María Teresa Blaconá*
María del Carmen García*
María Gabriela Borgognone*
Javier Bussi*
Nora Ventroni*
José Luis Pellegrini**
** Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística*
*** Instituto de Investigaciones Económicas*

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

I.- Introducción

La Encuesta Permanente de Hogares (EPH) se lleva a cabo por el Instituto Nacional de Estadística y Censos (INDEC) en forma periódica (dos veces al año) en veintiocho 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), por lo que constituye una de las fuentes más importante de información socio-laboral de la República Argentina.

La captación de información se realiza mediante dos cuestionarios: uno familiar y otro individual. Esta información permite el conocimiento de características demográficas básicas, laborales, migratorias, educativas, habitacionales y de ingresos.
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 veinticinco por ciento de los hogares.

Las características de la Encuesta, enunciadas en los párrafos anteriores, hacen que los datos registrados constituyan medidas repetidas en el tiempo para cada unidad, sea esta última una persona, un hogar o un aglomerado. Este tipo de datos en la literatura econométrica se conoce como datos de panel (Hsiao,1986; Baltagi,1995).

Usar los datos de panel donde se dispone de observaciones repetidas sobre el mismo individuo permite especificar y estimar modelos más complicados y a veces más realistas, ya que se tienen en cuenta tanto la sección cruzada en un momento, como la serie de tiempo. Esto implica varios beneficios, entre los que se pueden destacar (Baltagi, 1995):

i) controla la heterogeneidad individual, ya que en los datos de panel los individuos, en este caso aglomerados, hogares o personas, son heterogéneos. Si no se controla esta heterogeneidad se corre el riesgo de obtener resultados sesgados. Por ejemplo, las tasas de desocupación varían de aglomerado a aglomerado y a su vez varían en el tiempo;

ii) los datos son más informativos, presentan distintos tipos de variabilidades, se tiene menos colinealidad entre las variables, más grados de libertad y más eficiencia. Por ejemplo, la mayoría de los estudios de series de tiempo presenta el problema de multicolinealidad. Esto es menos probable si se adicionan las secciones cruzadas, ya que incorpora variabilidad y son más informativas. Por ejemplo, se consigue más información si en lugar de analizar sólo las series de las tasas totales de desocupación y de actividad de la República Argentina, se estudian las series de dichas tasas teniendo en cuenta el comportamiento de cada aglomerado. De hecho, la variación en los datos se puede descomponer en variación entre aglomerados de distintas características y variación dentro de los aglomerados. Por tal motivo, se pueden hallar estimaciones más confiables de los parámetros. Por supuesto, si se cumple la misma relación para todos los aglomerados los datos se pueden amalgamar;

iii) son mejores para estudios de ajustes dinámicos debido a que son necesarios los paneles para la estimación de relaciones intemporales;

iv) permiten identificar y medir efectos que simplemente no se detectan con sólo una sección cruzada o con sólo una serie de tiempo. Tal es el caso del efecto de los individuos sobre el ingreso cuando cambia alguna característica en su condición, por ejemplo de activo a inactivo;

v) permiten construir y probar modelos de comportamiento más complicados que los modelos que son para datos de sólo sección cruzada o de series de tiempo.

Sin embargo, también se deben considerar algunas desventajas que presentan los datos de panel, entre las que se pueden mencionar:

i) problemas de diseño y recolección de datos, como los de cobertura (cantidades incompletas de la población de interés), no respuesta (debido a la falta de cooperación del entrevistado, cansancio por tener que responder lo mismo en varias oportunidades), olvido (el entrevistado no recuerda correctamente), frecuencia de las entrevistas, período de referencia y otros;

ii) distorsión de los errores de medida, estos errores pueden surgir por falta de respuesta debido a que la pregunta no está clara, errores de memoria, distorsiones deliberadas de respuestas, informantes inapropiados, mal registro de la respuesta y efectos de la entrevista. Por ejemplo, un mismo individuo en dos ondas consecutivas responde que posee distinto años de escolaridad;

iii) problemas de selección, los que incluyen varios aspectos: a) propia selección, por ejemplo una persona que no trabaja porque tiene reserva de dinero muy alta y declara ingreso cero, b) no respuesta, hogares que se niegan a responder desde el principio y c) desgaste, aunque la no respuesta también ocurre en las secciones cruzadas, en este caso es un problema más serio, porque en las sucesivas ondas son más los individuos que no responden, por ejemplo por muerte, mudanza, cansancio, etc.. Por lo general, la no respuesta de un mismo individuo va aumentando en las sucesivas ondas;

iv) la corta dimensión de las series de tiempo, para el caso de la EPH, cada hogar sólo permanece cuatro ondas. Esto puede ocasionar un problema serio para los argumentos asintóticos. De hecho esto acrecienta la dificultad de cálculo para modelos de datos de panel con variables dependientes.

Entre los modelos que se utilizan para analizar este tipo de datos se encuentran los modelos lineales, del tipo modelos de regresión de componente de error a uno o a dos criterios. En los primeros se piensa que existe un efecto de individuo no observable que estaría incluido en el error. En los segundos, se incorpora otra fuente de error dada por el efecto 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 que se dan entre las mismas. También puede ser necesario tener que formular varias ecuaciones, para captar comportamientos propios de cada individuo, como por ejemplo, habilidades personales que no pueden ser capturadas en modelos uni-ecuacionales.

Los modelos de ecuaciones simultáneas para datos de panel (Hsiao,1986 y Baltagi,1996) se caracterizan por la presencia de variables endógenas (respuesta) entre las variables explicativas de algunas o de todas las ecuaciones y la presencia de correlación de algunos regresores con los errores de la ecuación de regresión. Este último efecto, denominado endogeneidad de los regresores, puede causar serios problemas de inferencia en econometría. Los estimadores de los parámetros utilizando mínimos cuadrados ordinarios resultan inconsistentes y, por eso, se requieren otros métodos de estimación con variables instrumentales, como el de mínimos cuadrados a dos o tres etapas.

En este trabajo en la sección 2 se realiza una breve descripción de los modelos uniecuacionales y sistemas de ecuaciones para datos de panel cuando la variable respuesta de interés es continua y se presentan métodos de estimación que brindan estimadores eficientes y asintóticamente consistentes. En la sección 3 se muestran dos estudios realizados con datos de la EPH y por último en la sección 4 se mencionan algunas consideraciones finales.

II. Modelos lineales para datos de panel

Los modelos lineales para datos de panel utilizan, en general, una estructura de error que posee varios componentes. El modelo en forma matricial se expresa como

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \qquad \text{(II.1)}$$

donde

y : es el vector de dimensión NT x 1 de las respuestas,

$\mathbf{Z} = [\mathbf{1}_{NT}\, \mathbf{X}]$: es la matriz de diseño, siendo **X** una matriz de dimensión NTxK de las variables explicativas y $\mathbf{1}_{NT}$ un vector de unos,

$\boldsymbol{\delta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}')$: es el vector de los parámetros de dimensión (K+1) x 1 y

u : es el término de error de dimensión NT x 1.

