuación es homocedástico.

Así, la matriz  $\Omega$  toma la siguiente forma

$$\text{Var}(U) = \Omega = \begin{bmatrix} \sigma_1^2 I_{n \times n} & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{n \times n} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_g^2 I_{n \times n} \end{bmatrix}$$

y el sistema por estimar es

$$y_1 = X_1 B_1 + u_1$$

$$y_2 = X_2 B_2 + u_2$$

.

.

$$y_g = X_g B_g + u_g$$

Cada una de las ecuaciones de este sistema cumplen las condiciones del modelo lineal general. Sin embargo, existe heterocedasticidad grupal. Por ello, el método de estimación por desarrollar deberá “corregir” este problema por lo que se plantea una estimación de MCG. En forma matricial

$$y_{g \times 1} = X_{g \times g} B_{g \times 1} + U_{g \times 1}$$

Luego se define el estimador

$$\hat{B}_{g \times 1} = (X_{g \times g}^T \Omega^{-1} X_{g \times g})^{-1} (X_{g \times g}^T \Omega^{-1} y_{g \times 1})$$

Para el caso en que  $g=2$

$$\begin{aligned} \hat{B} &= \left[ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 I & 0 \\ 0 & \sigma_2^2 I \end{pmatrix}^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right]^{-1} \left[ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 I & 0 \\ 0 & \sigma_2^2 I \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right] \\ &= \begin{bmatrix} (\sigma_1^2)^{-1} X_1^T I X_1 & 0 \\ 0 & (\sigma_2^2)^{-1} X_2^T I X_2 \end{bmatrix}^{-1} \begin{bmatrix} (\sigma_1^2)^{-1} X_1^T I y_1 \\ (\sigma_2^2)^{-1} X_2^T I y_2 \end{bmatrix} \\ &= \begin{bmatrix} (\sigma_1^2) & (\sigma_1^2)^{-1} (X_1^T X_1)^{-1} X_1^T y_1 \\ (\sigma_2^2) & (\sigma_2^2)^{-1} (X_2^T X_2)^{-1} X_2^T y_2 \end{bmatrix} \\ &= \begin{bmatrix} (X_1^T X_1)^{-1} X_1^T y_1 \\ (X_2^T X_2)^{-1} X_2^T y_2 \end{bmatrix} \end{aligned}$$

$$\hat{B} = \begin{bmatrix} \hat{B}_{1,MCO} \\ \hat{B}_{2,MCO} \end{bmatrix}$$

Se demuestra que bajo estas circunstancias es posible estimar por MCO cada una de las ecuaciones del sistema por separado. Tales estimadores ofrecerán resultados consistentes y eficientes (ver Greene, 2003).

b) Caso 2: MCG ecuación por ecuación

Sea  $cov(u_i, u_j) = 0$ ;  $Var(u_i) = \sigma_i^2 \Sigma_i$ ;  $W_i = [X_i]$ ; por lo que en un sistema como este no hay correlación entre los errores de diferentes ecuaciones y no hay variables endógenas en el sistema. De este modo, no existe ninguna fuente de simultaneidad entre las ecuaciones del sistema por ser estimado. Sin embargo, difiere del caso anterior en la medida que los errores de las ecuaciones individuales presentan heterocedasticidad.

La matriz  $\Omega$  toma la siguiente forma

$$Var(U) = \Omega = \begin{bmatrix} \sigma_1^2 \Sigma_{1(n,xn)} & 0 & \dots & 0 \\ 0 & \sigma_2^2 \Sigma_{2(n,xn)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_g^2 \Sigma_{g(n,xn)} \end{bmatrix}$$

y el sistema por estimar es

$$\begin{aligned} y_1 &= B_1 X_1 + u_1 \\ y_2 &= B_2 X_2 + u_2 \\ &\cdot \\ &\cdot \\ y_g &= B_g X_g + u_g \end{aligned}$$

Nótese que en este caso existe tanto heterocedasticidad grupal como individual. El método de estimación por desarrollar deberá “corregir” ambos problemas. Para ello, se plantea un estimador de MCG similar al del caso anterior. En forma decir,

$$\hat{B}_{g,x1} = (X_{g,xg}^T \hat{\Omega}_{g,xg}^{-1} X_{g,xg})^{-1} (X_{g,xg}^T \hat{\Omega}_{g,xg}^{-1} y_{g,x1})$$

