

USO DE MODELOS DE SISTEMAS DE ECUACIONES PARA DATOS DE PANEL CON INFORMACIÓN DE LA EPH

Blaconá, María Teresa*; **Garcia, María del Carmen***; **Borgognone, María Gabriela***; **Bussi, Javier***; **Ventroni, Nora***; **Pellegrini, José Luis****
**Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística. ** Instituto de Investigaciones de la Escuela de Economía. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario.*

RESUMEN

Los modelos de regresión con componente de error a uno o dos criterios constituyen una opción para analizar datos de panel. En ciertas ocasiones interesa explicar el comportamiento simultáneo de ciertas variables a través de un conjunto de ecuaciones que reflejen las relaciones que se dan entre las mismas. En este trabajo se describen tanto los modelos uniecuacionales como los de sistemas de ecuaciones para datos de panel y se presentan métodos de estimación que brindan estimadores eficientes y asintóticamente consistentes. Se presentan dos estudios realizados con datos de la EPH donde, en el primero, se describe el comportamiento del mercado de trabajo en la República Argentina y en el segundo se analiza el ingreso del varón, para aquellos hogares donde cohabita una pareja conyugal.

INTRODUCCIÓN

La EPH es una de las fuentes más importantes de información socio-laboral de la República Argentina.

- Es llevada a cabo por el INDEC en forma periódica (dos veces al año) en 28 aglomerados urbanos.
- Estos aglomerados representan el 70% de la población urbana y el 98% de la población residente en centros de más de 100.000 habitantes (INDEC, 1998).
- En su diseño de muestreo, la EPH utiliza un esquema de panel rotativo, donde cada vivienda seleccionada, permanece en la muestra cuatro ondas (dos años) por lo tanto en cada onda se renueva el 25% de los hogares.

Debido a estas características los datos registrados constituyan medidas repetidas en el tiempo para cada unidad, sea esta última un individuo, hogar o aglomerado. Este tipo de datos en la literatura econométrica se conoce como datos de panel (Hsiao, 1986, Baltagi, 1995).

El uso de los datos de panel permite especificar y estimar modelos más complicados y muchas veces más realistas, ya que se tienen en cuenta tanto la sección cruzada en un momento como la serie de tiempo.

En ciertas ocasiones, una única ecuación es suficiente para estimar relaciones de interés entre distintas variables medidas en la EPH. Pero en muchos casos puede ser interesante explicar el comportamiento simultáneo de ciertas variables a través de la formulación de un conjunto de ecuaciones que reflejen las relaciones existentes entre ellas.

METODOLOGÍA

Sistema de ecuaciones para datos de panel estimación del modelo.

Sea el sistema de ecuaciones definido como:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \quad (1) \quad \text{donde} \quad \mathbf{y}' = (\mathbf{y}'_1, \dots, \mathbf{y}'_M) \quad \mathbf{Z} = \text{diag}[\mathbf{Z}_j]$$

$\mathbf{Z}_j = [\mathbf{y}_j, \mathbf{x}_j]$ de dimensión NTx(g_j+k_j), j = 1, ..., M, g_j: tamaño de y_j del segundo miembro, k_j: tamaño de X_j

$$\boldsymbol{\delta}' = (\boldsymbol{\delta}'_1, \dots, \boldsymbol{\delta}'_M) \quad \mathbf{u}' = (\mathbf{u}'_1, \dots, \mathbf{u}'_M)$$

Modelo:

- Sistema de ecuaciones con datos de panel.
- En las ecuaciones, variables endógenas pueden aparecer como explicativas.

Los estimadores **mínimos cuadrados ordinarios** no son consistentes esencialmente por dos motivos:

- a) por la estructura de correlación que puede existir entre observaciones de un mismo individuo y
- b) porque puede existir correlación, en las distintas ecuaciones, entre el error y las variables endógenas que actúan como explicativas.

Para resolver estos problemas se utiliza **mínimos cuadrados a tres etapas para datos de panel**, propuesto por Baltagi (1995).

APLICACIÓN 1

Objetivo: Detectar el signo y la magnitud de las interacciones en el mercado de trabajo en la República Argentina mediante un modelo de ecuaciones simultáneas para datos de panel.

Sistema de dos ecuaciones:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_1 + \beta_2 y_2 + u_1 \\ y_2 = \beta_3 + \beta_4 x_2 + \beta_5 x_3 + \beta_6 y_1 + u_2 \end{cases} \quad \mathbf{u} \sim N(0, \Sigma)$$

VARIABLES ENDÓGENAS:

y_1 : incremento del log de la tasa de desocupación en el momento t,

y_2 : incremento del log de la tasa de actividad en el momento t.

VARIABLES EXÓGENAS:

x_1 : incremento del log de la tasa de desocupación en el momento t-1,

x_2 : incremento del log de la tasa de actividad en el momento t-1,

x_3 : incremento del log de la tasa de empleo en el momento t.

Parámetros a estimar: β_i : i= 1-6

u_1 y u_2 : errores de la primera y segunda ecuación respectivamente

VARIABLES Y ERRORES: vectores 550x1, 25 (N) aglomerados por 22 (T) momentos.