De acuerdo al número de elementos que forman el error se tienen los modelos con componente de error a uno y dos criterios.

- ***Componente de error a un criterio***

La estructura de los disturbios es

$$\mathbf{u} = \mathbf{Z}_{\mu}\boldsymbol{\mu} + \mathbf{v} \ , \qquad \text{(II.2)}$$

donde

$\boldsymbol{\mu}' = (\mu_1, \mu_2, ..., \mu_N)$ denota los efectos específicos no observables del individuo que no son explicados por la regresión y se los considera invariantes en el tiempo,

$\mathbf{v}' = (v_{11}, v_{12}, ..., v_{1T}, ..., v_{N1}, ..., v_{NT})$ representa los restantes errores, que varían con el individuo (i=1,...,N) y el tiempo (j=1,...,T) y se los puede pensar como el error usual de la regresión,

$\mathbf{Z}_{\mu} = (\mathbf{I}_N \otimes \ell_T)$, $\mathbf{I}_N$: matriz identidad de dimensión N, ℓ_T : vector de unos de dimensión T.

- ***Componente de error a dos criterios***

Los disturbios de este tipo se expresan como

$$\mathbf{u} = \mathbf{Z}_{\mu}\boldsymbol{\mu} + \mathbf{Z}_{\lambda}\boldsymbol{\lambda} + \mathbf{v} , \qquad \text{(II.3)}$$

donde $\mathbf{Z}_{\mu}$, $\boldsymbol{\mu}$ y $\mathbf{v}$ se definieron anteriormente,

$\boldsymbol{\lambda}$ denota el efecto tiempo no observable e invariante para los individuos, explica todos los otros efectos tiempo no incluidos en la regresión,

$\mathbf{Z}_{\lambda} = (\mathbf{I}_N \otimes \ell_T)$, con $\mathbf{I}_N$ y ℓ_T definidos anteriormente.

De acuerdo a las características de las componentes del término de disturbios los modelos se clasifican en:

- ***Modelos a efectos fijos***

En este caso los μ_j y λ_j se suponen parámetros fijos a estimar y el resto de los errores aleatorios con v_{it} independiente e idénticamente distribuídos $N(0, \sigma_v^2)$. Estos modelos son útiles cuando el interés se centra en el análisis de un conjunto específico de individuos o unidades y la inferencia se restringe solamente a ese conjunto.

♦ ***Modelo a efectos aleatorios***

Estos modelos constituyen una especificación apropiada cuando los individuos representan una muestra de una población acerca de la cual se desean hacer inferencias. En este caso N es grande y los modelos a efectos fijos producirán una gran pérdida de grados de libertad. Los efectos individuales se caracterizan como aleatorios y la inferencia pertenece a la población a partir de la cual se extrajo la muestra.

La matriz de covariancias para el modelo a un criterio resulta

$$\mathbf{\Omega} = E(\mathbf{u}\,\mathbf{u}') = \sigma_\mu^2\,(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_v^2\,(\mathbf{I}_N \otimes \mathbf{I}_T)\,. \qquad \text{(II.4)}$$

Esto implica homocedasticidad de la variancia de los errores para todo i y t, como así también una matriz en bloque diagonal equicorrelacionada la cual presenta sólo correlación serial sobre el tiempo entre los errores del mismo individuo. Es decir,

$$\begin{aligned} Cov(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_v^2, \quad i = j, t = s \\ Cov(u_{it}, u_{js}) &= \sigma_\mu^2, \quad i = j, t \neq s \end{aligned}, \qquad \text{(II.5)}$$

y cero en otro caso.

Para el modelo a dos criterios la matriz de covariancias es

$$\mathbf{\Omega} = E(\mathbf{u}\,\mathbf{u}') = \sigma_\mu^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\lambda^2(\mathbf{J}_N \otimes \mathbf{I}_T) + \sigma_v^2(\mathbf{I}_N \otimes \mathbf{I}_T)\,. \qquad \text{(II.6)}$$

Los disturbios son homocedásticos con variancia

$Var(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2, \quad \forall i, t$ y covariancias

$$\begin{aligned} Cov(u_{it}, u_{js}) &= \sigma_\mu^2, \quad i = j, t \neq s \\ Cov(u_{it}, u_{js}) &= \sigma_\lambda^2, \quad i \neq j, t = s \end{aligned} \qquad \text{(II.7)}$$

y cero en otro caso.

II.1 Inferencia

♦ ***Modelo a efectos fijos***

La inferencia es condicional a las unidades o individuos particulares que se observan. Para el modelo con componente de error a un criterio se sustituyen los errores en el modelo resultando

$$\mathbf{y} = \alpha\ell_{NT} + \mathbf{X\beta} + \mathbf{Z}_\mu\boldsymbol{\mu} + \mathbf{v} = \mathbf{Z\delta} + \mathbf{Z}_\mu\boldsymbol{\mu} + \mathbf{v} \qquad \text{(II.1.1)}$$

y se aplica el método de mínimos cuadrados ordinarios obteniendo los estimadores de $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ y $\boldsymbol{\mu}$. La dimensión de las matrices, definida anteriormente, depende de N, T y el número de covariables. Si el número de individuos (N) es grande, la matriz a invertir es de gran dimensión.

Una forma más simple de encontrar los estimadores de $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$ se logra premultiplicando (II.1.1) por la matriz de desviaciones del individuo respecto de la media $\mathbf{Q} = \mathbf{I}_{NT} - \mathbf{P}$, siendo $\mathbf{P} = \mathbf{I}_N \otimes \overline{\mathbf{J}}_T$ matriz que promedia las observaciones a través del tiempo de cada individuo y $\overline{\mathbf{J}}_T = \mathbf{J}_T / T$, y sobre el modelo transformado aplicar mínimos cuadrados ordinarios

$$\mathbf{Qy} = \mathbf{QX\beta} + \mathbf{Qv}$$
$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{v}}$$

con término genérico

$$(y_{it} - \bar{y}_{i.}) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})\,\boldsymbol{\beta} + (v_{it} - \bar{v}_{i.}), \qquad \text{(II.2.2)}$$

resultando el estimador $\tilde{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}\tilde{\mathbf{y}}$.

Esta matriz **Q** elimina los efectos individuales o las variables invariantes en el tiempo.

La estadística para probar la hipótesis $H_0)\mu_1 = \ldots = \mu_{N-1} = 0$ es

$$F = \frac{(SSER - SSEC)/(N-1)}{SSEC/(NT - N - K)} \sim F_{N-1, N(T-1)-K} \text{, si se cumple } H_0,$$

siendo:

SSER la suma de cuadrados del error del modelo reducido o bajo la hipótesis nula y
SSEC la suma de cuadrados del error del modelo completo o sin restricción.
Este test se conoce como Test de Chow.

