

Modelos de ecuaciones simultáneas

Gustavo A. García

ggarci24@eafit.edu.co

Econometría II

Programa de Economía

Universidad EAFIT

Link slides formato **html**

Link slides formato **PDF**

En este tema

- Motivación
- Sesgo de simultaneidad en MCO
- Identificación
- Prueba de simultaneidad
- Estimación
- El modelo aparentemente no relacionado (SUR)
- Ejercicio aplicado en R: comportamiento del gasto público de US

Lecturas

- Wooldridge, J. (2013). *Introducción a la econometría. Un enfoque moderno*. 5a edición. Cengage Learning [Cap 16](#)
- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. [Cap. 18, 19 y 20](#)
- Griffiths, W., Carter, R. y Judge, G. (1993). *Learning and practicing econometrics*. John & Sons, Inc. [Cap. 17](#)
- Baltagi, B. (2011). *Econometrics*. Fifth Edition, Springer. [Cap 10](#)

Motivación

- En el tema anterior se mostró que el método de variables instrumentales puede resolver dos tipos de problemas: variables omitidas y error de medición
- Otra forma importante de endogeneidad de las variables explicativas es la **simultaneidad**
- **Simultaneidad**: ocurre cuando una o más de las variables explicativas se **determinan conjuntamente** con la variable dependiente
- Es necesario entonces estudiar los métodos para estimar modelos de ecuaciones simultáneas (MES)

Motivación

El ejemplo clásico de un MES es una ecuación de oferta y demanda para algún bien o insumo de producción. Sea, por ejemplo, h_s las horas anuales de trabajo y w el salario por hora. Una función simple de oferta de mano de obra es

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1$$

z_1 : variable que afecta la oferta de mano de obra (por ejemplo, el salario del sector en la región analizada)

u_1 : término de error, contiene otros factores que afectan la oferta de mano de obra

α_1 : mide cómo cambia la oferta de mano de obra cuando el salario cambia. Si h_s y w están en logaritmos, representa la elasticidad de la mano de obra

La anterior ecuación representa una **ecuación estructura**. Este nombre proviene del hecho de que la función de oferta de mano de obra es **derivable de la teoría económica y tiene una interpretación causal**

En términos económicos, esta elasticidad es importante ya que permite determinar cómo cambiará el número de horas que los trabajadores desean trabajar cuando las tasas impositiva sobre el ingreso salarial cambian

Motivación

La ecuación de oferta de mano de obra parece similar a las vistas antes, pero presenta una diferencia: **el salario no varía exogenamente**

Se debe entender que existe una interacción entre oferta y demanda de mano de obra y de acuerdo a la teoría, **se iguala esta oferta y demanda, determinando los valores de los salarios y las horas trabajadas en equilibrio**

Para describir cómo se determinan los salarios y las horas en equilibrio, es necesario introducir la función de demanda de mano de obra:

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2$$

h_d : horas demandadas (demanda de mano de obra)

z_2 : puede ser la extensión de tierra

La anterior ecuación también es una ecuación estructural, ya que proviene de la teoría: maximización de utilidad de los empleadores

Se tiene que:

- Oferta de mano de obra: ecuación del comportamiento de los trabajadores
- Demanda de mano de obra: ecuación del comportamiento de los empleadores

Motivación

En el análisis econométrico, las dos ecuaciones se vinculan sólo debido a que el salario y las horas observadas se determinan por la intercepción entre la oferta y la demanda. En otras palabras, para cada región i , las horas observadas h_i y el salario observado w_i están determinados por la condición de equilibrio

$$h_{is} = h_{id}$$

Dado a que se observan sólo las horas laboradas del equilibrio, entonces las horas observadas se denotan con h_i

Al combinar la condición de equilibrio con las ecuaciones de oferta y demanda se obtiene el sistema:

$$\begin{aligned}h_i &= \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \\h_i &= \alpha_2 w_i + \beta_2 z_{i2} + u_{i2}\end{aligned}$$

Estas dos ecuaciones constituyen un **modelo de ecuaciones simultáneas (MES)**

Características del MES:

- h_i y w_i son **variables endógenas**. Las dos ecuaciones determinan estas dos variables
- z_{i1} y z_{i2} son variables exógenas al determinarse fuera del modelo. **Algo importante sobre estas variables es que si no son diferentes, las ecuaciones de oferta y demanda serán las mismas y el sistema no se podrá identificar**

Motivación

Otro ejemplo de MES: índice de homicidios y tamaño de la fuerza policial

Cuánta más fuerza policial los índices de crime disminuyen. Esta relación se puede escribir como:

$$murdpc = \beta_{10} + \alpha_1 polpc + \beta_{11} incpc + u_1$$

murdpc: homicidios

polpc: número de policias

incpc: ingreso (se asume exógeno)

La pregunta que surge es: ¿es el número de policías es exógena? Probablemente no. **El gasto en policias está determinado, al menos parcialmente, por su nivel de crimen.** Esto implica la siguiente ecuación:

$$polpc = \beta_{20} + \alpha_2 murdpc + otros factores$$

Aunque sólo interesa la primera ecuación, es necesaria saber con precisión como se especifica la segunda con el fin de estimar la primera. Surge entonces un MES