Para el caso en que  $g=2$

$$\hat{B} = \left[ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 \Sigma_1 & 0 \\ 0 & \sigma_2^2 \Sigma_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right]^{-1} \left[ \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 \Sigma_1 & 0 \\ 0 & \sigma_2^2 \Sigma_2 \end{pmatrix}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right]$$

$$\begin{aligned}
&= \begin{bmatrix} (\sigma_1^2)^{-1} X_1^T \Sigma_1^{-1} X_1 & 0 \\ 0 & (\sigma_2^2)^{-1} X_2^T \Sigma_2^{-1} X_2 \end{bmatrix}^{-1} \begin{bmatrix} (\sigma_1^2)^{-1} X_1^T \Sigma_1^{-1} y_1 \\ (\sigma_2^2)^{-1} X_2^T \Sigma_2^{-1} y_2 \end{bmatrix} \\
&= \begin{bmatrix} (\sigma_1^2) & (\sigma_1^2)^{-1} (X_1^T \Sigma_1^{-1} X_1)^{-1} (X_1^T \Sigma_1^{-1} y_1) \\ (\sigma_2^2) & (\sigma_2^2)^{-1} (X_2^T \Sigma_2^{-1} X_2)^{-1} (X_2^T \Sigma_2^{-1} y_2) \end{bmatrix} \\
&= \begin{bmatrix} (X_1^T \Sigma_1^{-1} X_1)^{-1} (X_1^T \Sigma_1^{-1} y_1) \\ (X_2^T \Sigma_2^{-1} X_2)^{-1} (X_2^T \Sigma_2^{-1} y_2) \end{bmatrix} \\
\hat{B} &= \begin{bmatrix} \hat{B}_{1,MCG} \\ \hat{B}_{2,MCG} \end{bmatrix}
\end{aligned}$$

Se demuestra que bajo estas circunstancias es posible estimar por MCG cada una de las ecuaciones del sistema por separado. En términos prácticos, el procedimiento estándar es estimar los parámetros con MCO corrigiéndolos por heterocedasticidad (por ejemplo, utilizando el procedimiento de White/Huber) si es que en planteamiento original se demuestra la existencia de esta perturbación no esférica (por ejemplo, mediante el desarrollo de pruebas de hipótesis sobre la constancia de la varianza del error)<sup>5</sup>. La matriz de corrección que utiliza este procedimiento es muy similar al caso estándar de MCG factibles. De este modo, la estimación de White/Huber no es otra cosa que la matriz de errores cuadráticos que se obtiene de la estimación MCO.

$$\hat{\Sigma}_i = \begin{bmatrix} e_{i1}^2 & 0 & \dots & 0 \\ 0 & e_{i2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e_{in}^2 \end{bmatrix}$$

Los estimadores resultantes luego de aplicar MCG (factibles) serán consistentes y eficientes (ver Greene, 2003).

A partir del caso 1 y caso 2 ya se puede ensayar una conclusión: “cuando no haya relaciones de simultaneidad entre las ecuaciones no es necesario realizar una estimación del sistema; basta con estimar utilizando MCO ecuación por ecuación o por MCG factibles (cuando haya que corregir por heterocedasticidad individual) para obtener estimadores consistentes y eficientes”.

c) Caso 3: MC2E

---

<sup>5</sup> Por ejemplo, algunas pruebas de hipótesis disponibles son el LR test, la prueba de Spearman, la prueba de Goldfeld y Quandt, la prueba de Glejser, la prueba de Park, la prueba de Harvey, el test de White, la prueba de Breusch-Pagan. El lector no familiarizado con estas pruebas puede revisar Castro y Rivas-Losa (2003).

Sea  $\text{cov}(u_i, u_j) = 0$ ;  $\text{Var}(u_i) = \sigma_i^2 I$ ;  $W_i = [Y_i \ X_i]$  ; por lo que en un sistema como este no hay correlación entre los errores de diferentes ecuaciones y los errores de cada ecuación individual son homocedásticos. Sin embargo, existe una fuente de simultaneidad entre las ecuaciones al existir variables endógenas al lado derecho de la ecuación.