Para el modelo con componente de error a dos criterios el procedimiento es el mismo que se usó anteriormente. No se necesita invertir una gran matriz. Los estimadores de los efectos fijos $\boldsymbol{\beta}$ se pueden obtener aplicando la transformación intra a través de la matriz **Q** que elimina los efectos $\boldsymbol{\mu}$ y $\boldsymbol{\lambda}$. Después de la transformación se realiza la regresión de $\tilde{\mathbf{y}} = \mathbf{Qy}$ versus $\tilde{\mathbf{X}} = \mathbf{QX}$ obteniéndose el estimador intra

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X'QX})^{-1}\mathbf{X'Qy}\,.$$

Un término genérico del modelo transformado es

$$(y_{it} - y_{i.} - y_{.t} + y_{..}) = (\mathbf{x}_{it} - \mathbf{x}_{i.} - \mathbf{x}_{.t} + \mathbf{x}_{..})\,\boldsymbol{\beta} + (v_{it} - v_{i.} - v_{.t} + v_{..}) \quad \text{(II.2.3)}$$

Se observa que el estimador intra no puede estimar el efecto de las variables del individuo que son invariantes en el tiempo, porque la transformación **Q** las elimina.

Como en el modelo a un criterio, se puede probar la siguiente hipótesis:

$$\mathrm{H}_0)\ \mu_1 = \ldots = \mu_{N-1} = 0 \ \text{ y } \ \lambda_1 = \ldots = \lambda_{T-1} = 0$$

La suma de cuadrados del error del modelo reducido (SCER) se obtiene de amalgamar los estimadores, mientras que la del modelo completo (SCEC) es la que se obtiene a partir de la regresión intra en (II.2.3). En este caso

$$F = \frac{(SSER - SSEC)/(N+T-2)}{SSEC/(N-1)(T-1)-K} \sim F_{N+T-2,(N-1)(T-1)-K}$$, si se cumple H_0.

Luego se puede probar la existencia de los efectos individuales dado los efectos tiempo, es decir,

$$\mathrm{H}_0)\ \mu_1 = \ldots = \mu_{N-1} = 0 \quad \text{dado } \lambda_t \neq 0, t = 1,..,T-1\,.$$

Similarmente, se puede probar la existencia de los efectos tiempo dado los efectos individuos,

$$H_0)\ \lambda_1 = \ldots = \lambda_{T-1} = 0 \quad \text{dado } \mu_i \neq 0, i = 1,..,N-1.$$

- ***Modelos a efectos aleatorios***

Para obtener los estimadores de los parámetros de interés Fuller y Battese (1973,1974) sugieren premultiplicar la ecuación (II.1) por $\sigma_v \boldsymbol{\Omega}^{-1/2} = \mathbf{Q} + (\frac{\sigma_v}{\sigma_1})\mathbf{P}$,

y aplicar mínimos cuadrados ordinarios sobre el modelo transformado. La matriz de covariancias se puede expresar como

$$\boldsymbol{\Omega} = \sigma_1^2\ \mathbf{P} + \sigma_v^2\ \mathbf{Q}$$

siendo, $\sigma_1^2 = T\,\sigma_\mu^2 + \sigma_v^2$.

En este caso, $\mathbf{y}^* = \sigma_v \boldsymbol{\Omega}^{-1/2}\mathbf{y}$ tiene elemento genérico $y_{it} - (1 - (\sigma_v / \sigma_1))\overline{y}_{i.}$.

Para obtener los estimadores de las componentes de variancia, Swamy y Arora (1972) sugieren realizar dos regresiones. La primera es la regresión "intra", dada en (II.2.2), la cual produce

$$\hat{\sigma}_v^2 = \left[\mathbf{y'Qy} - \mathbf{y'QX}(\mathbf{X'QX})^{-1}\mathbf{X'Qy}\right] / (N(T-1) - K).$$

La segunda es la regresión "entre" la cual realiza la regresión de los promedios a través del tiempo. Se logra premultiplicando el modelo (II.2.1) por **P** y luego aplicando mínimos cuadrados ordinarios al modelo transformado. La estimación de la variancia es

$$\hat{\sigma}_1^2 = \left[\mathbf{y'Py} - \mathbf{y'PZ}(\mathbf{Z'PZ})^{-1}\mathbf{Z'Py}\right] / (N - K - 1).$$

Para el caso del modelo de componente de error a dos criterios el procedimiento es el mismo.

II.1 Sistema de ecuaciones para datos de panel

Sea el sistema de ecuaciones definido como

$$\mathbf{y} = \mathbf{Z\delta} + \mathbf{u} \tag{II.1.1}$$

donde

$\mathbf{y}' = (y_1^{'},\ldots,y_M^{'})$,

$\mathbf{Z} = diag\lfloor \mathbf{Z}_j \rfloor$,

$\mathbf{Z}_j = \lfloor \mathbf{y}_j, \mathbf{x}_j \rfloor$, de dimensión NTx(g_j+k_j), j = 1,..., M, g_j :tamaño de $\boldsymbol{y}_j$ del segundo miembro, k_j:tamaño de $\mathbf{X}_j$,

$\boldsymbol{\delta}' = (\boldsymbol{\delta}_1^{'},\ldots,\boldsymbol{\delta}_M^{'})$,

$\mathbf{u}'=(\mathbf{u}'_1,\ldots, \mathbf{u}'_M)$.

La diferencia entre este modelo con un modelo de sistema aparentemente no correlacionado (SUR), es que en los segundos miembros de las ecuaciones del primero aparecen variables endógenas.

Para un modelo de componente de error a un criterio, el error de la j-ésima ecuación es

$$\mathbf{u}_j = \mathbf{Z}_\mu \boldsymbol{\mu}_j + \mathbf{v}_j \qquad j = 1,\ldots,M \ , \tag{II.1.2}$$

donde

$\mathbf{Z}_\mu = (\mathbf{I}_N \otimes \ell_T)$, $\mathbf{I}_N$:matriz identidad de dimensión N, ℓ_T: vector de unos de dimensión T,

$\boldsymbol{\mu}_j^{'} = (\mu_{1j},\mu_{2j},\ldots,\mu_{Nj})$ y $\mathbf{v}_j^{'} = (v_{11j},v_{12j},\ldots,v_{1Tj},\ldots,v_{N1j},\ldots,v_{NTj})$ son vectores aleatorios con media cero y matriz de variancias y covariancias

$$E\begin{bmatrix}\boldsymbol{\mu}_j \\ \mathbf{v}_j\end{bmatrix}(\boldsymbol{\mu}_l^{'}\mathbf{v}_l^{'}) = \begin{bmatrix}\sigma_{\mu_{jl}}^2 I_N & 0 \\ 0 & \sigma_{v_{jl}}^2 I_{NT}\end{bmatrix} \quad \text{j,l =1,...,M.}$$

Siguiendo los supuestos clásicos para SUR (Zellner, 1962), se puede mostrar que cada componente de error tiene

$$\boldsymbol{\mu} \sim (\mathbf{0}, \boldsymbol{\Sigma}_\mu \otimes \mathbf{I}_N) \text{ y } \mathbf{v} \sim (\mathbf{0}, \boldsymbol{\Sigma}_v \otimes \mathbf{I}_{NT}),$$

donde $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_M'), \mathbf{v}' = (\mathbf{v}_1', ..., \mathbf{v}_M')$, $\boldsymbol{\Sigma}_\mu = \lfloor \sigma^2_{\mu_{jl}} \rfloor$ y $\boldsymbol{\Sigma}_v = \lfloor \sigma^2_{v_{jl}} \rfloor$ para j = 1,..., M.