Sesgo de simultaneidad en MCO

Cuando una variable explicativa se determina en forma simultánea con la variable dependiente, dicha variable explicativa estará correlacionada con el término de error, con lo cual **las estimaciones por MCO serán sesgadas e inconsistentes**

Consideremos el siguiente modelo estructural de dos ecuaciones

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \quad (1)$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2 \quad (2)$$

z_1 y z_2 son exógenas (no se correlacionan con los errores)

Para mostrar que y_2 se correlaciona con u_1 , se resuelven las dos ecuaciones para y_2 en términos de las variables exógenas y el término de error

Insertando (1) en (2) y despejando y_2 se obtiene:

$$y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$$

$$(1 - \alpha_2 \alpha_1) y_2 = \alpha_2 \beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2$$

Sesgo de simultaneidad en MCO

Suponiendo que $\alpha_2\alpha_1 \neq 1$ para que se puede resolver para y_2 y dividiendo por $1 - \alpha_2\alpha_1$, se tiene:

$$y_2 = \pi_{21}z_1 + \pi_{22}z_2 + \nu_2 \quad (3)$$

$$\pi_{21} = \alpha_2\beta_1/(1 - \alpha_2\alpha_1), \pi_{22} = \beta_2/(1 - \alpha_2\alpha_1) \text{ y } \nu_2 = (\alpha_2u_1 + u_2)/(1 - \alpha_2\alpha_1)$$

y_2 está expresada en términos de las variables exógenas y los términos de error, y representa la **ecuación de la forma reducida** para y_2

π_{21} y π_{22} representan los **parámetros de la forma reducida** y están en función, no lineal, de los **parámetros estructurales**. El **error de la forma reducida** ν_2 es una función lineal de los términos de error estructurales u_1 y u_2

Recordemos que el interés principal es estimar consistentemente la ecuación (1) a partir de los MCO y esto sólo se lograría si y_2 y u_1 no están correlacionados. Sin embargo, como muestra la ecuación (3), y_2 y u_1 están correlacionados, ya que ν_2 es una función lineal de u_1 y u_2

Cuando y_2 está correlacionado con u_1 debido a la simultaneidad, se dice que MCO sufre de **sesgo de simultaneidad**

Identificación

Sabemos que MCO es sesgado e inconsistente cuando se aplica a una ecuación estructural en un sistema de ecuaciones

Del tema anterior se explicó que el método de MC2E se puede utilizar para resolver el problema de variables explicativas endógenas

La pregunta que surge al aplicar MC2E a un MES, teniendo en cuenta que se tiene una ecuación estructural para cada variable endógena, es si hay suficientes variables instrumentales para estimar el sistema \implies si el sistema se encuentra identificado

Identificación

Cuando se estima un modelo mediante MCO, la condición clave de identificación es que ninguna variable explicativa esté correlacionada con el término de error

No obstante, si se tiene algunas variables instrumentales, aún se pueden identificar (o estimar de manera consistente) los parámetros en una ecuación de MES, tal como con las variables omitidas o el error de medición

Supongamos que tenemos un sistema de oferta y demanda en equilibrio (esto es, $q_s = q_d = q$)

$$q = \alpha_1 p + \beta_1 z_1 + u_1 \text{ oferta}$$

$$q = \alpha_2 p + u_2 \text{ demanda}$$

La pregunta entonces es ¿cuál de estas dos ecuaciones está identificada? La ecuación de demanda está identificada, mientras que la ecuación de oferta no

La idea es que se puede emplear z_1 (esta variable se supone exógena a las ecuaciones de oferta y demanda) como VI para el precio en la ecuación de demanda. Pero ya que z_1 aparece en la ecuación de oferta, no se tiene VI alguna para el precio en la ecuación de oferta

Identificación

En forma general para un sistema de dos ecuaciones:

$$\begin{aligned}y_1 &= \beta_{10} + \alpha_1 y_2 + \mathbf{z}_1 \boldsymbol{\beta}_1 + u_1 \\y_2 &= \beta_{20} + \alpha_2 y_1 + \mathbf{z}_2 \boldsymbol{\beta}_2 + u_2\end{aligned}$$

y_1 y y_2 son variables endógenas

$\mathbf{z}_1 = (z_{11}, z_{12}, \dots, z_{1k_1})$ y $\mathbf{z}_2 = (z_{21}, z_{22}, \dots, z_{2k_2})$ conjuntos de variables exógenas

El hecho de que \mathbf{z}_1 y \mathbf{z}_2 contengan, por lo general, diferentes variables exógenas significa que se tienen **restricciones de exclusión** en el modelo \implies variables exógenas que aparecen en la primera ecuación pero no en la segunda y viceversa

La pregunta, entonces, ahora es si es posible estimar los parámetros, por ejemplo, de la primera ecuación? Las **condiciones de rango y orden para la identificación nos ayuda a responder esta pregunta**

Condición de rango: La primera ecuación en un modelo de dos ecuaciones simultáneas se identifica si, y sólo si, la segunda ecuación contiene al menos una variable exógena (con un coeficiente diferente de cero) excluida de la primera

Condición de orden: para identificar la primera ecuación es necesario que al menos una variable exógena se excluya de esta ecuación

Identificación

La condición de rango requiere más que la condición de orden, ya que es necesario que al menos una variable exógena excluida de la primera ecuación debe tener un coeficiente diferente de cero en la segunda ecuación \implies empíricamente esto se puede verificar mediante una prueba t o F