La matriz  $\Omega$  toma la siguiente forma

$$\text{Var}(U) = \Omega = \begin{bmatrix} \sigma_1^2 I_{1(nxn)} & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{2(nxn)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_g^2 I_{g(nxn)} \end{bmatrix}$$

Conocemos que el modelo toma la forma

$$y_1 = \Gamma_1' Y_1 + B_1 X_1 + u_1$$

$$y_2 = \Gamma_2' Y_2 + B_2 X_2 + u_2$$

.

.

$$y_g = \Gamma_g' Y_g + B_g X_g + u_g$$

o en términos matriciales

$$y_{gxl} = \Gamma_{g \times g}' Y_{gxl} + B_{g \times g} X_{gxl} + U_{gxl}$$

de este modo para la ecuación 1, se puede demostrar que

$$y_{1(nxl)} = Y_{1(n \times g)} \Gamma_{1(g \times 1)}' + X_{1(n \times k)} B_{1(k \times 1)} + u_{1(n \times 1)}$$

Sin embargo, dado que existen regresores estocásticos al lado derecho de la ecuación, resultará útil expresarlo en términos de su forma reducida. Para ello, conviene recuperar el modelo original (antes de la normalización).

$$\Gamma_{g \times g}' Y_{gxl} + B_{g \times g} X_{gxl} + U_{gxl} = 0_{g \times 1}$$

y expresarlo en su forma reducida

$$Y_{gxl} = \Pi_{g \times g} X_{gxl} + V_{gxl}$$

que para la ecuación 1, toma la forma

$$y_{1(n \times l)} = X_{1(n \times k)} \Pi_{1(k \times l)} + v_{1(n \times l)}$$

La forma reducida ya puede ser estimada mediante MCO para hallar los parámetros  $\Pi_1$  de modo que

$$\hat{\Pi}_1 = (X_1^T X_1)^{-1} (X_1^T y_1)$$

Luego se procede a hallar las proyecciones  $\hat{y}_1$

$$\hat{y}_1 = X_1 \hat{\Pi}_1$$

Este procedimiento se puede repetir para cada una de las variables  $y_i$  de modo que se construye la matriz  $\hat{Y}_1$ . Así, reemplazando en la ecuación original

$$y_{1(n \times l)} = \hat{Y}_{1(n \times g)} \Gamma'_{1(g \times l)} + X_{1(n \times k)} B_{1(k \times l)} + u_{1(n \times l)}$$

esta expresión ya estaría en términos de variables exógenas por lo que se procede a construir

$$\hat{W}_{1[n \times (g+k)]} = [\hat{Y}_1 \quad X_1]$$

$$\Psi_{1[(g+k) \times l]} = \begin{bmatrix} \Gamma'_1 \\ B_1 \end{bmatrix}$$

y la ecuación queda definida de la siguiente manera

$$y_1 = \hat{W}_1 \Psi_1 + u_1$$

replicando este procedimiento para cada una de las ecuaciones del sistema, se puede hallar

$$y_{g \times l} = \hat{W}_{g \times g} \Psi_{g \times l} + U_{g \times l}$$

De este modo, el estimador MC2E queda definido como

$$\hat{\Psi}_{MC2E} = (\hat{W}^T \hat{W})^{-1} (\hat{W}^T y)$$

Para el caso en que  $g=2$

$$\hat{\Psi}_{1,MC2E} = (\hat{W}_1^T \hat{W}_1)^{-1} (\hat{W}_1^T y_1)$$

$$\hat{\Psi}_{1,MC2E} = \left[ \begin{pmatrix} \hat{Y}_1 & X_1 \end{pmatrix}^T \begin{pmatrix} \hat{Y}_1 & X_1 \end{pmatrix} \right]^{-1} \left[ \begin{pmatrix} \hat{Y}_1 & X_1 \end{pmatrix}^T y_1 \right]$$

$$\hat{\Psi}_{1,MC2E} = \left[ \begin{pmatrix} \hat{Y}_1^T \\ X_1^T \end{pmatrix} (\hat{Y}_1 \ X_1) \right]^{-1} \left[ \begin{pmatrix} \hat{Y}_1^T \\ X_1^T \end{pmatrix} y_1 \right]$$

$$\hat{\Psi}_{1,MC2E} = \begin{bmatrix} \hat{Y}_1^T \hat{Y}_1 & \hat{Y}_1^T X_1 \\ X_1^T \hat{Y}_1 & X_1^T X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1^T y_1 \\ X_1^T y_1 \end{bmatrix}$$