Usando (II.1.2) sigue que

$$\boldsymbol{\Omega}_{jl} = E(\mathbf{u}_j \mathbf{u}_l') = \sigma^2_{\mu_{jl}} (\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma^2_{v_{jl}} (\mathbf{I}_N \otimes \mathbf{I}_T). \qquad \text{(II.1.3)}$$

siendo $\boldsymbol{J}_T$ una matriz de unos de dimensión T.

En este caso, la matriz de covariancia entre errores de diferentes ecuaciones tiene la misma forma que la del componente de error a un criterio, pero ahora, se tienen los componentes de variancia de ecuaciones cruzadas, los cuales se deben estimar.

La matriz de variancias y covariancias para el conjunto de M ecuaciones, está dada por

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}(\mathbf{u}_j \mathbf{u}_l') = \boldsymbol{\Sigma}_\mu \otimes (\mathbf{I}_N \otimes \mathbf{J}_T) + \boldsymbol{\Sigma}_v \otimes (\mathbf{I}_N \otimes \mathbf{I}_T), \qquad \text{(II.1.4)}$$

donde $\mathbf{u}' = (\mathbf{u}_1', ..., \mathbf{u}_M')$ es un vector de errores de dimensión 1xMNT con $\mathbf{u}_j$ definido en (II.1.2) para j = 1,..., M. $\boldsymbol{\Sigma}_\mu = \lfloor \sigma^2_{\mu_{jl}} \rfloor$ y $\boldsymbol{\Sigma}_v = \lfloor \sigma^2_{v_{jl}} \rfloor$ son ambas matrices MxM.

Reemplazando $\mathbf{J}_T$ por $T\bar{\mathbf{J}}_T$ e $\mathbf{I}_T$ por $\mathbf{E}_T + \bar{\mathbf{J}}_T$ (donde $\bar{\mathbf{J}}_T = \mathbf{J}_T / T$ y $\mathbf{E}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T$) y reagrupando términos se tiene

$$\begin{aligned} \boldsymbol{\Omega} &= (T\boldsymbol{\Sigma}_\mu + \boldsymbol{\Sigma}_v) \otimes (\mathbf{I}_N \otimes \bar{\mathbf{J}}_T) + \boldsymbol{\Sigma}_v \otimes (\mathbf{I}_N \otimes \mathbf{E}_T) \\ &= \boldsymbol{\Sigma}_1 \otimes \mathbf{P} + \boldsymbol{\Sigma}_v \otimes \mathbf{Q}, \end{aligned} \qquad \text{(II.1.5)}$$

donde $\mathbf{\Sigma}_1 = T\mathbf{\Sigma}_\mu + \mathbf{\Sigma}_v$;   y $\mathbf{Q} = \mathbf{I}_{NT} - \mathbf{P}$ (matriz de desviaciones del individuo respecto de la media). **P** y **Q** matrices idempotentes.

La expresión (II.1.5) es la descomposición espectral de $\mathbf{\Omega}$ deducida por Baltagi(1980), por lo que

$$\mathbf{\Omega}^r = \mathbf{\Sigma}_1^r \otimes \mathbf{P} + \mathbf{\Sigma}_v^r \otimes \mathbf{Q}\,, \tag{II.1.6}$$

donde r es un escalar arbitrario. Para r=-1, se tiene la inversa y para r=-(1/2) se obtiene

$$\mathbf{\Omega}^{-\frac{1}{2}} = \mathbf{\Sigma}_1^{-\frac{1}{2}} \otimes \mathbf{P} + \mathbf{\Sigma}_v^{-\frac{1}{2}} \otimes \mathbf{Q}\,. \tag{II.1.7}$$

Kinal y Lahiri(1990) sugieren obtener la descomposición de Cholesky de $\mathbf{\Sigma}_v$ y $\mathbf{\Sigma}_1$ en (II.1.7) para reducir los cálculos y simplificar las transformaciones del sistema.

Se pueden obtener estimadores consistentes de $\mathbf{\Sigma}_v$ y $\mathbf{\Sigma}_1$, usando los residuos que surgen de aplicar MC2EI(intra) y MC2EE(entre), de la siguiente manera

$$\hat{\sigma}_{vjl}^2 = (\mathbf{y}_j - \mathbf{Z}_j\tilde{\boldsymbol{\delta}}_{j,MC2EI})'\mathbf{Q}(\mathbf{y}_l - \mathbf{Z}_l\tilde{\boldsymbol{\delta}}_{l,MC2EI})/N(T-1)\,, \tag{II.1.8}$$

$$\hat{\sigma}_{1_{jl}}^2 = (\mathbf{y}_j - \mathbf{Z}_j\hat{\boldsymbol{\delta}}_{j,MC2EE})'\mathbf{P}(\mathbf{y}_l - \mathbf{Z}_l\hat{\boldsymbol{\delta}}_{l,MC2EE})/N\,. \tag{II.1.9}$$

Se debe chequear si $\hat{\mathbf{\Sigma}}_\mu = (\hat{\mathbf{\Sigma}}_1 - \hat{\mathbf{\Sigma}}_v)/T$ es definida positiva.

Para encontrar los estimadores $\tilde{\boldsymbol{\delta}}_{j,MC2EI}$ y $\hat{\boldsymbol{\delta}}_{j,MC2EE}$ se siguen los siguientes pasos:

a) Se transforma (II.1.1) mediante

$$\mathbf{Qy}_j = \mathbf{QZ}_j\boldsymbol{\delta}_j + \mathbf{Qu}_j\,,\ \ j = 1, ..., M; \tag{II.1.10}$$

b) Se toma $\tilde{\mathbf{y}}_j = \mathbf{Q}\mathbf{y}_j$ y $\tilde{\mathbf{Z}}_j = \mathbf{Q}\mathbf{Z}_j$ y se desarrolla MC2E sobre (II.1.10), con $\tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X}$ como instrumento, se obtiene el estimador MC2E intra

$$\tilde{\boldsymbol{\delta}}_{j,MC2EI} = (\tilde{\mathbf{Z}}_j^{'}\mathbf{P}_{\tilde{X}}\tilde{\mathbf{Z}}_j)^{-1}\tilde{\mathbf{Z}}_j^{'}\mathbf{P}_{\tilde{X}}\mathbf{y}_j, \qquad \text{(II.1.11)}$$

con Var($\tilde{\delta}_{j,MC2EI}$)= $\sigma^2_{v_{11}}(\tilde{\mathbf{Z}}_j^{'}\mathbf{P}_{\tilde{X}}\tilde{\mathbf{Z}}_j)^{-1}$;