Otros ejemplos:

Oferta de mano de obra de las mujeres casadas

$$\begin{aligned} \text{hours} &= \beta_{10} + \alpha_1 \log(\text{wage}) + \beta_{11} \text{educ} + \beta_{12} \text{age} + \beta_{13} \text{kidslt6} + \beta_{14} \text{nwifeinc} + u_1 \\ \log(\text{wage}) &= \beta_{20} + \alpha_2 \text{hours} + \beta_{21} \text{educ} + \beta_{22} \text{exper} + \beta_{23} \text{exper}^2 + u_2 \end{aligned}$$

La primera ecuación satisface la condición de orden, ya que las dos variables exógenas exper y exper^2 se omiten de esta ecuación

La segunda ecuación de salarios, se identifica si al menos una de las variables age , kidslt6 o nwifeinc tiene coeficiente diferente de cero en la primera ecuación

Inflación y apertura

$$\begin{aligned} \text{inf} &= \beta_{10} + \alpha_1 \text{open} + \beta_{11} \log(\text{pcinc}) + u_1 \\ \text{open} &= \beta_{20} + \alpha_2 \text{inf} + \beta_{21} \log(\text{pcinc}) + \beta_{22} \log(\text{land}) + u_2 \end{aligned}$$

Bajo la condición de rango la primera ecuación está identificada siempre y cuando $\beta_{22} \neq 0$, mientras que la segunda no. Pero lo que interesa es la primera ecuación, así que el sistema se podría estimar para determinar los efectos de la apertura sobre la inflación

Identificación

En términos generales para un sistema de M ecuaciones las condiciones de orden y rango pueden expresarse de la siguiente forma, teniendo en cuenta la siguiente notación:

M = número de variables endógenas en el modelo

m = número de variables endógenas en una ecuación dada

K = número de variables exógenas en el modelo (incluyendo el intercepto)

k = número de variables exógenas en una ecuación dada

Condición de orden (condición necesaria pero no suficiente)

En un modelo de M ecuaciones simultáneas, para que una ecuación esté identificada, el número de variables exógenas excluidas de esa ecuación no debe ser menor que el número de variables endógenas incluidas en la ecuación menos 1, es decir:

$$K - k \geq m - 1$$

Si $K - k = m - 1$, la ecuación está exactamente identificada, pero si $K - k > m - 1$, estará sobreidentificada

Condición de rango (condición necesaria y suficiente)

En un modelo que contiene M ecuaciones con M variables endógenas, una ecuación está identificada si y sólo si puede construirse por lo menos un determinante diferente de cero, de orden $(M - 1)(M - 1)$, a partir de los coeficientes de las variables (endógenas y exógenas) excluidas de esa ecuación particular, pero incluidas en las otras ecuaciones del modelo

Identificación

Consideremos el siguiente ejemplo. Sea el siguiente sistema de ecuaciones donde Y son las variables endógenas y X las exógenas:

$$Y_{1t} - \beta_{10} - \beta_{12}Y_{2t} - \beta_{13}Y_{3t} - \gamma_{11}X_{1t} = u_{1t} \quad (1)$$

$$Y_{2t} - \beta_{20} - \beta_{23}Y_{3t} - \gamma_{21}X_{1t} - \gamma_{22}X_{2t} = u_{2t} \quad (2)$$

$$Y_{3t} - \beta_{30} - \beta_{31}Y_{1t} - \gamma_{31}X_{1t} - \gamma_{32}X_{2t} = u_{3t} \quad (3)$$

$$Y_{4t} - \beta_{40} - \beta_{41}Y_{1t} - \beta_{42}Y_{2t} - \gamma_{43}X_{3t} = u_{4t} \quad (4)$$

Para facilitar la identificación se construyen las siguientes tablas:

Ecuación	1	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
(1)	$-\beta_{10}$	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
(2)	$-\beta_{20}$	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
(3)	$-\beta_{30}$	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
(4)	$-\beta_{40}$	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

Condición de orden:

Ecuación	# de exógenas excluidas (K-k)	# de endógenas incluidas menos 1 (m-1)	¿Identificada?
(1)	2	2	Exactamente
(2)	1	1	Exactamente
(3)	1	1	Exactamente
(4)	2	2	Exactamente

Identificación

Se observa que desde la condición de orden todas las ecuaciones se encuentran identificadas. Ahora es necesario aplicar la condición de rango para determinar en efecto que cada ecuación esté identificada

Consideremos la primera ecuación. Para que esta ecuación esté identificada, se debe obtener por lo menos un determinante diferente de cero de orden 3×3 , a partir de los coeficientes de las variables excluidas de esta ecuación, pero incluidas en otras: Y_4 , X_2 y X_3 . Se tiene entonces la siguiente matriz:

$$A = \begin{bmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{bmatrix}$$

El determinante de A es cero, por lo que su rango es menor que 3, con lo cual **no se satisface la condición de rango y por tanto no está identificada**. Queda para demostrar la condición de rango para las otras ecuaciones y determinar si están o no identificadas (ayuda: la ecuación 2 y 3 no están identificadas, mientras que la 4 si lo está)