Por extensión

$$\hat{\Psi}_{2,MC2E} = \begin{bmatrix} \hat{Y}_2^T \hat{Y}_2 & \hat{Y}_2^T X_2 \\ X_2^T \hat{Y}_2 & X_2^T X_2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_2^T y_2 \\ X_2^T y_2 \end{bmatrix}$$

De este modo

$$\hat{\Psi}_{MC2E} = \begin{bmatrix} \hat{\Psi}_{1,MC2E} \\ \hat{\Psi}_{2,MC2E} \end{bmatrix}$$

El estimador de MC2E resultará aquel eficiente y consistente bajo las condiciones iniciales del modelo expuesto (ver Greene, 2003).

En resumen, el estimador MC2E toma su nombre debido a que su cálculo resulta de realizar dos etapas en el proceso de estimación.

En la primera etapa, se instrumentalizan las variables endógenas del modelo. Esta instrumentalización es muy similar al procedimiento de variables instrumentales (VI) para modelos uniecuacionales. Para ello, el modelo multi-ecuacional debe presentarse en forma reducida; es decir, las variables endógenas deben presentarse en términos de las variables exógenas. De este modo, cada una de las endógenas del sistema estará expresada en términos de TODAS las exógenas. Es importante notar en este paso que originalmente NO TODAS las variables exógenas del sistema habían sido incluidas en las ecuaciones estructurales individuales. Por ello, la forma reducida permitirá que se cumplan las condiciones de sobreidentificación, necesarias para la estimación de VI<sup>6</sup>. Dadas estas condiciones de sobreidentificación, se podrá estimar por MCO la forma reducida y obtener las proyecciones de las variables endógenas. A cada una de estas proyecciones se les conoce como variables endógenas “exogeneizadas”.

En la segunda etapa los regresores estocásticos de las ecuaciones estructurales son reemplazados por las nuevas variables exogeneizadas. Dado que ahora todo el sistema está en término de regresores no estocásticos, puede aplicarse nuevamente MCO para hallar el estimador final. La estimación conjunta de la etapa 1 y 2 es lo que se conoce como los estimadores MCO/VI o en este contexto el MC2E.

Al igual que en el caso de MCO/VI, en la aplicación de MC2E deben cumplirse los supuestos de relevancia y exogeneidad de instrumentos. Es decir, se deben realizar las

---

<sup>6</sup> Por ejemplo, las variables exógenas incluidas en la ecuación 2 pero omitidas en la ecuación 1 actuarán como los “instrumentos” del procedimiento estándar.

pruebas de hipótesis previas necesarias para comprobar la validez del procedimiento. En el primer caso, se recomienda el uso de la “rule of thumb” de Stock y Watson (2003) y en el segundo caso una prueba de sobreidentificación a lo Sargan o el test J. Una explicación detallada de cómo realizar estas pruebas se explica en Stock y Watson (2003). Asimismo, como se explica en Greene (2003) utilizar la prueba de Hausman para verificar la exogeneidad de las variables es siempre un procedimiento útil.

Una propiedad interesante de este estimador es que puede ser estimado para una ecuación sin tomar en cuenta que sucede con el resto de ecuaciones (Gujarati, 2001). Por ejemplo, si es que es posible expresar una de la ecuaciones en función de sus endógenas y exógenas y se contienen suficientes instrumentos para “instrumentalizar” las endógenas presentes en el sistema la aplicación de este estimador de VI será consistente y eficiente para los parámetros de la ecuación de interés. Sin embargo, los parámetros estructurales del resto de ecuaciones no podrían ser estimados.

