



Blaconá, María Teresa*
García, María del Carmen*
Bussi, Javier**
Ventroni, Nora**

**Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias Económicas y Estadística. Consejo de Investigaciones, Universidad Nacional de Rosario. E-mail: mblacona@agatha.unr.edu.ar*

*** Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario.*

UN ANALISIS ESTADISTICO COMPARATIVO DE LOS AGLOMERADOS CAPITAL FEDERAL Y ROSARIO RESPECTO DE LA FORMACION DE LOS INGRESOS INDIVIDUALES EN LA PAREJA CONYUGAL

I. INTRODUCCIÓN

La información relativa a los individuos que se releva mediante la Encuesta Permanente de Hogares (EPH), que lleva a cabo el Instituto Nacional de Estadísticas y Censos (INDEC), ofrece importantes ventajas para el análisis empírico de la conformación del ingreso de la pareja conyugal. Las variables demográficas registradas se pueden incorporar a ecuaciones estimables, conjuntamente con otras características propias de los individuos que posibilitan capturar la heterogeneidad de los mismos.

También es conocido que los comportamientos de los aglomerados pueden variar según las variables que se estudien. En el caso del ingreso individual de las parejas conyugales existen diferencias apreciables que es interesante analizar.

Sin embargo, los datos microeconómicos poseen ciertas características que complican la modelación econométrica y su estimación. Por ejemplo, para la comparación propuesta aparecen algunos problemas estadísticos que no permiten utilizar directamente los métodos de estimación clásicos. Uno de ellos surge cuando se estiman simultáneamente ambas ecuaciones debido a que los residuos están correlacionados entre sí y a su vez, con algunos de los regresores. El otro se debe a que muchas mujeres no participan en el mercado laboral y en consecuencia un gran porcentaje de la variable respuesta es cero (variable limitada), hecho que se debe tener en cuenta para obtener estimaciones consistentes.



Una solución es realizar la estimación mediante modelos Tobit multivariados por el método de máxima verosimilitud clásico, lo cual es computacionalmente trabajoso, debido a que el cálculo de las probabilidades normales truncadas de dimensión alta es dificultoso. Por tal motivo, se desarrollaron modelos más accesibles para sistemas simultáneos con información limitada y se propusieron distintos estimadores para los coeficientes de las ecuaciones (Amemiya 1974, 1979, Nelson y Olson 1978) .

La comparación de la eficiencia de varios métodos de estimación se ha llevado a cabo en contextos especiales de modelos para variables respuestas limitadas. Amemiya (1978,1979) y Lee (1993), entre otros, mostraron que los estimadores en dos etapas que usan valores predichos (Nelson y Olson,1978) pueden ser ineficientes respecto a la versión del estimador mínimo cuadrático generalizado de Amemiya (AGLS).

En este trabajo se realiza una comparación entre los aglomerados Capital Federal y Rosario sobre la formación de los ingresos individuales de la pareja conyugal. Dicha comparación se efectúa mediante un breve estudio descriptivo y la estimación de un modelo de sistema de ecuaciones. La estimación se realiza por métodos que tienen en cuenta la censura y, por otro lado, por métodos convencionales analizando sólo aquellas parejas en que ambos cónyuges tienen ingreso positivo. Se emplea el método econométrico propuesto por Nelson y Olson para la estimación de sistemas de ecuaciones con variables censuradas y se detallan algunas limitaciones que surgen al utilizar la estimación según Amemiya (1979). Se utiliza la información sumistrada por la EPH para los aglomerados urbanos Rosario y Capital Federal correspondiente a la primera onda de 1998.

En la sección II se presenta una breve reseña sobre propuestas existente en la bibliografía para la especificación y estimación de modelos multiecuacionales con variables dependientes censuradas. En la sección III se definen las variables que se incorporan al modelo para explicar el ingreso de la pareja conyugal. En la sección IV se realiza una descripción resumida de los aglomerados en base a las variables definidas en la sección III. En la sección V se presentan las estimaciones de los modelos para los dos aglomerados y se analizan las posibles causas de fallas que presentan algunos métodos de estimación en el presente estudio. Por último, en la sección VI se discuten los resultados hallados.