En resumen, **la condición de rango dice si la ecuación bajo análisis está identificada o no, en tanto que la condición de orden expresa si dicha ecuación está exactamente identificada o sobreidentificada**

Prueba de simultaneidad

Si existe simultaneidad, los estimadores MCO serán inconsistentes, por lo que un primer análisis puede ser verificar la presencia de simultaneidad

En una prueba de simultaneidad se intenta averiguar si una variable explicativa está correlacionada con el término de error. En este caso se utilizará la prueba de error de especificación de Hausman

Analicemos esta prueba con el siguiente modelo del comportamiento del gasto gubernamental de Estados Unidos. Se tiene el siguiente modelo de ecuaciones simultáneas:

$$EXP = \beta_1 + \beta_2 AID + \beta_3 INC + \beta_4 POP + u_i$$

$$AID = \delta_1 + \delta_2 EXP + \delta_3 PS + v_i$$

EXP = gasto público de los gobiernos estatales y locales (endógena); AID = nivel de ayuda mediante subsidio federal (endógena); INC = ingreso de los estados (exógena); POP = población estatal (exógena); PS = población estudiantil de primaria y secundaria (exógena); u y v términos de error

Aquí el interés es estimar el efecto de las ayudas federales sobre el gasto público estatal, es decir, que la ecuación principal es la primera. Sin embargo, es necesario corroborar la existencia de simultaneidad entre EXP y AID

El test de Hausman comprende los siguientes pasos:

- **Paso 1:** Efectúe la regresión de la forma reducida de AID y obtenga los residuales de esta regresión (\hat{w}_i)
- **Paso 2:** Efectúe la regresión de EXP sobre AID , INC , POP y \hat{w}_i , y realice una prueba t sobre el coeficiente de \hat{w}_i . Si éste es significativo, no se rechaza la hipótesis de simultaneidad

Estimación

Mínimos Cuadrados Indirectos (MCI)

Cuando una ecuación estructural se encuentra exactamente identificada, el método para obtener las estimaciones de los coeficientes estructurales a partir de las estimaciones por MCO de los coeficientes en forma reducida se conoce como **Método de Mínimos Cuadrados Indirectos (MCI)**

MCI comprende tres pasos:

- **Paso 1:** Se obtienen las ecuaciones en la forma reducida
- **Paso 2:** Se aplica MCO individualmente a las ecuaciones de la forma reducida. Esta operación es permisible puesto que las variables explicativas en estas ecuaciones son exógenas y, por tanto, no están correlacionadas con las perturbaciones estocásticas. Las estimaciones así obtenidas son consistentes
- **Paso 3:** Se obtienen estimaciones de los coeficientes estructurales originales a partir de los coeficientes en forma reducida estimados

Limitaciones

- Las ecuaciones deben estar exactamente identificadas
- En algunos casos no son posible los MCI, y en otros conducen a varias estimaciones

Estimación

Mínimos Cuadrados en Dos Etapas (MC2E)

- Los MC2E proporciona un procedimiento de estimación muy útil para obtener los valores de los parámetros estructurales en [ecuaciones sobreidentificadas](#)
- La estimación de MC2E utiliza información disponible de la especificación de un sistema de ecuaciones para obtener una estimación única para cada parámetro estructural
- Intuición del método: la primera etapa implica la creación de un instrumento, mientras que la segunda etapa implica una variante de la estimación de variables instrumentales

Estimación

Mínimos Cuadrados en Dos Etapas (MC2E)

Consideremos el siguiente ejemplo de oferta y demanda para explicar el método

Modelo estructural

$$\begin{aligned} q &= \alpha_2 p + \epsilon \text{ oferta} \\ q &= \beta_2 p + \beta_3 y + \beta_4 w + u \text{ demanda} \end{aligned}$$

Forma reducida

$$\begin{aligned} q &= \pi_{12} y + \pi_{13} w + v_1 \\ p &= \pi_{22} y + \pi_{23} w + v_2 \end{aligned}$$

Se observa que la ecuación de oferta se encuentra sobreidentificada, por lo que los MCI no llevarán a estimaciones únicas de los parámetros. La ecuación de demanda no se encuentra identificada, pero asumamos que el parámetro α_2 de la ecuación de oferta es el de interés

Otro proceso de estimación de α_2 es el de variables instrumentales. Sin embargo, si elegimos éste enfoque, tendríamos que elegir entre dos estimadores del parámetro α_2 : el primero utilizará y como instrumento y el segundo usaría w

Dado que ambos estimadores son consistentes, es necesario un criterio para elegir entre los dos. Un procedimiento razonable (y eficiente) implica elegir como un instrumento un promedio ponderado de las dos variables exógenas, eligiendo los pesos para maximizar la correlación entre el instrumento nuevo y p

Para obtener este instrumento, se realiza la regresión de p sobre y y w , y se calculan los valores ajustados \hat{p} que sería el instrumento. Luego se obtiene el estimador de α_2 :

$$\hat{\alpha}_2 = \frac{\sum \hat{p} q}{\sum \hat{p} p}$$

Estimación

Mínimos Cuadrados en Dos Etapas (MC2E)

De manera formal, MC2E funciona de la siguiente forma:

- **Etapa 1:** la ecuación de la forma reducida para p se estima usando MCO, esto es, la regresión de la variable endógena sobre todas las variables exógenas del sistema. A partir de esta regresión se determina los valores ajustados de la variable endógena (\hat{p})

