REGRESIÓN APARENTEMENTE NO RELACIONADOS

Verónica Cuenca Ramallo

Desarrolla una técnica econométrica conocida como Regresión Aparentemente No Relacionados o modelos SUR, la cual constituye un caso muy especifico de un sistema de ecuaciones simultaneas en el que la correlación entre las ecuaciones se origina entre los errores de éstas y no la incorporación de variables endógenas (dependientes) como variables predeterminadas en otras ecuaciones del sistema.

1.- INTRODUCCION.-

Usualmente los modelos de ecuaciones simultáneas, se piensa de inmediato en aquellos sistemas en los que se especifican variables endógenas (dependientes) en algunas ecuaciones como predeterminadas en otras ecuaciones del mismo modelo.

Bajo esta especificación existe una correlación identificable de los términos de error entre las ecuaciones del sistema donde no debe confundirse con el problema de autocorrelación, el cual toma lugar cuando existe correlación de los términos de error dentro de cada ecuación.

Para las ecuaciones simultaneas, la aplicación de mínimos cuadrados ordinarios(MCO) resulta en estimadores sesgados y con errores cuadrados medios que pueden ser bastante elevados, especialmente en muestras pequeñas, se utilizan procedimientos tales como los mínimos cuadrado s ordinarios y la estimació n máximo verosímil.

2.- MODELO SUR.-

Las siglas **SUR** provienen del nombre que recibe en inglés este sistema de ecuaciones (Seemingly Unrelated Regressions) que en castellano significa Regresiones aparentemente no relacionados.

Sin embargo, considérese un conjunto de ecuaciones de regresión de la siguiente forma.

$$
\begin{gathered}
y_{1i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{\ i} + \cdots + \beta_k x_{Ki} + u_i \\
y_{2i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{2i} \\
y_{3i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{3i} \\
\cdots \\
y_{Mi} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{Mi}
\end{gathered}
$$

Donde i= 1,2,3 .... N

En este sistema existen **M** variables endógenas (dependientes) denotadas como $\mathbf{y}_{Mi}$, cada una asociada con un término de error $\mathbf{u}_{Mi}$, así como con un conjunto de **K** variables exógenos (independientes) $\mathbf{X}_{Ki}$.

La relación entre las variables endógenas y las exógenos está dada por los coeficientes b_k,el número de observaciones ***i*** para cada variable (endógenas, exógenas y términos de error) es de **N**.

En principio se podría razonar que, al no observarse variables endógenas $\mathbf{y}_{Mi}$ como variables predeterminadas en otras ecuaciones del sistema, cada una de las ecuaciones podría ser estimada con el uso de mínimos cuadrados ordinarios.

Esto sería posible si las ecuaciones fueran completamente independientes en el sentido de que la variabilidad de alguna de las variables endógenas no afectara el comportamiento de alguna otra ecuación.

En el vocabulario econométrico ello sería equivalente a decir que la matriz de varianzas ycovarianzas del sistema de ecuaciones tiene triángulos (inferior y superior) iguales a cero.

En otras palabras, sería una matriz con una diagonal diferente de cero y cuyas entradas serían las variancias de los términos de error de cada ecuación.

Esta conjetura, sin embargo, podría ser incorrecta si se detectara algún tipo de movimiento simultáneo de todas las ecuaciones originado por una supuesta relación contemporánea entre los términos de error que no se origina por la presencia de variables endógenas como variables predeterminadas en las ecuaciones.

Es decir, las regresiones que no están aparentemente correlacionadas, sí lo estarían por medio de correlaciones implícitas, sin modelar específicamente, entre los términos de error.

3.- LA ESPECIFICACIÓN DEL MODELO GENERAL.-

Suponga que la representación matricial de la m-ésima ecuación del sistema (1) es de la siguiente forma:

$$Y_m = X_m \beta_m + U_m \quad m=1, 2,3,\ldots\ldots\ldots\ldots, M$$

Donde los vectores **Y** y **U** son de orden (N x 1), **X** es una matriz de orden (N x Km), donde Km es el número de variables exógenas en la m-ésima ecuación, el numero de variable exógena no tienen que ser el mismo para cada una de las ecuaciones, es más, la eficiencia del método SUR aumenta en el tanto que las variables X muestren una menor asociación entre ecuaciones y **βm** es un vector de parámetros de orden (Km x 1). Al considerar las **M** ecuaciones de forma matricial se tiene la siguiente representación:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_M \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ . \\ . \\ U_M \end{bmatrix}$$

La cual se puede expresar, en forma matricial, como:

$$Y = XB + U \qquad (4)$$

Donde la dimensión de Y es (MN x 1), la de X es (MN x K), la de b es

(K x 1) y la de U es (MN x ¡) en este caso, $K = \sum_{m=1}^{M} K_m$

Dado que $U_{m\,i}$ es el valor observado del término de error de la m-ésima ecuación en el i-ésimo período, el supuesto de **correlación contemporánea** de los errores, pero no correlación serial, implica que, E [Umi, Ujs] = s_{mj} si i=s, pero es igual a cero si i#s, en otras palabras, cuando los períodos **i** y **s** coinciden existe una covariancia diferente de cero entre los errores de las ecuaciones **m** y **j**.