c) Similarmente se puede transformar (II.1.1) haciendo $\bar{\mathbf{y}}_j = \mathbf{P}\mathbf{y}_j$ y $\bar{\mathbf{Z}}_j = \mathbf{P}\mathbf{Z}_j$;

d) Se desarrolla MC2E con $\bar{\mathbf{X}} = \mathbf{P}\mathbf{X}$ como instrumento y se obtiene el estimador MC2E entre

$$\hat{\boldsymbol{\delta}}_{j,MC2EE} = (\bar{\mathbf{Z}}_j^{'}\mathbf{P}_{\bar{X}}\bar{\mathbf{Z}}_j)^{-1}\bar{\mathbf{Z}}_j^{'}\mathbf{P}_{\bar{X}}\bar{\mathbf{y}}_j, \qquad \text{(II.1.12)}$$

con var($\hat{\delta}_{j,MC2EE}$) $= \sigma^2_{j,11}(\bar{\mathbf{Z}}_j^{'}\mathbf{P}_{\bar{X}}\bar{\mathbf{Z}}_j)^{-1}$ donde $\sigma^2_{j,11} = T\sigma^2_{\mu_{11}} + \sigma^2_{v_{11}}$.

Por último para encontrar el estimador MC3E eficiente, se transforma (II.1.1) usando $\mathbf{\Omega}^{-\frac{1}{2}}$ de (II.1.7), obteniendo

$$\mathbf{y}^* = \mathbf{Z}^*\boldsymbol{\delta} + \mathbf{u}^*, \qquad \text{(II.1.13)}$$

con $\mathbf{y}^* = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{y}$; $\mathbf{Z}^* = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{Z}$ y $\mathbf{u}^* = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{u}$. Para un conjunto arbitrario de instrumento **A**, el estimador MC3E de (II.1.13) resulta

$$\hat{\boldsymbol{\delta}}_{MC3E} = (\mathbf{Z}^{*'}\mathbf{P}_A\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\mathbf{P}_A\mathbf{y}^*. \qquad \text{(II.1.14)}$$

Usando los resultados de White(1986), el conjunto de instrumento óptimo es

$$\mathbf{X}^* = \mathbf{\Omega}^{-\frac{1}{2}}(\mathbf{I}_2 \otimes \mathbf{X}) = (\mathbf{\Sigma}_v^{-\frac{1}{2}} \otimes \mathbf{QX}) + (\mathbf{\Sigma}_1^{-\frac{1}{2}} \otimes \mathbf{PX}). \qquad \text{(II.1.15)}$$

Tomando **A**=**X*** se obtiene el estimador mínimo cuadrático tres etapas eficiente (MC3EEFI)

$$\hat{\boldsymbol{\delta}}_{MC3EEFI} = (\mathbf{Z}^{*\prime}\mathbf{P}_{x^*}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{P}_{X^*}\mathbf{y}^* . \qquad \text{(II.1.16)}$$

Baltagi (1981, b) muestra que el estimador MC3EEFI se reduce al MC2EEFI cuando los errores de las diferentes ecuaciones estructurales no están correlacionados con los de las otras, pero no necesariamente, cuando todas las ecuaciones están exactamente identificadas. Esto es diferente de las condiciones análogas entre estimadores MC2E y MC3E en modelos clásicos de ecuaciones simultáneas.

III. Aplicaciones

Aplicación 1:

Estudios realizados en varios países y regiones del mundo sugieren la existencia de una relación negativa entre la participación laboral y el desempleo, tanto en niveles como en variaciones (Summers y Clark, 1990; Elmeskov y Pichelman, 1993, Elhorst, 1996). En Argentina, por el contrario, la mera inspección de los datos sugiere que la relación entre ambas tasas *puede ser positiva*.

En un trabajo anterior (Pellegrini, Blaconá, García y Ventroni, 1999) se sostuvo que las variaciones de la participación laboral se explican por las variaciones del empleo y de la desocupación. Pero mientras la relación con la primera de estas dos variables debe ser unívocamente positiva, no ocurre lo mismo con la segunda, ya que operan dos efectos de signo contrario (los efectos trabajador desalentado y trabajador adicional) y el signo de la relación depende del que prevalezca.

Para detectar el signo y la magnitud de las interacciones en el mercado de trabajo, en esta aplicación se plantea un modelo de ecuaciones simultáneas para datos de panel para representar las variaciones de la tasa de desocupación y la tasa de actividad en la República Argentina.

Para estudiar el mercado de trabajo en la República Argentina se tienen en cuenta: la tasa de desocupación (número de desocupados dividido la población activa por 100); la tasa de actividad (la población activa divida la población total por 100) y la tasa de empleo (número de ocupados divididos la población total por 100).
Dichas tasas se calculan a partir de la EPH, para 25 aglomerados urbanos en el período 1985 - 1996 (24 ondas en 25 aglomerados). Se cuenta con 594 observaciones para cada tasa, a las que se agregan 6 estimaciones correspondientes a datos faltantes de diferentes aglomerados, cuyos valores se imputan mediante un modelo autorregresivo de orden uno.

El conjunto de datos disponibles para el análisis tiene la estructura de un conjunto de observaciones repetidas sobre cada aglomerado.

Se desea describir el comportamiento del mercado de trabajo en la República Argentina, por lo cual se cree conveniente usar un sistema de ecuaciones, el cual permite captar las interrelaciones que existen en dicho mercado. En este caso, los modelos de datos de panel para una única ecuación, por lo general no resultan satisfactorios, por la misma razón que los estimadores mínimos cuadráticos ordinarios tienen problemas de eficiencia y consistencia en el caso de sistemas de ecuaciones, que no tienen en cuenta datos de panel.

El requisito básico que deben satisfacer los modelos de sistemas de ecuaciones económicos, es que el número de variables que se pretende explicar sea igual al número de relaciones independientes en el modelo, pues de lo contrario los valores de estas variables no se podrían determinar. Estos modelos, además, pueden contener variables cuyos valores no se ven afectados de inmediato por el mecanismo presentado en el modelo. La importancia de estas variables radica en su función como factores explicativos.

Se debe hacer una distinción entre las variables cuyos valores se desean explicar por el modelo y las variables que contribuyen a proporcionar esta explicación; las primeras se llaman endógenas y las segundas predeterminadas, las que pueden subdividirse en exógenas y endógenas rezagadas. Los valores de las variables exógenas se determinan fuera del sistema, mientras que las otras están representadas por los valores previos de las variables endógenas del modelo.