Los valores ajustados por construcción serán independientes de los términos de error ϵ y u . El proceso de la primera etapa permite construir una variable que se relaciona en forma lineal con la variable explicativa que es endógena y es depurada de cualquier correlación con el término de error en la ecuación de la oferta

- **Etapa 2:** Se estima la ecuación estructural de interés (ecuación de oferta) reemplazando la variable endógena (p) con la variable ajustada calculada en la primera etapa (\hat{p})

El uso de MCO en esta segunda etapa producirá un estimador consistente del parámetro de oferta α_2 , pero es necesario ajustar los errores estándar de las estimaciones por la inclusión de variables predichas en el modelo

MC2E elimina el problema de excesivos instrumentos con el uso de combinaciones de las variables exógenas para crear un nuevo instrumento

Cuando una ecuación es identificada exactamente, la estimación por MC2E es idéntica a los MCI y a la de VI

Cuando una ecuación no está identificada, MC2E no es posible

Estimación

Mínimos Cuadrados en Dos Etapas (MC2E)

Limitaciones

- MC2E genera estimaciones ineficientes ya que sólo se aplican a una ecuación dentro del sistema de ecuaciones. El método toma en cuenta el hecho de que una o más variables exógenas son omitidas de la ecuación que se va a estimar, pero no se toma en cuenta el hecho de que también puede haber variables exógenas omitidas de otras ecuaciones
- Una fuente alternativa de ineficiencia surge debido a que la estimación de una sola ecuación no explica la correlación de ecuación cruzada entre errores

Estimación

Mínimos Cuadrados en Tres Etapas (MC3E)

Zellner y Theil (1962) consideran la ineficiencia de los MC2E e implementan el mecanismo denominado Mínimos Cuadrados en Tres Etapas

MC3E es más eficiente asintóticamente que el estimador por MC2E

MC3E implica la aplicación de estimación de mínimos cuadrados generalizados a un sistema de ecuaciones, cada una de las cuales se ha estimado primero usando MC2E

MC3E funciona de la siguiente forma:

- **Etapa 1:** Se estima la forma reducida del sistema de modelos. Se calculan los valores ajustados de las variables endógenas y se usan luego para obtener estimaciones MC2E de todas las ecuaciones en el sistema
- **Etapa 2:** Una vez que se han calculado los parámetros MC2E, se usan los residuales de cada ecuación para estimar las varianzas y covarianzas de ecuaciones cruzadas
- **Etapa 3:** Se obtienen las estimaciones de parámetros de mínimos cuadrados generalizados

Consideraciones

- La aplicación de la tercera etapa del procedimiento no alterará las estimaciones MC2E en el caso especial en el que todas las covarianzas de ecuaciones cruzadas son 0
- Cualquier ecuación que no esté identificada debe eliminarse del sistema de ecuaciones antes de se apliquen los MC3E

El modelo aparentemente no relacionado (SUR)

El modelo aparentemente no relacionado (SUR, *seemingly unrelated model*) consiste de una serie de variables endógenas que son consideradas como un grupo debido a que tienen una relación conceptual estrecha entre sí

Como ejemplo, consideremos 5 firmas de acero: GM, Chrysler, GE, Westinghouse y US steel. Los datos consisten de series de tiempo de 20 años para las siguientes variables:

I_{it} = inversión bruta

F_{it} = valor de mercado de la firma al final del año previo

C_{it} = valor de las existencias de planta y equipo de la firma al final del año previo

donde $i = 1, \dots, 5$, $t = 1935, \dots, 1954$. En cualquier momento t se pueden escribir las 5 ecuaciones:

$$I_{1t} = \beta_{10} + \beta_{11}F_{1t} + \beta_{12}C_{1t} + \epsilon_{1t}$$

.....

$$I_{5t} = \beta_{50} + \beta_{51}F_{5t} + \beta_{52}C_{5t} + \epsilon_{5t}$$

Cada ecuación puede ser estimada individualmente tomando $t = 1935, \dots, 1954$ (cada firma lleva a cabo su propia política de inversión, por tanto se puede tratar estas ecuaciones como no relacionadas)

Por otro lado, todas las actividades económicas toman lugar en el mismo ambiente económico, por tanto puede esperarse que los 5 shocks $\epsilon_{1t}, \dots, \epsilon_{5t}$ estén correlacionados

El modelo aparentemente no relacionado (SUR)

El procedimiento que toma en cuenta la correlación entre los errores en un sistema de ecuaciones, es llamado SUR

Generalmente, éste método de estimación difiere de los MCO excepto en dos caso:

- las ecuaciones son realmente no correlacionadas, es decir, $Cov(\epsilon_{it}, \epsilon_{js}) = 0, t \neq s$
- todas las ecuaciones en el sistema tienen la mismas variables explicatorias, es decir, $F_{1t} = \dots = F_{5t} = F_t$

Usando SUR para estimar conjuntamente las ecuaciones del sistema, permitiendo por correlación entre los errores de las ecuaciones, mejorará la eficiencia de la estimación

Limitación

Si existe una variable omitida o endógena en una de las ecuaciones, la estimación por SUR no será insesgada y consistente para cualquiera de las ecuaciones

El modelo aparentemente no relacionado (SUR)