En forma matricial se puede especificar una matriz de variancias y covariancias para las ecuaciones **m** y **j** como E [Um U_j^t]=s_{mj} I_N, donde **I** es la matriz identidad de orden **N**.

Para los M vectores de términos de errores existe una matriz de varianzas y covarianzas que asume la siguiente forma:

$$\Phi = E[UU^t] = \begin{bmatrix} \sigma_{11}I_N & \sigma_{12}I_N & \dots & \sigma_{1M}I_N \\ \sigma_{21}I_N & \sigma_{22}I_N & \dots & \sigma_{2M}I_N \\ . & & & \\ . & & & \\ \sigma_{M1}I_N & \sigma_{M1}I_N & \dots & \sigma_{MM}I_N \end{bmatrix} = \Sigma \otimes I_N$$

Donde $\otimes$ es el producto Kronecker y Σ la matriz de varianzas y covarianzas de la forma:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ . & & & \\ . & & & \\ \sigma_{M1} & \sigma_{M1} & \dots & \sigma_{MM} \end{bmatrix}$$

La matriz I es simétrica y se supone que es definida positiva y que no es singular, el producto Kronecker es la expansión de cada una de las entradas de una matriz (en este caso las variancias y covariancias de la matriz I) por una matriz, se dice que una matriz es simétrica cuando las entradas de sus triángulos superior e inferior son iguales.

En una estructura matricial el producto $S_{mj}I_N$ denota los términos de error dentro de cada ecuación son homocedásticos (varianza del error constante) y que no tienen autocorrelación en el tiempo (existe casos que se trata de datos de corte transversal).

4.- ESTIMACIÓN CON MATRIZ Σ CONOCIDA.-

Dado que el modelo no involucra la simultaneidad de las variables endógenas en el sentido de los sistemas de ecuaciones sim ultáneas, el procedimiento de estimaciones de un modelo SUR, cuando son conocidas las varianzas y covarianzas de la matriz Σ , es el correspondiente al de mínimos cuadrados generalizados. En este caso, el estimador toma la forma:

$$\hat{\beta} = [X^1\Phi^{-1}X]^{-1}X^1\Phi^{-1}Y$$

La cual es equivalente a:

$$\hat{\beta} = [X^1(\Sigma^{-1}\otimes I)^{-1}X]^{-1}X^1(\Sigma^{-1}\otimes I)Y$$

Es decir, el mejor estimador lineal insesgado. El estimador expresado detalladamente es de la forma:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_M \end{bmatrix} = \begin{bmatrix} \sigma^{11}X_1^1X_1 & \sigma^{12}X_1^1X_2 & \dots & \sigma^{1M}X_1^1X_M \\ \sigma^{21}X_2^1X_1 & \sigma^{22}X_2^1X_2 & \dots & \sigma^{2M}X_2^1X_M \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^{M1}X_M^1X_1 & \sigma^{M2}X_M^1X_2 & \dots & \sigma^{MM}X_M^1X_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{m=1}^{M} \sigma^{1m}X_1^1y_m \\ \sum_{m=1}^{M} \sigma^{2m}X_2^1y_m \\ \vdots \\ \sum_{m=1}^{M} \sigma^{Mm}X_M^1y_m \end{bmatrix}$$

Donde los escalare σ^{mj} corresponden al elemento (m,j) de la matriz de? .
La matriz de varianzas y covarianzas del estimador mínimo cuadrático generalizado de β es:

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

En la discusión de los estimadores SUR hay tres aspectos que destacar sobre la eficiencia (entendido como varianza mínima).

1) Cuanto más elevada la correlación contemporánea de los términos de error entre ecuaciones mayor será la ganancia en eficiencia del estimador generalizado.
2) Si la correlación contemporánea es muy baja (los triángulos de la matriz E se aproxima a cero) no hay una ganancia importante por aplicar la regresión SUR a las ecuaciones en vez de utilizar los mínimos cuadrados ordinarios en cada ecuación.
3) Si cada una de las ecuaciones del sistema tiene las rrismas variables exógenas, entonces los estimadores SUR son equivalentes a los mínimos cuadrados ordinarios.

En general, cualquier ganancia en eficiencia (mínima varianza) tiende a ser mayor cuando las variables explicativas en las diferentes ecuaciones no están altamente correlacionadas.

5.- ESTIMACION CON MATRIZ Σ DESCONOCIDA.-

En la práctica es poco probable contar con los verdaderos valores de la varianzas y covarianzas (σ_{mj}) de los errores, por lo que es necesario recurrir a una estimación preliminar de los errores.

Este procedimiento se logra mediante la aplicación del método de mínimos cuadrados ordinarios a cada una de las ecuaciones del sistema.