Finalmente, es útil notar que ante la presencia de heterocedasticidad en las ecuaciones individuales es posible realizar una corrección por MCG sin alterar el procedimiento descrito anteriormente. La alternativa a la mano es corregir las matrices de varianza covarianza mediante el estimador robusto de White/Huber.

d) Caso 4: SUR, regresiones aparentemente no correlacionadas

Sea  $\text{cov}(u_i, u_j) = \sigma_{ij}I$ ;  $\text{Var}(u_i) = \sigma_i^2 I$ ;  $W_i = [X_i]$ . En este sistema los errores de cada una de las ecuaciones individuales son homocedásticos. Asimismo, aunque no existen variables endógenas como explicativas, como fuente de simultaneidad se identifica una correlación entre los errores de las ecuaciones estructurales. Justamente por esta razón es que el estimador se llama SUR (regresión aparentemente no correlacionada, por sus siglas en inglés): la fuente de simultaneidad no es tan evidente como en el caso anterior.

La matriz  $\Omega$  toma la siguiente forma

$$\text{Var}(U) = \Omega = \begin{bmatrix} \sigma_1^2 I & \sigma_{12} I & \dots & \sigma_{1g} I \\ \sigma_{21} I & \sigma_2^2 I & \dots & \sigma_{2g} I \\ \dots & \dots & \dots & \dots \\ \sigma_{g1} I & \sigma_{g2} I & \dots & \sigma_g^2 I \end{bmatrix}$$

Luego, el sistema por estimar toma la siguiente forma

$$\begin{aligned} y_1 &= X_1 B_1 + u_1 \\ y_2 &= X_2 B_2 + u_2 \\ &\cdot \\ &\cdot \\ y_g &= X_g B_g + u_g \end{aligned}$$

Cada una de las ecuaciones de este sistema cumplen las condiciones del modelo lineal general y al igual que en el caso 1, existe heterocedasticidad grupal. Por ello, de modo similar a este caso se plantea una estimación por MCG como estrategia para superar este problema. En forma matricial el sistema es

$$y_{gxl} = X_{g \times g} B_{g \times l} + U_{g \times l}$$

Luego se define el estimador

$$\hat{B}_{SUR} = (X_{g \times g}^T \Omega_{g \times g}^{-1} X_{g \times g})^{-1} (X_{g \times g}^T \Omega_{g \times g}^{-1} y_{g \times l})$$

Sin embargo, en la práctica la matriz  $\Omega$  es desconocida por lo que se requiere estimar. El procedimiento utilizado en el contexto SUR es muy similar a la corrección por White/Huber ya discutida en el caso 2. Así, se plantea la siguiente matriz

$$\hat{\Omega} = \begin{bmatrix} e_1^2/n & e_{12}/n & \dots & e_{1g}/n \\ e_{21}/n & e_2^2/n & \dots & e_{2g}/n \\ \dots & \dots & \dots & \dots \\ e_{g1}/n & e_{g2}/n & \dots & e_g^2/n \end{bmatrix}$$

que hace referencia a la matriz de varianzas – covarianzas de los errores de estimados de cada una de las ecuaciones individuales. Por ello, en base a esta característica es que normalmente el estimador SUR también se define como un estimador en dos etapas.

En la primera, se estiman cada una de las ecuaciones individuales por MCO (nótese que el procedimiento es posible ya que no hay regresores estocásticos). En base a estos resultados se computa la matriz  $\hat{\Omega}$ . Así, para el sistema se tiene que

$$\begin{aligned} y_1 - X_1 B_{1,MCO} &= e_1 \\ y_2 - X_2 B_{2,MCO} &= e_2 \\ &\cdot \\ &\cdot \\ y_g - X_g B_{g,MCO} &= e_g \end{aligned}$$

donde  $\hat{B}_{i,MCO} = (X_i^T X_i)^{-1} (X_i^T y_i)$ , lo que permite construir  $\hat{\Omega}$

En la segunda etapa se calcula un estimador de MCG incorporando la matriz  $\hat{\Omega}$ . Al realizar de manera conjunta ambas etapas permite obtener estimador  $\hat{B}_{SUR}$

$$\hat{B}_{SUR} = (X_{g \times g}^T \hat{\Omega}_{g \times g}^{-1} X_{g \times g})^{-1} (X_{g \times g}^T \hat{\Omega}_{g \times g}^{-1} y_{g \times l})$$

Sin embargo, ¿cuándo es necesario realizar una estimación SUR? La respuesta es bastante obvia: “cuando se demuestre la existencia de correlación entre los errores”. Para llegar a esa conclusión, se dispone de la siguiente prueba de hipótesis (Breusch – Pagan).

$$\lambda = n \sum_i \sum_j r_{ij}^2 \sim \chi_{\frac{g(g-1)}{2}}^2$$

donde

$r_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$  es el coeficiente de correlación entre los errores estimados por MCO de las ecuaciones  $i$  y  $j$ .