El modelo general

La estimación SUR es la aplicación de la estimación de mínimos cuadrados generalizados a un grupo de ecuaciones aparentemente no relacionadas

Las ecuaciones están relacionadas por medio de las covarianzas no cero asociadas con los términos de error a lo largo de diferentes ecuaciones en un punto determinado en el tiempo

Generalizando el modelo a un sistema de G ecuaciones se tiene:

$$\mathbf{Y}_i = \mathbf{X}_i\beta_i + \mathbf{u}_i, i = 1, 2, \dots, G$$

Combinando las G ecuaciones se tiene

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \dots \\ \mathbf{Y}_G \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_G \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_G \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \dots \\ \mathbf{u}_G \end{bmatrix}$$

El modelo aparentemente no relacionado (SUR)

El modelo general

De acuerdo con los supuestos del modelo aparentemente no relacionado, no hay autocorrelación dentro de las ecuaciones, pero existe correlación de ecuaciones cruzadas, es decir

$$\begin{aligned}\mathbf{\Omega} = E(\mathbf{uu}') &= \begin{bmatrix} E(\mathbf{u}_1\mathbf{u}'_1) & E(\mathbf{u}_1\mathbf{u}'_2) & \dots & E(\mathbf{u}_1\mathbf{u}'_G) \\ E(\mathbf{u}_2\mathbf{u}'_1) & E(\mathbf{u}_2\mathbf{u}'_2) & \dots & E(\mathbf{u}_2\mathbf{u}'_G) \\ \dots & \dots & \dots & \dots \\ E(\mathbf{u}_G\mathbf{u}'_1) & E(\mathbf{u}_G\mathbf{u}'_2) & \dots & E(\mathbf{u}_G\mathbf{u}'_G) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \dots & \sigma_{1G}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \dots & \sigma_{2G}\mathbf{I} \\ \dots & \dots & \dots & \dots \\ \sigma_{G1}\mathbf{I} & \sigma_{G2}\mathbf{I} & \dots & \sigma_{GG}\mathbf{I} \end{bmatrix}\end{aligned}$$

Se observa que es una matriz no diagonal como en el caso de la matriz var-cov de los errores en la estimación MCO

Algunas aplicaciones importantes en economía de la estimación de modelos por SUR incluye:

- **La estimación de un sistema de ecuaciones de demanda:** este sistema explica el consumo de bienes de los hogares. La correlación entre las ecuaciones podría deberse a específicos atributos no observables de los hogares que influyen el consumo de dichos bienes
- **Función de costos translog junto con la ecuación de participación de insumos:** Este sistema de ecuaciones se basa en datos de firmas y la correlación entre ecuaciones podría ser debida a efectos no observables específicos a las firmas que influyen la elección de insumos y costos en las decisiones de producción

El modelo aparentemente no relacionado (SUR)

El modelo general

La estimación más eficiente se obtiene aplicando la estimación de mínimos cuadrados generalizados (MCG) para obtener:

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

Para un sistema de dos ecuaciones tendremos:

$$\hat{\beta}_{MCG} = \begin{bmatrix} \sigma_{11}\mathbf{X}'_1\mathbf{X}_1 & \sigma_{12}\mathbf{X}'_1\mathbf{X}_2 \\ \sigma_{21}\mathbf{X}'_2\mathbf{X}_1 & \sigma_{22}\mathbf{X}'_2\mathbf{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11}\mathbf{X}'_1\mathbf{Y}_1 & \sigma_{12}\mathbf{X}'_1\mathbf{Y}_2 \\ \sigma_{21}\mathbf{X}'_2\mathbf{Y}_1 & \sigma_{22}\mathbf{X}'_2\mathbf{Y}_2 \end{bmatrix}$$

En la práctica, deben estimarse los elementos de $\mathbf{\Omega}$. Zellner (1962) recomienda el siguiente procedimiento de estimación por [MCG factibles](#):

- estimar cada ecuación por MCO
- obtener los residuales de cada ecuación
- calcular las diferentes $\hat{\sigma}_{ii}$ y $\hat{\sigma}_{ij}$:

$$\hat{\sigma}_{ii} = \sum_{n=1}^N e_{in}^2 / (N - K_i), i = 1, \dots, G$$

$$\hat{\sigma}_{ij} = \sum_{n=1}^N e_{in}e_{jn} / N, i, j = 1, \dots, G, i \neq j$$

donde e_{in} representan los residuales MCO de la i -ésima ecuación. $\hat{\sigma}_{ii}$ es el estimador insesgado para σ_{ii} , y $\hat{\sigma}_{ij}$ es el estimador consistente y asintóticamente eficiente (aunque no insesgado) de σ_{ij}

El modelo aparentemente no relacionado (SUR)

Pruebas para detectar correlación contemporánea

Ya que la no diagonalidad de Ω es el principal supuesto del método de estimación SUR, es importante probar $H_0 : \Omega$ es diagonal ($\sigma_{12} = \sigma_{13} = \dots = \sigma_{1G} = \sigma_{23} = \dots = \sigma_{(G-1)G} = 0$)

Breusch y Pagan (1980) derivaron una forma simple para realizar esta prueba y se basa en construir un estadístico de multiplicadores de lagrange (LM) sobre los coeficientes de correlación muestral de los residuales MCO. El estadístico tiene la forma:

$$LM = N \sum_{i=2}^G \sum_{j=1}^{i-1} r_{ij}^2 \sim \chi_{G(G-1)/2}^2$$

donde N es el número de observaciones, G es el número de ecuaciones y $r_{ij} = \hat{\sigma}_{ij} / (\hat{\sigma}_{ii} \hat{\sigma}_{jj})^{1/2}$. Las $\hat{\sigma}$'s son calculadas de los residuales MCO

Ejercicio aplicado: comportamiento del gasto público de US

En este ejercicio aplicado se va a analizar el comportamiento del gasto gubernamental para los estados de los Estados Unidos.