II.- MODELOS MULTIVARIADOS CON VARIABLE DEPENDIENTE LIMITADA

Se puede definir un modelo de ecuación única para variable dependiente limitada como

$$Y_t^* = \beta' X_t + u_t \quad (II.1)$$

y

$$Y_t = h(Y_t^*) \quad (II.2)$$

donde, X_t es un vector de variables exógenas; Y_t^* es una variable aleatoria latente (es decir, no directamente observable); Y_t es la variable dependiente observada; u_t es un disturbio no observado que se supone no correlacionado e independiente de X_t , con distribución $N(0, \sigma^2)$, β es un vector de parámetros desconocidos; y h es una función conocida que especifica la estructura de quiebre de la variable dependiente Y_t . Por ejemplo para la estructura Tobit (Tobin, 1958) la forma de $h(\cdot)$ es

$$h(Y_t^*) = \begin{cases} Y_t^* & \text{si } Y_t^* > c_t \\ c_t & \text{si } Y_t^* \leq c_t \end{cases} \quad (II.3)$$

El valor del quiebre, c_t , debe ser una constante observada o una variable exógena.

Para el caso de un sistema de ecuaciones simultáneas de dos variables endógenas con una de ellas censurada, por ejemplo, una variable no restringida y la segunda restringida a ser no negativa, Amemiya (1974) y otros autores posteriores proponen un modelo que se puede escribir como

$$Y_{1t} = \gamma_1 Y_{2t} + \beta_1' X_t + u_{1t} \quad (II.4)$$

$$Y_{2t} = \begin{cases} \gamma_2 Y_{1t} + \beta_2' X_t + u_{2t} & \text{si RHS} > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (II.5)$$

donde, RHS (right-hand side) significa que el segundo miembro de la ecuación es positiva.

Las correspondientes ecuaciones reducidas son

$$\left. \begin{aligned} Y_{1t} &= \pi_1' X_t + v_{1t} \\ Y_{2t} &= \pi_2' X_t + v_{2t} \end{aligned} \right\} \text{ si } \pi_2' X_t + v_{2t} > 0 \quad (II.6)$$



$$\left. \begin{array}{l} Y_{1t} = \pi_1' X_t + v_{1t} \\ Y_{2t} = 0 \end{array} \right\} \text{ si } (1 - \gamma_1 \gamma_2)(\pi_2' X_t + v_{2t}) \leq 0 \quad (II.7)$$

En forma matricial las ecuaciones II.4 y II.5 se pueden expresar como

$$\Gamma Y_t \geq B X_t + u_t; \quad \text{si } Y_t \geq 0 \quad (II.8)$$

y

$$(\Gamma Y_t)_i = (B X_t + u_t)_i, \quad \text{toda vez que } Y_{it} > 0.$$

Amemiya (1979) observa los inconvenientes que presenta el sistema (II.8)

- a.- se deben restringir los coeficientes de las variables endógenas a $(1 - \gamma_1 \gamma_2) > 0$ por problemas de coherencia,
- b.- para que el modelo tenga una única solución cada menor principal de Γ debe ser positivo, por lo tanto la condición suficiente es que $\Gamma + \Gamma'$ sea definida positiva.

Nelson y Olson (1978) introducen una expresión diferente del modelo de ecuaciones simultáneas con variables endógenas latentes. Suponiendo el mismo tipo de variables que en el modelo (II.4) y (II.5) de Amemiya, estos autores proponen la siguiente forma estructural del modelo

$$Y_{1t} = \gamma_1 Y_{2t}^* + \beta_1' X_t + u_{1t} \quad (II.9)$$

$$Y_{2t}^* = \gamma_2 Y_{1t} + \beta_2' X_t + u_{2t} \quad (II.10a)$$

$$Y_{2t} = \begin{cases} Y_{2t}^* & \text{si } Y_{2t}^* > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (II.10b)$$