Luego, siendo

$H_0 : \Omega$  es diagonal

$H_1 : \Omega$  no es diagonal

Se pretende verificar

$$\lambda > \chi_{\frac{g(g-1)}{2}}^2 \rightarrow SUR$$

$$\lambda < \chi_{\frac{g(g-1)}{2}}^2 \rightarrow MCO$$

Solo al rechazar la  $H_0$  (o verificar que la matriz  $\hat{\Omega}$  es diagonal) se aplica el estimador SUR que resulta consistente y eficiente (ver Greene, 2003).

Es importante notar tres características de este procedimiento. Primero, si es que no existe correlación entre los errores, el estimador colapsa a un estimador MCO ecuación por ecuación. Segundo, si es que las variables exógenas incluidas en cada una de las ecuaciones individuales son las mismas, los estimadores SUR y MCO son idénticos (incluso en valor) por lo que la estimación SUR también puede ser reemplazada por la estimación MCO ecuación por ecuación en este caso. Tercero, si es que los errores de las ecuaciones individuales son heterocedásticos, el estimador SUR debe extenderse para incorporar una estimación por MCG en la primera etapa (en lugar de MCO). Este nuevo estimador MCG/SUR será aquel eficiente y consistente (ver Wooldridge, 2002 y Greene, 2003 para las demostraciones de estos argumentos).

Vale la pena ampliar que es lo que ocurre en este último caso. Ahora las características del modelo son  $cov(u_i, u_j) = \sigma_{ij}I$ ;  $Var(u_i) = \sigma_i^2 \Sigma_i$ ;  $W_i = [X_i]$

La matriz  $\Omega$  toma la siguiente forma

$$Var(U) = \Omega = \begin{bmatrix} \sigma_1^2 \Sigma_1 & \sigma_{12} I & \dots & \sigma_{1g} I \\ \sigma_{21} I & \sigma_2^2 \Sigma_2 & \dots & \sigma_{2g} I \\ \dots & \dots & \dots & \dots \\ \sigma_{g1} I & \sigma_{g2} I & \dots & \sigma_g^2 \Sigma_g \end{bmatrix}$$

En este caso, el estimador tomará una forma bastante similar

$$\hat{B}_{MCG/SUR} = (X_{g \times g}^T \Omega^{-1} X_{g \times g})^{-1} (X_{g \times g}^T \Omega^{-1} y_{g \times 1})$$

y las diferencias provienen al momento de estimar  $\hat{\Omega}$

$$\hat{\Omega} = \begin{bmatrix} e_1^2 / n \hat{\Sigma}_1 & e_{12} / n & \dots & e_{1g} / n \\ e_{21} / n & e_2^2 / n \hat{\Sigma}_2 & \dots & e_{2g} / n \\ \dots & \dots & \dots & \dots \\ e_{g1} / n & e_{g2} / n & \dots & e_g^2 / n \hat{\Sigma}_g \end{bmatrix}$$

donde

$$\hat{\Sigma}_i = \begin{bmatrix} e_{i1}^2 & 0 & \dots & 0 \\ 0 & e_{i2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e_{in}^2 \end{bmatrix}$$

que es la matriz que se utilizó para corregir por heterocedasticidad en el caso 2.

e) Caso 5: MC3E

Sea  $cov(u_i, u_j) = \sigma_{ij} I$ ;  $Var(u_i) = \sigma_i^2 I$ ;  $W_i = [Y_i \ X_i]$ . En este sistema los errores de cada una de las ecuaciones individuales son homocedásticos. Asimismo, existen dos fuentes de simultaneidad: presencia de variables endógenas como explicativas y correlación entre los errores de las ecuaciones estructurales. El estimador de MC3E propone una metodología para “corregir” por ambas fuentes de manera conjunta. Como se intuye, este estimador combinará los procedimientos del MC2E y del modelo SUR. Justamente, esta es la naturaleza de las etapas que definen al procedimiento: las dos primeras similares al MC2E y la tercera similar al SUR.