El sistema de ecuaciones que se va a estimar es:

$$EXP_i = \beta_1 + \beta_2 AID_i + \beta_3 INC_i + \beta_4 POP_i + u_{1i} \quad (1)$$

$$AID_i = \delta_1 + \delta_2 EXP_i + \delta_3 PS_i + u_{2i} \quad (2)$$

EXP: gasto público de los gobiernos estatales y locales (endógena)

AID: nivel de ayuda mediante subsidio federal (endógena)

INC: ingreso de los estados (exógena)

POP: población estatal (exógena)

PS: población estudiantil de primaria y secundaria (exógena)

u_{1i} y u_{2i} : términos de error

En los siguientes links se encuentran los datos y el código utilizado en R:

- [Datos](#)
- [Datos SUR](#)
- [Código en R](#)

Ejercicio aplicado: comportamiento del gasto público de US

$$EXP_i = \beta_1 + \beta_2 AID_i + \beta_3 INC_i + \beta_4 POP_i + u_{1i} \quad (1)$$

$$AID_i = \delta_1 + \delta_2 EXP_i + \delta_3 PS_i + u_{2i} \quad (2)$$

Determinando si las ecuaciones están identificadas:

- **Condición de orden:** para que una ecuación esté identificada es necesario que al menos una variable exógena se excluya de la ecuación analizada

La ecuación (1) satisface la condición de orden ya que PS se encuentra excluida \implies exactamente identificada

La ecuación (2) satisface la condición de orden ya que INC y POP se encuentran excluidas \implies sobreidentificada

- **Condición de rango:** una ecuación está identificada si, y sólo si, la segunda ecuación contiene al menos una variable exógena (con un coeficiente diferente de cero) excluida de la primera

Para corroborar esta condición se estiman por MCO las ecuaciones de la forma reducida:

$$EXP_i = \pi_1 + \pi_2 INC_i + \pi_3 POP_i + \pi_4 PS_i + \epsilon_1$$

$$AID_i = \pi_5 + \pi_6 INC_i + \pi_7 POP_i + \pi_8 PS_i + \epsilon_2$$

Y luego se prueba que los coeficientes de las variables excluidas sean estadísticamente significativas

Ejercicio aplicado: comportamiento del gasto público de US

Cargando las librerías

```
library(stargazer); library(car); library(readxl); library(AER); library(systemfit)
```

Leyendo los datos

```
setwd("C:/Users/ggarci24/OneDrive - Universidad EAFIT/EAFIT/Cursos EAFIT/Econometria II/R/Tema 10")  
data <- read_excel("EX73.xlsx")
```

Estimamos por MCO la ecuación (1) y (2) en la forma reducida y corroboramos que el coeficiente asociado a *PS*, *INC* y *POP* sean o no estadísticamente significativos

```
EXP.reducida <- lm(EXP~INC+POP+PS, data=data)  
AID.reducida <- lm(AID~INC+POP+PS, data=data)
```

```
linearHypothesis(AID.reducida, "PS = 0")
```

```
linearHypothesis(EXP.reducida, c("INC = 0", "POP = 0"))
```

Linear hypothesis test

Hypothesis:
PS = 0

Model 1: restricted model
Model 2: AID ~ INC + POP + PS

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	47	2496640				
2	46	2265748	1	230893	4.6877	0.0356 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Linear hypothesis test

Hypothesis:
INC = 0
POP = 0

Model 1: restricted model
Model 2: EXP ~ INC + POP + PS

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	48	108094974				
2	46	27978053	2	80116921	65.862	3.157e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

El coeficiente de *PS* es estadísticamente significativo, lo cual indica que [la ecuación \(1\) está identificada](#). Los coeficientes de *INC* y *POP* son estadísticamente significativos, lo cual indica que [la ecuación \(2\) está identificada](#)

Ejercicio aplicado: comportamiento del gasto público de US

Prueba de simultaneidad

El test de Hausman comprende los siguientes pasos:

- **Paso 1:** Estime la regresión de la forma reducida de AID y obtenga los residuales de esta regresión (w)
- **Paso 2:** Estime la regresión de EXP sobre AID, INC, POP y w , y realice una prueba t sobre el coeficiente de w . Si éste es significativo, no se rechaza la hipótesis de simultaneidad

Obtenemos los residuales de las ecuaciones estimadas de la forma reducida

```
data$w <- residuals(AID.reducida)
data$v <- residuals(EXP.reducida)
```