Suponen que Y_{1t} e Y_{2t} son variables aleatorias observables, Y_{2t}^* es una variable latente no observable, X_t matrices de constantes conocidas y todos los otros términos son los que se definen generalmente para modelos de ecuaciones simultáneas. La correspondiente forma reducida del modelo es

$$Y_{1t} = \pi_1' X_t + v_{1t} \quad (II.11)$$

$$Y_{2t}^* = \pi_2' X_t + v_{2t} \quad (II.12a)$$

$$Y_{2t} = \begin{cases} Y_{2t}^* & \text{si } Y_{2t}^* > 0 \\ 0 & \text{en otro caso} \end{cases} \quad (II.12b)$$



La diferencia de este último modelo con el de Amemiya, es que en las ecuaciones estructurales, la variable endógena del segundo miembro aparece en forma latente no restringida. Este cambio produce modificaciones importantes sobre las implicaciones del modelo, entre las que se pueden destacar: i) la forma reducida del modelo no requiere ninguna restricción sobre los coeficientes de las variables endógenas, ii) la ecuación de la forma reducida para Y_{1t} (no restringida) no depende de si la variable Y_{2t} (restringida) toma el valor o está por encima del punto de quiebre y iii) esta especificación permite que la magnitud de la variable latente Y_{2t}^* afecte directamente el valor de Y_{1t} .

Las modificaciones planteadas en los puntos ii) y iii) pueden producir interpretaciones distintas según el modelo que se emplee en el caso de los ingresos de la pareja conyugal. Si se considera que Y_{1t} es el logaritmo del ingreso de la fuente laboral del varón e Y_{2t} el logaritmo del ingreso de la fuente laboral de la mujer, existe un porcentaje muy grande de mujeres con punto de quiebre para la censura $c_t=0$, por lo tanto preferir Y_{2t} como variable explicativa, puede afectar el desempeño de Y_{1t} . Se puede esperar que el desempeño de Y_{1t} sea diferente cuando se usa Y_{2t}^* .

II.1 Estimación de los modelos

Existen dificultades bien conocidas, tanto en la literatura estadística como econométrica, de la aplicación usual de máxima verosimilitud con información limitada (MVIL) a la función normal bivariada truncada.

La estimación del modelo de ecuaciones (II.9)-(II.10) no es sencilla ni viable por el procedimiento de MVIL. Para solucionar este problema varios autores plantean distintos métodos de estimación.

II.1.1 Estimador de Nelson y Olson

Nelson y Olson (1978) proponen un procedimiento análogo a mínimos cuadrados en dos etapas. El mismo se puede expresar brevemente de la siguiente manera: en el primer paso se estiman los parámetros de la forma reducida para cada ecuación por separado, la primera por mínimos cuadrados, y la segunda por máxima verosimilitud Tobit. Las estimaciones se usan para crear las variables instrumentales $\hat{Y}_{1t} = \hat{\pi}'_1 X_t$ e $\hat{Y}_{2t}^* = \hat{\pi}'_2 X_t$, las cuales son al menos asintóticamente no correlacionadas con los términos de los disturbios.



Estos instrumentos se sustituyen por sus contrapartes en el segundo miembro de las ecuaciones (II.9) y (II.10a). Las ecuaciones resultantes son tratadas como una ecuación lineal simple y un modelo Tobit, respectivamente. Luego, MCO y el análisis Tobit proveen los elementos para estimar, en la segunda etapa, los coeficientes estructurales, aplicando máxima verosimilitud a cada ecuación.

En la primera etapa de la estimación se considera que las covariancias entre los disturbios son cero. Se pueden obtener estimadores más eficientes si se supone que dicha covariancia no es cero, hecho que tienen en cuenta los siguientes métodos de estimación.

II. 1. 2 Estimador de Amemiya

Amemiya (1979) introdujo un método general para la estimación de los parámetros estructurales de sistemas de ecuaciones simultáneas, a partir de los parámetros de la forma reducida de las mismas. Es un método a dos etapas, que puede ser considerado como una generalización del estimador mínimo cuadrados en dos etapas (MC2E). Se estiman los parámetros del modelo de la forma reducida por algún método consistente. Luego los parámetros estructurales se estiman por mínimos cuadrados o mínimos cuadrados generalizados (AGLS), que tiene en cuenta la covariancia entre los errores.