La matriz  $\Omega$  toma la siguiente forma

$$\text{Var}(U) = \Omega = \begin{bmatrix} \sigma_1^2 I & \sigma_{12} I \dots & \sigma_{1g} I \\ \sigma_{21} I & \sigma_2^2 I \dots & \sigma_{2g} I \\ \dots & & \\ \sigma_{g1} I & \sigma_{g2} I \dots & \sigma_g^2 I \end{bmatrix}$$

Conocemos que el modelo toma la forma

$$y_1 = \Gamma_1' Y_1 + B_1 X_1 + u_1$$

$$y_2 = \Gamma_2' Y_2 + B_2 X_2 + u_2$$

.

.

$$y_g = \Gamma_g' Y_g + B_g X_g + u_g$$

Luego, haciendo las mismas transformaciones que en el caso 3, se obtiene la forma estructural del sistema

$$\Gamma_{g \times g} Y_{g \times 1} + B_{g \times g} X_{g \times 1} + U_{g \times 1} = 0_{g \times 1}$$

que en su forma reducida es

$$Y_{g \times 1} = \Pi_{g \times g} X_{g \times 1} + V_{g \times 1}$$

y para la ecuación 1, toma la forma

$$y_{1(n \times 1)} = X_{1(n \times k)} \Pi_{1(k \times 1)} + v_{1(n \times 1)}$$

Así, la **primera etapa** de la estimación consiste en estimar mediante MCO los parámetros  $\Pi_1$  de modo que

$$\hat{\Pi}_1 = (X_1' X_1)^{-1} (X_1' y_1)$$

y se procede a hallar las proyecciones  $\hat{y}_1$

$$\hat{y}_1 = X_1 \hat{\Pi}_1$$

procedimiento que se repite para cada una de las variables  $y_i$  de modo que se construye la matriz  $\hat{Y}_1$ .

En la **segunda etapa**, se reemplazan las expresiones  $\hat{Y}_1$  por las  $Y_1$  en la ecuación original y se obtiene

$$y_{1(nxl)} = \hat{Y}_{1(nxg)} \Gamma'_{1(gxl)} + X_{1(nxk)} B_{1(kxl)} + u_{1(nxl)}$$

esta expresión ya estaría en términos de variables exógenas por lo que se procede a construir

$$\hat{W}_{1[nx(g+k)]} = [\hat{Y}_1 \ X_1]$$

$$\Psi_{1[(g+k)x1]} = \begin{bmatrix} \Gamma'_1 \\ B_1 \end{bmatrix}$$

y la ecuación queda definida de la siguiente manera

$$y_1 = \hat{W}_1 \Psi_1 + u_1$$

replicando este procedimiento para cada una de las ecuaciones del sistema, se puede hallar

$$y_{gxl} = \hat{W}_{gxl} \Psi_{gxl} + U_{gxl}$$

En base a este modelo el estimador MC2E queda definido como

$$\hat{\Psi}_{MC2E} = (\hat{W}^T \hat{W})^{-1} (\hat{W}^T y)$$

Nótese que estas dos primeras etapas son justamente las que se siguieron para hallar el estimador MC2E.

Finalmente, para desarrollar la **tercera etapa**, debe entenderse que el estimador objetivo tomará la forma

$$\hat{\Psi}_{MC3E} = (\hat{W}^T \hat{\Omega}^{-1} \hat{W})^{-1} (\hat{W}^T \hat{\Omega}^{-1} y)$$

debido a la correlación entre los errores de las diferentes ecuaciones individuales. Así, el objetivo es obtener la matriz  $\hat{\Omega}$  que en este caso se define como

$$\hat{\Omega} = \begin{bmatrix} e_{1,MC2E}^2 / n & e_{12,MC2E} / n \dots & e_{1g,MC2E} / n \\ e_{21,MC2E} / n & e_{2,MC2E}^2 / n \dots & e_{2g,MC2E} / n \\ \dots & & \\ e_{g1,MC2E} / n & e_{g2,MC2E} / n \dots & e_{g,MC2E}^2 / n \end{bmatrix}$$