El sistema para el caso de mercado de trabajo consta de dos ecuaciones, a saber:

$$\begin{cases} \mathbf{y}_1 = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{y}_2 + \mathbf{u}_1 \\ \mathbf{y}_2 = \beta_3 + \beta_4 \mathbf{x}_2 + \beta_5 \mathbf{x}_3 + \beta_6 \mathbf{y}_1 + \mathbf{u}_2 \end{cases} \tag{III.1}$$

donde

las *variables endógenas* representan:

$\mathbf{y}_1$: el incremento del logaritmo de la tasa de desocupación en el momento t ,

$\mathbf{y}_2$: el incremento del logaritmo de la tasa de actividad en el momento t;

las *variables exógenas* representan:

$\mathbf{x}_1$: el incremento del logaritmo de la tasa de desocupación en el momento t-1,

$\mathbf{x}_2$: el incremento del logaritmo de la tasa de actividad en el momento t-1,

$\mathbf{x}_3$: el incremento del logaritmo de la tasa de empleo en el momento t;

β_i: i= 1-6 parámetros a estimar y

$\mathbf{u}_1$ y $\mathbf{u}_2$ son los errores de la primera y segunda ecuación respectivamente (cada uno vector de dimensión 550x1).

Todas la variables constituyen vectores de dimensión 550x1, dado que se cuenta con 25 (N) aglomerados medidos en 22 (T) momentos[1].

Se trabaja con el incremento del logaritmo de las tasas para evitar el problema de raíces unitarias que se presenta en algunos aglomerados para algunas tasas, lo que implica que los estimadores no sean adecuados. Por otro lado, trabajar con esta transformación de las tasas, no trae problemas de interpretación, ya que en esta forma se están modelando las variaciones porcentuales de dichas tasas.

Como se dijo anteriormente, el hecho de tener datos de panel y un sistema de ecuaciones donde en la primera ecuación aparece la segunda variable endógena como explicativa, hace que los tradicionales estimadores mínimos cuadráticos ordinarios no sean consistentes esencialmente por dos motivos: a) por la estructura de correlación que puede existir entre observaciones de un mismo aglomerado y b) porque puede existir correlación, en la primer ecuación entre el error e $\mathbf{y}_2$ y en la segunda entre el error e $\mathbf{y}_1$.
Para resolver estos problemas se usa un estimador mínimo cuadrático a tres etapas para datos de panel, propuesto por Baltagi (1995).

Se estima el modelo (III.1) por diferentes métodos para mostrar como varían las estimaciones cuando se usan aquéllos que no tiene en cuenta todas las particularidades que tienen los datos, en especial la de constituir una estructura de datos de panel y la de ser un sistema de ecuaciones.

Como se dijo en el punto II, el método apropiado de estimación es el de mínimos cuadrados en tres etapas eficiente (MC3EEFI) propuesto por Baltagi (1995). Para realizar esta estimación se realiza un programa en SAS combinando el procedimiento SYSLYN para calcular los estimadores MC2E y MC3E y el procedimiento IML para realizar todas las transformaciones y cálculos de matrices de variancias y covariancias siguiendo los pasos enunciados en el punto (II.1).

Los dos métodos de estimación inapropiados son: i) mínimos cuadrados a tres etapas que tiene en cuenta el sistema de ecuaciones pero no la estructura de datos de panel (se realiza mediante el procedimiento SYSLIN de SAS) y ii) estimar por separado cada ecuación del sistema por un método que contempla la estructura longitudinal de los datos (mediante el procedimiento GENMOD de SAS).

En la Tabla 3.1 se observan los cambios que se producen en los coeficientes cuando se aplican los tres métodos. En especial en la ecuación uno, en la cual interviene la variable endógena $\boldsymbol{y}_2$ como variable explicativa, se ve que el coeficiente cambia sustancialmente, incluso de signo, cuando se estima por cualquiera de los métodos de estimación en tres etapas en lugar de MLG. Esto se debe principalmente a la correlación que existe entre la variable explicativa y los residuos (Tabla 3.2), hecho que no contempla el método MLG. Al comparar el resultado del mismo coeficiente MC3E se aprecia que el estimador MC3EEFI es mucho más chico que el MC3E, esto se debe a que este último no tiene en cuenta la existencia de correlación entre las medidas repetidas de cada aglomerado.

Tabla 3.1 Estimación del Modelo II.1

[1] Si bien inicialmente se cuenta con 24 repeticiones para cada aglomerado, se pierden dos, una por trabajar con los incrementos de los logaritmos y otra por usar como variables exógenas a variables rezagadas un período.

Ecuación	Regresor	Método de Estimación					
		MC3EEFI		MC3E		MLG	
		Coef.	P	Coef.	p	Coef.	p
$\boldsymbol{y}_1$	Constante	0.5464	0.0001	0.0243	0.011	0.0182	0.0001
	$\boldsymbol{x}_1$	-0.1442	0.0001	-0.4770	0.0001	-0.3572	0.0001
	$\boldsymbol{y}_2$	-0.6789	0.0001	-5.1800	0.0001	1.7245	0.0001
$\boldsymbol{y}_2$	Constante	0.2957	0.0005	0.0005	0.0103	0.0004	0.0330
	$\boldsymbol{x}_2$	-0.1593	0.0001	-0.0721	0.0001	-0.065	0.0001
	$\boldsymbol{x}_3$	1.2823	0.0001	0.8914	0.0001	0.9368	0.0001
	$\boldsymbol{y}_1$	1.3572	0.0001	0.0643	0.0001	0.0761	0.0001

En la Tabla 3.1 también se puede observar que el coeficiente que menos cambia al estimarlo por los tres métodos es el que corresponde a $\boldsymbol{x}_3$ de la segunda ecuación, variable exógena que no presenta correlación con los residuos (Tabla 3.2). En esta ecuación también cambia sustancialmente el coeficiente de la variable endógena ($\boldsymbol{y}_1$).

Tabla 3.2: Correlaciones residuos versus variables explicativas

Residuos	Correlación	p	Correlación	p	Correlación	p
Ecuación 1	$\boldsymbol{y}_2$		$\boldsymbol{y}_1$			
	0.6336	0.0001	-0.1909	0.0001		
Ecuación 2	$\boldsymbol{x}_2$		$\boldsymbol{x}_3$		$\boldsymbol{y}_1$	
	-0.0235	0.5818	-0.0003	0.9947	-0.2229	0.0001

De acuerdo al modelo estimado eficientemente, se puede decir 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 con las variaciones del logaritmo de la tasa de empleo y de la desocupación.

Aplicación 2:

(Becker, 1974).

Esta aplicación, que está en su faz inicial, analiza el ingreso del varón, para aquellos hogares donde cohabita una pareja conyugal[i], teniendo en cuenta las diferencias entre los niveles de escolarización de los cónyuges.