Este estimador es asintóticamente más eficiente que el de Nelson y Olson (1978), aunque computacionalmente sea más difícil.

II. 1. 3 Estimación de Huang, Sloan y Adamache

Huang, Sloan y Adamache (1987) proponen aplicar el algoritmo de maximización de la esperanza (EM) (Dempster, Laird, y Rubin, 1977) para calcular la función de verosimilitud en el modelo reducido resultante de suponer que las variables endógenas se distribuyen como una normal bivariada truncada. Consideran la estimación de dos regresiones Tobit aparentemente no relacionadas.

La maximización de la función de verosimilitud se puede realizar por el método iterativo EM, el cual consiste en dos etapas: 1º) encontrar la esperanza condicional de las estadísticas suficientes y 2º) maximizarla por un método iterativo.



II. 1. 4 Estimación por mínimos cuadrados en tres etapas

En caso de considerar sólo las parejas en donde ambos miembros tienen ingreso positivo, la estimación simultánea de las ecuaciones se realiza por el convencional método de mínimos cuadrados en tres etapas (MC3E).

III.- Determinación y descripción de las variables que intervienen en el modelo

En este estudio se desea modelar el logaritmo del ingreso de la fuente laboral de cada uno de los cónyuges, en función de ciertas variables de interés. A continuación se presenta una descripción de las mismas.

Variables endógenas

Se decide trabajar con el ingreso laboral neto de cada uno de los integrantes de la pareja conyugal, debido a que es la variable registrada en la EPH. No se trabaja con la variable ingreso por hora, como suele ser costumbre en trabajos sobre ingreso, por dos motivos: i) se considera que dividir por el número de horas trabajadas, cuando el período de referencia de esta variable no es el mismo que para la variable ingreso puede incluir una nueva fuente de error en la variable en estudio, la cual ya tiene problemas de registro; ii) el salario neto puede estar influenciado en algunos casos por el subsidio conocido como salario familiar.

Se consideran en forma conjunta los ingresos de los dos integrantes de la pareja conyugal.

Se trabaja con estas variables transformadas, que se denominan:

- **Y₁: liflv** : logaritmo del ingreso de fuente laboral del cónyuge (varón).
- **Y₂: liflm**: logaritmo del ingreso de fuente laboral de la cónyuge (mujer).

En los casos en que los ingresos presentan valores iguales a cero, no se aplica la transformación y permanecen como tales. La variable Y₂ presenta una distribución asimétrica con datos censurados a la izquierda, correspondientes a ingresos cero.

La consideración conjunta de los ingresos de los dos integrantes de la pareja conyugal, mediante la estimación de un sistema de ecuaciones simultáneas, puede ayudar a atenuar el problema de endogeneidad, al tiempo que permite tener en cuenta las relaciones entre los ingresos de ambos.

Existen estudios econométricos que consideran algunas de las características del cónyuge en una ecuación de ingreso individual e introducen el ingreso como variable exógena, al igual que la educación o alguna otra característica. Este criterio fue objetado tempranamente



por Griliches(1977), en relación con la educación y por Boulier y Rosenzweig (1984) en lo que se refiere a las características llevadas por la persona al mercado matrimonial y tratadas como si fueran exógenas. Lam (1988) resaltó que una implicación empírica de cualquier modelo de ordenamiento marital es que las características de los esposos son endógenas al mercado matrimonial. Sin embargo, Jacobsen y Rayack(1996) no encontraron diferencias sustanciales entre los ingresos de los maridos independientemente del hecho que sus esposas trabajen o no.

Variables exógenas

Las variables registradas en la EPH y que son utilizadas en la literatura económica para explicar el ingreso de fuente laboral (o su transformación logarítmica) denominadas variables exógenas (X), se definen para cada uno de los cónyuges. Para el caso de los varones se agrega a cada variable el sufijo v y para las mujeres se agrega el sufijo m.