Ecuac.	Regr.	Método de Estimación					
		MC3EEFI		MC3E		MLG	
		Coef.	p	Coef.	p	Coef.	p
Y ₁	Const	0.5464	0.0001	0.0243	0.011	0.0182	0.0001
	X ₁	-0.1442	0.0001	-0.4770	0.0001	-0.3572	0.0001
	Y ₂	-0.6789	0.0001	-5.1800	0.0001	1.7245	0.0001
Y ₂	Const.	0.2957	0.0005	0.0005	0.0103	0.0004	0.0330
	X ₂	-0.1593	0.0001	-0.0721	0.0001	-0.065	0.0001
	X ₃	1.2823	0.0001	0.8914	0.0001	0.9368	0.0001
	Y ₁	1.3572	0.0001	0.0643	0.0001	0.0761	0.0001

Residuos	Correl.	p	Correl.	p	Correl.	p
Ecuación 1	Y ₂		X ₁			
	0.6336	0.0001	-0.1909	0.0001		
Ecuación 2	X ₂		X ₃		Y ₁	
	-0.0235	0.5818	-0.0003	0.9947	-0.2229	0.0001

El modelo estimado correctamente (MC3EEFI) muestra que las variaciones del logaritmo de la tasa de desocupación están inversamente relacionadas con la misma variable rezagada un período y con las variaciones del logaritmo de la tasa de actividad. Esta última, a su vez, se halla inversamente relacionada con ella misma en un período anterior y directamente relacionada con el logaritmo de la tasa de empleo y el logaritmo de la tasa de la desocupación.

APLICACIÓN 2 (en su etapa inicial)

Se analiza el ingreso del varón, para aquellos hogares donde cohabita una pareja conyugal, teniendo en cuenta las diferencias entre los niveles de escolarización de los cónyuges.

Pueden plantearse dos hipótesis alternativas sobre el origen de una posible relación entre los niveles de ingresos de mercado del varón y las diferencias de escolarización entre ambos miembros de la pareja.

- La primera se basa en la complementariedad de características que entran como factores en la función de producción para el mercado.
- La segunda se basa en los resultados de los procesos de búsqueda en el curso de los cuales se ponen en práctica estrategias orientadas a lograr parejas más valiosas.

Objetivo: Verificar empíricamente cuál de las dos hipótesis alternativas se cumple en las parejas de la Argentina.

Modelo. El estudio empírico se realiza sobre datos de hogares de aglomerados urbanos del país, relevados en las ondas de mayo de 1998 a octubre 1999 por la EPH. Los hogares incluidos son aquellos en los que se registró la presencia de dos personas de diferente sexo, clasificadas como jefe de hogar y cónyuge. Las posibles variables a tener en cuenta son:

VARIABLES ENDÓGENAS: y_1 : logaritmo del ingreso del varón, y_2 : diferencia de escolaridad entre varón y mujer (escolaridad varón – escolaridad mujer)

VARIABLES EXÓGENAS: x_1 : edad varón, x_2 : edad varón al cuadrado, x_3 : escolaridad varón, x_4 : total horas trabajadas en la última semana por el varón, x_5 : antigüedad en el trabajo del varón, x_6 : número hijos menores de 18 años en el hogar,

x_7 : cant. ocupaciones del varón, x_8 : vble “dummy” condición actividad varón (1: ocupado, 0: otro caso), x_9 : vble “dummy” condición actividad varón (1: desocupado, 0: otro caso), x_{10} : ingreso mujer, x_{11} : edad mujer,

x_{13} : vble “dummy” condición actividad mujer (1: activa, 0: otro caso).

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{13} + \beta_{12} y_2 + u_1 \\ y_2 = \beta_{13} + \beta_{14} x_1 + \beta_{15} x_{11} + \beta_{16} x_{10} + u_2 \end{cases} \quad \mathbf{u} \sim N(0, \Sigma)$$

Este modelo aún no ha sido estimado. En este momento se están depurando las bases de datos, las cuales presentan inconsistencias y valores perdidos, que se deben analizar para determinar la mejor manera de tratarlos.

CONSIDERACIONES FINALES

En este trabajo se muestra la utilidad de los Modelos de Sistema de Ecuaciones para el análisis de datos de panel. Ésta se ve reflejada en el análisis de interrelaciones, tanto entre individuos como a través del tiempo, de distintos aspectos que se miden en la EPH. En la primera aplicación el uso de los modelos propuestos permite determinar el signo de la relación y su magnitud, mediante estimaciones eficientes. También se puede apreciar el cambio que se produce en las estimaciones cuando no se emplea el método de estimación apropiado. En la segunda aplicación, si bien se encuentra en su etapa inicial, permite apreciar que se pueden plantear modelos más complejos que los de una única ecuación, para tener en cuenta características propias de cada individuo. En esta aplicación también quedan de manifiesto los inconvenientes que se presentan al trabajar con datos de panel.

REFERENCIAS

Baltagi, B. (1995) “Econometric Analysis of Panel Data”, John Wiley & Sons. **Hsiao, Cheng (1986)** “Analysis of Panel Data”, Econometric Society Monographs N°11. Cambridge University Press. **INDEC (1998)** “Encuesta de Hogares: Reformulación de la Encuesta Permanente de Hogares de Argentina”, INDEC-ISI.