En numerosas poblaciones estudiadas se observó que ciertas características de los integrantes de cada pareja conyugal están correlacionadas. Así, la edad, educación, inteligencia, religión, etnia, procedencia geográfica y hasta la altura están positivamente correlacionadas, mientras que algunas características de personalidad, como el carácter dominante, agresivo o tranquilo parecen mostrar correlaciones negativas (Beker, 1983; Epstein y Gutman, citados por Lamb, 1988). La asociación de características en las parejas puede tener importantes consecuencias económicas, afectando los ingresos individuales, y asumiendo una responsabilidad primaria en el papel de la familia como determinante de la posición económica y social de sus integrantes

La educación, una variable que afecta críticamente a los ingresos por su papel en la formación del capital humano, es una de la que presentan correlaciones positivas más fuertes, y los ingresos mismos están correlacionados con signo positivo, una vez controlados por otras covariables. Sin embargo, este es un punto sujeto a controversia (Lamb, 1988), porque la autoselección en el mercado de trabajo puede alejar de la fuerza laboral a las mujeres de ingresos potencialmente bajos casadas con varones de ingresos más elevados, y de esta forma, convertir en positiva una correlación que de otro modo sería negativa (Becker, 1973). Behrman *et al* (1994) también llegan a la conclusión de que la correlación es en realidad negativa.

La explicación más generalmente aceptada entre los economistas para aquellos fenómenos se funda en la hipótesis de que la formación de parejas está sujeta a un proceso de selección a partir del ordenamiento de los potenciales cónyuges según sus características. Las uniones basadas en la asociación de características (*assortative matting*[ii]) maximizan, en condiciones de equilibrio del mercado matrimonial, el producto doméstico total del conjunto de las parejas constituidas (Becker, 1973, 1974).

En las primeras aplicaciones de aquel marco analítico todas las características llevadas por las personas al mercado matrimonial fueron tratadas como exógenas. Boulier y Rosenzweig (1984) propusieron un esquema alternativo, en el cual las inversiones en educación, la búsqueda de pareja y la formación de las uniones conyugales son respuestas a las condiciones del mercado matrimonial y a las características personales de los individuos, algunas de los cuales son inobservables para el investigador.

Para aquellos autores, la educación es valiosa incluso cuando la participación laboral es baja en un marco de fuerte división sexual del trabajo entre actividades de mercado y fuera del mercado, porque los niveles educativos de ambos esposos son factores complementarios en la función de producción doméstica.

Sin embargo, el papel de la educación femenina puede ir más allá de esa complementariedad señalada. Si la dotación de capital humano afecta la eficiencia en la producción doméstica (Becker, 1983), también influiría en el flujo de servicios de apoyo que las esposas proporcionan a sus esposos. Así, los niveles educativos de ambos cónyuges pueden ser *también* factores complementarios en las funciones de producción para el mercado, y en consecuencia, afectar los ingresos que los maridos obtienen en sus ocupaciones. La participación laboral de la esposa sería un factor contrarrestante, debido a que tendería a reducir la provisión de aquellos servicios de apoyo (Jacobsen y Rayack, 1996).

Si como sostuvieron Boulier y Rosenzweig en el trabajo ya citado, la escolarización y la búsqueda de pareja son variables endógenas influenciadas directa o indirectamente por las condiciones del mercado matrimonial, una mayor escolarización sería un factor capaz compensar características propias menos deseables, o de contribuir a la obtención de una pareja más atractiva.

La proposición de que las características complementarias se asocian *positivamente* en las parejas es central para la teoría económica del matrimonio. Sin embargo, la asociación no tiene por qué ser perfecta para ninguna característica en particular, en virtud de la heterogeneidad de los individuos respecto de las demás. De hecho, debe esperarse que una proporción de parejas esté formada por personas con diferentes niveles de educación, y si hay una relativa paridad entre ambos sexos en cuanto a escolarización, estas parejas mostrarían diferencias de signo tanto positivo como negativo.

En condiciones de participación laboral relativamente baja de las mujeres casadas, las características asociadas a la capacidad de un hombre para obtener ingresos son seguramente muy importantes. Bajo una hipótesis análoga a la de Boulier y Rosenzweig (1984), el nivel de ingresos del marido podría estar relacionado con un diferencial entre los niveles de escolarización propios y los de su esposa, debido, al menos parcialmente, a características no observadas de aquél que tienen un valor en el mercado matrimonial *porque* lo tienen en el mercado de trabajo (o, tal vez, debido a características que tienen un valor en el mercado matrimonial *tanto como* en el de trabajo).

De este modo, pueden plantearse al menos dos hipótesis alternativas sobre el origen de una posible relación entre los niveles de ingresos de mercado del marido y las diferencias de escolarización entre ambos miembros de la pareja. La primera de ellas se basa en la complementariedad de características que entran como factores en la función de producción para el mercado, y la segunda, 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.

En ambos casos, un mayor ingreso del marido debería estar asociado a un menor valor de la diferencia entre su nivel de educación y el de su esposa (*i.e.* hombres de ingresos relativamente alto, dadas sus demás características, pueden estar casados con mujeres más educados que ellos mismos, y en consecuencia, se observaría una diferencia de signo negativo).

Un criterio que permitiría diferenciar una hipótesis de otra estaría dado por la participación laboral de las esposas. Según la hipótesis *de la productividad*, los maridos con esposas económicamente activas deberían tener ingresos significativamente menores que aquellos que las tienen inactivas, pero esto no tiene por qué ocurrir en el caso de que se sostuviera la hipótesis *de los procesos de búsqueda*.

El estudio empírico se realiza sobre datos de hogares del Aglomerados urbanos del país, relevados entre las ondas de mayo de 1998 a octubre 1999 por la Encuesta Permanente de Hogares, en los que se registró la presencia de dos personas, de diferente sexo, clasificadas como jefe de hogar y cónyuge, a los que se supuso marido y esposa, según correspondiera a su sexo.

Para construir el modelo de sistema de ecuaciones se tienen en cuenta las variables del Cuadro 3.1.

Cuadro Nº 3.1

Variables consideradas en los modelos

Variables endógenas
Y_1 : logaritmo ingresos esposo,
Y_2: diferencia de escolaridad entre esposo y esposa (escolaridad esposo – escolaridad esposa).
Variables exógenas
X_1: edad del esposo
X_2: edad del esposo al cuadrado,
X_3: escolaridad del esposo,
X_4: total horas trabajada en la última semana por el esposo,
X_5: antigüedad en el trabajo del esposo,
X_6: número de hijos del esposo menores de 18 años,
X_7: cantidad de ocupaciones,
X_8: variable "dummy" de condición de actividad del esposo (igual 1 si es ocupado, cero en otro caso),
X_9: variable "dummy" de condición de actividad del esposo (igual 1 si es desocupado, cero en otro caso),
X_{10}: ingreso de la esposa,
X_{11}: edad de la esposa,
X_{13}: variable "dummy" de condición de actividad de la esposa (igual 1 si es activa, cero en otro caso).

Para cada pareja conyugal fueron definidas dos ecuaciones: una de ingresos del marido y otra de diferencia educativa en la pareja.

Para la primera, se tomó como referencia la conocida función de ingresos del capital humano de Mincer (Willis, 1997). En primer lugar, los niveles educativos incompletos se computaron como el promedio de los años prescriptos[iii], y se adoptó a la edad como *proxy* de la experiencia, debido calcular a ésta del modo convencional (edad actual *menos* años de escolarización *menos* seis) parece un procedimiento artificioso cuando se investiga un medio en el cual la asistencia al sistema educativo formal no es incompatible con la participación en la fuerza de trabajo. En segundo lugar, se tomaron en cuenta otras variables relativas al esposo, detalladas en el Cuadro Nº 3.1, que fueron consideradas susceptibles de influir sobre el ingreso corriente. Finalmente, se incluyen dos variables: una "dummy", que recoge la condición de actividad de la esposa, y otra que representa la diferencia educativa en la pareja, aproximada por la diferencia entre los años de escolarización del marido respecto de los de la esposa.