Estimamos los modelos incorporando los residuales para corroborar la existencia de simultaneidad. Notamos que los coeficientes de w y v son estadísticamente significativos al 10%, con lo cual [hay problemas de simultaneidad](#)

```
hausman1 <- lm(EXP~AID+INC+POP+w, data=data)
summary(hausman1)
```

```
Call:
lm(formula = EXP ~ AID + INC + POP + w, data = data)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1114.12  -186.07    26.29    87.91  1069.76
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.941e+01  8.602e+01  -1.040  0.3041
AID          4.501e+00  7.643e-01   5.889 4.57e-07 ***
INC          1.293e-04  4.222e-05   3.062  0.0037 **
POP         -5.181e-01  1.118e-01  -4.633 3.09e-05 ***
w           -1.391e+00  8.023e-01  -1.734  0.0898 .
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 267.3 on 45 degrees of freedom
```

```
hausman2 <- lm(AID~EXP+PS+v, data=data)
summary(hausman2)
```

```
Call:
lm(formula = AID ~ EXP + PS + v, data = data)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
 -294.040  -43.120   -5.014   72.491  188.044
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 37.32576   23.29099   1.603 0.115872
EXP          0.16408    0.01216  13.490 < 2e-16 ***
PS           0.12207    0.05612   2.175 0.034803 *
v            0.08774    0.02391   3.670 0.000628 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 108.9 on 46 degrees of freedom
Multiple R-squared:  0.9842    Adjusted R-squared:  0.9822
```

Ejercicio aplicado: comportamiento del gasto público de US

Estimación de sistemas de ecuaciones

Primero se crea el sistema

```
eqEXP <- EXP ~ AID + INC + POP
eqAID <- AID ~ EXP + PS
eq.sys <- list(EXP = eqEXP,
              AID = eqAID)
```

MCO: ignorando la estructura de sistema de ecuaciones. Esto es equivalente a estimar cada ecuación por separado

```
MCO <- systemfit(eq.sys, method = "OLS", data = data)
summary(MCO)
```

MC2E: equivalente a la estimación de variables instrumentales de la ecuación única. Para la estimación por MC2E es necesario tener variables instrumentales para la identificación. Los instrumentos son las variables exógenas

- Mismo instrumento para las dos ecuaciones

```
inst <- ~ INC + POP + PS
MC2E.1 <- systemfit(eq.sys, method = "2SLS", inst = inst, data = data)
summary(MC2E.1)
```

Note que lo anterior es lo mismo que hacer IV

```
IV <- ivreg(EXP ~ AID+INC+POP | INC+POP+PS, data=data)
summary(IV)
```

- Diferentes instrumentos por ecuación

```
inst1 <- ~ INC + POP + PS
inst2 <- ~ INC + POP
instlist <- list(inst1, inst2)
MC2E.2 <- systemfit(eq.sys, method = "2SLS", inst = instlist, data = data)
summary(MC2E.2)
```

Ejercicio aplicado: comportamiento del gasto público de US

Modelo SUR

En el modelo SUR los residuales son correlacionados a través de las ecuaciones. La estimación se hace la siguiente forma:

```
SUR <- systemfit(eq.sys, method = "SUR", data = data)
summary(SUR)
```

Test de Breusch-Pagan

Ho: la matrix de covarianza de los errores es diagonal, las ecuaciones son independientes. Recordemos que el estadístico tiene la forma

$$LM = N \sum_{i=2}^G \sum_{j=1}^{i-1} r_{ij}^2 \sim \chi_{G(G-1)/2}^2$$

donde $r = \sigma_{ij} / (\sigma_{ii}\sigma_{jj})^{1/2}$

Primero se calcula la matrix var-cov de los residuales estimados. Con la siguiente función nos muestra dicha matriz var-cov y la matriz de correlación de los residuales

```
summary(MCO, residCov = TRUE, equations = FALSE)
```

Ahora calculamos el r y lo comparamos con el coeficiente de correlación, deben dar igual

```
r <- (-25845.3)/((140757.9*14996.3)^(1/2))
```

Y procedemos calcular el estadístico de prueba

```
BP <- 50*((r)^2)
BP
pchisq(BP, df=1, lower.tail = F) # Pvalor de BP
qchisq(0.05, df=1, lower.tail = F) # Chi de la tabla al 5% y con 1 gdl
```

Pvalor = 0, rechazamos Ho, es decir que la estimación SUR es adecuada

Ejercicio aplicado: comportamiento del gasto público de US

Estimando por MC3E el modelo SUR

En la anterior estimación del modelo SUR no se consideró el problema de la variable explicativa endógena. en particular, si las ecuaciones continen variables explicativas endógenas, SUR será sesgado e inconsistente

Así que es necesario estimar por MC3E que toma en cuenta la correlación contemporanea entre los residuales y la existencia de una variable explicativa endógena. Esto es como IV en el SUR

Las tres etapas serían:

1. Regrese cada variable endógena sobre todas las variables exógenas y calcule los valores predichos de las variables endógenas
2. Estime las ecuaciones estructurales por MCO, reemplazando las variables endógenas que se encuentran como explicativas por sus valores predichos de la primera etapa
3. Calcule las varianzas y covarianzas estimadas de los residuales de la etapa 2, y reestime las ecuaciones estructurales usando el método SUR

La estimación la podemos hacer de la siguiente forma:

```
MC3E <- systemfit(eq.sys, method = "3SLS", inst = inst, data = data)
summary(MC3E)
```