que hace referencia a la matriz de varianzas – covarianzas de los errores de estimados de cada una de las ecuaciones individuales. Sin embargo, en este caso la estimación no se ha realizado por MCO sino por MC2E. Así,

$$\begin{aligned}
y_1 - W_1 \hat{\Psi}_{1,MC2E} &= e_{1,MC2E} \\
y_2 - W_2 \hat{\Psi}_{2,MC2E} &= e_{2,MC2E} \\
&\cdot \\
&\cdot \\
y_g - W_g \hat{\Psi}_{g,MC2E} &= e_{g,MC2E}
\end{aligned}$$

donde  $\hat{\Psi}_{MC2E} = (\hat{W}^T \hat{W})^{-1} (\hat{W}^T y)$ , lo que permite construir  $\hat{\Omega}$ . Un vez que se obtiene la matriz es posible ahora si construir el estimador de interés

$$\hat{\Psi}_{MC3E} = (X_{g \times g}^T \hat{\Omega}_{g \times g}^{-1} X_{g \times g})^{-1} (X_{g \times g}^T \hat{\Omega}_{g \times g}^{-1} y_{g \times 1})$$

Este estimador será consistente y eficiente de acuerdo con las condiciones iniciales del modelo (Greene, 2003).

Además, deberán tomarse en cuenta todas las consideraciones previamente discutidas en el resto de modelos. Primero, que se cumplan las condiciones de identificación, relevancia de instrumentos y exogeneidad de los mismos al momento de aplicar el estimador de MC2E. Segundo, que efectivamente exista correlación entre los errores, para lo cual se puede utilizar la misma prueba de hipótesis que se explicó en el caso 4. Tercero, es igualmente posible extender este estimador para el caso en que los errores de las ecuaciones individuales sean heterocedástico siguiendo un procedimiento similar al explicado en el caso 4.

#### 4. Prueba de hipótesis

Las pruebas de hipótesis sobre los parámetros tienen la misma estructura que en el caso uniecuacional. Así, puede definirse por ejemplo

$$\begin{aligned}
H_o : \delta_2 + \delta_3 + \delta_4 &= 0 \\
\delta_2 &= \delta_4 \\
\delta_3 &= \delta_1
\end{aligned}$$

$H_1 : alguna no se cumple$

en términos matriciales estas restricciones toman la forma

$$\begin{aligned}
\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
R \delta &= r
\end{aligned}$$

de forma general se definen las restricciones como

$$R_{q \times k} \delta_{k \times 1} = r_{q \times 1}$$

donde  $R$  es la matriz de restricciones,  $\delta$  es la matriz de parámetros,  $r$  es la matriz de constantes,  $q$  el número de restricciones que se imponen o hipótesis que se intentan testear y  $k$  el número de regresores en la estimación.

En base a estas definiciones, la prueba de hipótesis en términos matriciales toma la forma para un sistema de  $g$  ecuaciones tomará la forma

$$F = \frac{(R\hat{\delta} - r)^T [R(X^T \hat{\Omega} X)^{-1} R^T]^{-1} (R\hat{\delta} - r)}{q} \sim F_{[q, g(n-k)]}$$

simplemente bastaría con hacer la estimación, recuperar las varianzas covarianzas de los estimadores y realizar la siguiente comparación

$$F > F_{[q, g(n-k)]} \rightarrow \text{no acepto } H_0$$

$$F < F_{[q, g(n-k)]} \rightarrow \text{no rechazo } H_0$$

La utilidad de expresar la prueba de hipótesis de este modo que puede observarse que la estimación de ecuaciones simultáneas permiten realizar pruebas sobre parámetros en diferentes ecuaciones. Ello es de suma utilidad en la medida que en muchas aplicaciones económicas es de interés realizar inferencias sobre dos ecuaciones estructurales dentro de un mismo sistema. Nótese que la matriz  $\hat{\Omega}$  es la que permite identificar una matriz de varianzas – covarianzas para todo el set de regresores (en diferentes ecuaciones), por lo que será posible realizar las pruebas de hipótesis.