Considerando que tal diferencia puede ser endógena, se construyó un modelo biecuacional para estimarla simultáneamente con los ingresos del marido. Dada la información disponible, se supuso que ella depende de la edad de cada uno de los cónyuges y del ingreso de la esposa.

El modelo planteado en un principio es el siguiente:

$$\begin{cases} \mathbf{y}_1 = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \beta_3\mathbf{x}_8 + \beta_4\mathbf{x}_4 + \beta_5\mathbf{x}_5 + \beta_6\mathbf{x}_6 + \beta_7\mathbf{x}_7 + \beta_8\mathbf{x}_8 + \beta_9\mathbf{x}_9 + \beta_{10}\mathbf{x}_{10} + \\ \quad \beta_{11}\mathbf{x}_{13} + \beta_{12}\mathbf{y}_2 + \mathbf{u}_1 \\ \qquad \mathbf{y}_2 = \beta_{13} + \beta_{14}\mathbf{x}_1 + \beta_{15}\mathbf{x}_{11} + \beta_{16}\mathbf{x}_{10} + \mathbf{u}_2 \end{cases}$$

$$\mathbf{u} \sim N(\mathbf{0}, \Sigma). \qquad \text{(III.2)}$$

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. En esta aplicación se hacen más evidentes los inconvenientes, mencionados en la sección I, que se presentan al trabajar con datos de panel, por ello se debe ser muy cuidadoso en la construcción de la base de datos.

IV. Consideraciones Finales

Este trabajo muestra como los Modelos de Sistema de Ecuaciones para datos de panel, son una herramienta muy útil para analizar interrelaciones, tanto entre individuos como a través del tiempo, de distintos aspectos que se miden en la EPH.

En la primera aplicación, donde a cada aglomerado urbano se lo considera un individuo, el uso de los modelos propuestos permite determinar el signo de la relaciones 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, considerando como unidad el hogar donde cohabita una pareja conyugal y a la variable respuesta el ingreso del jefe, 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.

Se piensa en el futuro seguir con la aplicación del ingreso hasta conseguir un modelo adecuado que pruebe qué hipótesis, sobre el ingreso de los cónyuges, sería la que se cumple en Argentina y si es la misma para todos los aglomerados urbanos.

V.- Referencias

Baltagi, B. (1995) “Econometric Analysis of Panel Data”, John Wiley & Sons.

Becker, G. (1973): “A Theory of Marriage:Part 1”, *Journal of Political Economy*, Vol.81, Nº 4.

(1974):): “A Theory of Marriage:Part 2”, *Journal of Political Economy*, Vol.82, Nº 2 part.2.

(1983): *El Capital Humano*, Alianza Universidad, Madrid.

(1987): *Tratado sobre la Familia*, Alianza Universidad, Madrid.

Behrman, J., Rosenzweig, M. and Taubman, P. (1994): “Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment”, *Journal of Political Economy,* Vol 102, Nº6, 1131-1175.

Boulier, B. L. and Rosenzweig, M. (1984): “Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior”, *Journal of Political Economy, Vol.92m nº 4.*

Elhorst, J.P. (1996) “A Regional Analysis of Labour Force Participation Rates across the member States of the European Union”. Regional Studies, 30(5),455-465.

Elmeskov, J y Pichelman, K. (1993) "Interpreting Unemployment: the Role of Labour-Force Participation". OECD Economic Studies, Nº21, Winter, 139-160.

Fuller, W.A. y Battese, G.E. (1973) “Transformations for estimation of linear models with nested error structure. *Journal of the American Statistical Association* 68, 626-632.

(1974) “Estimation of linear models with cross-error structure”, *Journal of Econometrics* 2, 67-78.

Hsiao, Cheng (1986) “Analysis of Panel Data”, Econometric Society Monographs Nº11. Cambridge University Press.

INDEC (1998) “Encuesta de Hogares: Reformulación de la Encuesta Permanente de Hogares de Argentina”, INDEC-ISI.

Jacobsen, J. and Rayack, W. (1996): “Do Men Whose Wives Work Realy Earn Less”, *American Economic Review*, Vol 86(2), May, 269-273.

Kinal, T y Lahiri, K.(1990) “A computational algorithm for multiple equation models with panel data”.*Economics Letters* 34, 143-146.

Lamb, D. (1988): “Marriage Markets and Assortative Mating with Household Public Goods”, *The Journal of Human Resources*, XXIII, 4, 462-487.

Pellegrini, J.L.; Blaconá, M.T.; Ventroni, N.; Garcia, M.del C.,(1999) "Las fluctuaciones de la participación laboral en relación con las del empleo y el desempleo en la Argentina". 34º Reunión anual de la Asociación Argentina de Economía Política – volumen 1, pág. 335.

Summers, L.; Clark, K. (1990[a]) "Labour Market Dynamics and Unemployment. A reconsideration". En Summers, L. *Understanding Unemployment.* The MIT Press, Cambridge, Massachusetts y London, England, 3-47.

Swamy, P.A.V.B. y Arora S.S. (1972) “The exact finite sample properties of the estimators of coefficients in the error components regression models”. *Econometrica* 40, 261-275-

Theil, H.(1971) *Principles of Econometrics*, John Wiley, New York.

Willis, R. J. (1997): "Wage Determinants: a Survey and Reinterpretation of Human Capital Earnings Function", en Ashenfelter, O. C. and Layard, R. , *Handbook of Labour Economics,* North Holland, Amsterdam.

Zellner, A.(1962)"An efficient method of estimating seemingly unrelated regression and tests for aggregation bias". *Journal of the American Statistical Association 57,348-368.*

[i] Una pareja conyugal se define como aquella formada por dos personas de diferente sexo que habitan en el mismo hogar y se declaran cónyuges, sin considerar el carácter legal, consensual o casual de su unión.

[ii] La expresión *uniones basadas en la asociación entre características* se adoptó porque, aunque menos parca, parece reflejar mejor el significado de la expresión inglesa que su traducción como "emparejamiento selectivo", adoptada en la edición en castellano, de Alianza Editorial, del Tratado sobre la Familia (Becker, 1987).

[iii] El estudio de cuatro ondas sucesivas de la Encuesta Permanente de Hogares mostró numerosas inconsistencias la variable que recoge los años de escolarización después de haber terminado el último ciclo completo, incluyendo períodos de estudios incompleto mucho más largos que los prescriptos para el ciclo completo. La solución adoptada para el problema supone, además de una distribución normal de las duraciones de los ciclos incompletos, que los logros educativos no dependen tanto de los años cursados como de los niveles alcanzados.