El estimador es de la forma $b = (X_M^1 X)^{-1} X_M^1 y_m$, el cual permite obtener una primera estimación de errores mínimo cuadráticos $\hat{u}_m = y_m - X_m b_m$ para cada ecuación.

Si bien las variancias y covariancias calculadas con estos errores mínimo cuadráticos son sesgadas en muestras pequeñas , tienen la propiedad de consistencia que permite continuar con el procedimiento SUR.

La forma genérica de cálculo es la siguiente:

$$\hat{\sigma}_m = \frac{1}{N}\hat{u}_m^1\hat{u}_j = \frac{1}{N}\sum_{i=1}^{N}\hat{u}_{mi}\hat{u}_{ji}$$

Si se definiera $\hat{\Sigma}$ como la matriz de variancias y covariancias conformada por las estimaciones $\hat{\sigma}_{mj}$; , el correspondiente estimador mínimo cuadrático generalizado asume la forma:

$$\tilde{\beta} = \left[X^1\left(\hat{\Sigma}^{-1}\otimes I\right)\right]^{-1} X^1(\hat{\Sigma}^{-1}\otimes I)Y$$

Este estimador es de uso generalizado y en la literatura se le conoce como el estimador mínimo cuadrático generalizado de Séller.
Sin embargo, otro estimador con propiedades asintóticas más deseables es el carácter iterativo, es decir, una vez que se cuenta con los estimadores de la fórmula $\tilde{\beta} = \left[X^1\left(\hat{\Sigma}^{-1}\otimes I\right)\right]^{-1} X^1(\hat{\Sigma}^{-1}\otimes I)Y$, se calculan los errores y las varianzas a partir de ello, los cuales se utilizan en un nuevo cálculo de los mínimos cuadrados generalizados, el procedimiento se repite hasta que la función de verosimilitud alcance un máximo.

En tanto que las varianzas y covarianzas tiendan a permanecer constantes, ambos procedimientos tiene la misma distribución límite, esta distribución es aproximadamente normal con una media β y con una matriz de varianzas y covarianzas consistentemente estimada por:

$$\left[X^{1}\left(\hat{\Sigma}^{-1} \otimes I\right)X\right]^{-1}$$

Por consiguiente, el estimador de Zellner y el estimador iterado serán asintoticamente más eficiente que el estimador mínimo cuadrático.

En muestras finitas, sin embargo, existe una región en el espacio paramétrico de Σ , en laque el estimador mínimos cuadrático es más eficiente.

Es el caso cuando la pérdida de eficiencia originada por el uso de una estimación de Σ vez de la verdadera matriz, es mayor que la ganancia obtenida por utilizar un estimador mínimo cuadrático generalizado (especialmente si las correlaciones contemporáneas entre los errores son pequeñas), el caso extremo es cuando Σ es diagonal y el estimador mínimo cuadrático es el relevante.

6 -EJEMPLO DE LA APLICACIÓN DEL MODELO SUR.-

Como ejemplo de aplicación, puede ser los cuestionarios del Curso Prefacultativo, variables que se puede considerar son.

Donde vive, procedencia del colegio, edad, sexo, si trabaja o no trabaja, el factor de la economía de los alumnos y a la vez de los padres, que les afecta para que reprueben o aprueben, etc., estudiar si existe un impacto del conjunto de regresores sobre las puntuaciones en Matemáticas, Informática, Química y Física.

7-CONCLUSIÓN.-

Lo interesante de los modelos **SUR** (Modelo Aparentemente no Relacionados), es que se logra evaluar simultáneamente el efecto de las variables independientes sobre el conjunto de las variables dependientes, tomando en cuenta la correlación que existe entre las ecuaciones.

8.- Bibliografía

[1] Agresti Alan Wiley Jhon (1938), Análisis de Datos Categóricos, New Cork.

[2] J.Aníbal Angulo (2003), Estudio Socioeconómico de la Población Estudiantil Biblioteca de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales, La Paz- Bolivia.

[3] Dhrymes Phoebus J., Me Gran Hill (1995) Económetria, México.

[4] Greene William H. (1999), Análisis Econométrico Tercera edición, Madrid España

[5] Gujarati Damodar N. (1997), Econometría Básica, Santafé de Bogotá, Colombia.

[6] Hair Jr.Joseph F., Ralph E. Anderson (1999), Análisis Multivariante, Madrid España.

[7] Araya Monge Rigoberto, Muñoz Giro Juan E., Regresiones que Aparentemente no están Relacionados, disponible en internet.

“El éxito de la vida: si piensas que estas vencido, vencido estarás, Si piensas que no te atreves no lo harás, Si piensas que te gustaría ganar, pero no lo puedes no lo lograras, Si piensas que perderás, ya has perdido, **PORQUE EN EL MUNDO EL ÉXITO COMIENZA CON LA VOLUNTAD DEL HOMBRE**, Todo esto esta en el estado mental porque muchas carreras se ha perdido entes de haberse corrido y muchos cobardes han fracasado antes de haber su trabajo comenzado”.