Las principales dificultades del contraste empírico de la tradicional teoría del capital humano (Mincer, 1970), se encuentran en identificar las habilidades de las personas, entendidas como características que pueden ser consideradas dotaciones, y distinguirlas de las de inversión. Se definen las variables que pueden ser consideradas de inversión.

- **edad:** variable continua calculada a partir de la fecha de nacimiento declarada por el encuestado. Esta variable se toma como *proxy* de la experiencia de la persona.
- **edad2:** variable edad elevada al cuadrado. Se incluye para captar el comportamiento no lineal del retorno por año de actividad, que se supone creciente hasta una cierta edad y luego puede comenzar a decrecer lentamente.
- **medesco:** años de escolaridad declarados por el encuestado. Típica variable del capital humano que mide el retorno por cada año adicional de educación del individuo.

Otro aspecto a tener en cuenta es la relación entre los logros educativos de uno de los miembros de la pareja conyugal y los ingresos del otro miembro. Becker (1973) propuso una aplicación fundada en la teoría acerca de la formación de parejas basadas en la asociación de características. Sugirió que la educación de la esposa es una variable *proxy* de características no observadas que afectan la productividad del cónyuge en las actividades fuera del mercado.

En Estados Unidos se probó la existencia de una relación positiva entre ambas variables (Becker, 1983) y se interpreta como el resultado de la influencia de la educación en la calidad de flujos de servicios de apoyo que las esposas proporcionan a los maridos, y que incrementa la productividad de las actividades de mercado de estos, en el marco de una fuerte división sexual de trabajo. Akerlof (1997) señaló, por otra parte, que la educación es



una variable que no sólo tiene importancia económica intrínseca, sino también significado social pues revela la posición social relativa de las personas, pudiendo ser éste un factor atractivo por sí mismo.

- **lifnl**: logaritmo del ingreso de fuente no laboral. Variable continua que es el logaritmo de la suma de todos los ingresos no laborales declarados por el encuestado. Si es cero, permanece como tal.

Se cuenta con cada una de las siguientes variables exógenas relativas al hogar, que pueden influenciar el ingreso de ambos cónyuges.

- **cont6**: cantidad de hijos menores de 6 años.
- **cont18**: cantidad de hijos entre 6 y 18 años.

Esta categorización de la variable número de hijos, se debe a que ella puede estar relacionada con el ingreso por distintas razones: para el hombre por la percepción de un subsidio por los hijos (salario familiar) y para la mujer por la complementación entre el tiempo dedicado al cuidado de los hijos y el tiempo dedicado a trabajar fuera del hogar.

Con el objetivo de contar con una base de datos que contenga la información correspondiente a las variables de interés para la muestra de hogares, se programó una macro en SAS. La misma permite, de una manera automática, seleccionar exclusivamente los hogares de interés y generar un registro para cada pareja conyugal. Una vez seleccionados los hogares, se estudian posibles inconsistencias en algunas variables tales como edad, sexo y escolaridad, revisando la información de otras ondas en los casos que fue posible.

IV.- CARACTERÍSTICAS DE LA MUESTRA EN ESTUDIO

Para realizar el estudio empírico se seleccionan los aglomerados urbanos: Rosario y Capital Federal, debido a que las características de los mismos son muy disímiles (Tabla IV.1).

La muestra en estudio para cada aglomerado está compuesta por las parejas conyugales (sin importar la legalidad de la unión) que viven en un mismo hogar, entrevistadas en la EPH en la primera onda de 1998. Estas muestras no incluyen aquellas parejas donde el varón tiene ingreso cero o es inactivo o tiene 65 o más años de edad¹.

¹ Se eliminan los varones con ingreso cero porque su inclusión apareja problemas en la estimación. Además se presentan en un número pequeño de casos y no producen cambios importantes en los resultados.



Tabla IV.1: Porcentaje de mujeres con participación laboral, ocupadas y con ingreso cero discriminadas por aglomerado

Aglomerado	Número parejas conyugales	Participación laboral	Ocupadas (en relación con participación)	Ingreso cero (*)
Cap. Federal	331	61.6	90.0	47.4
Rosario	444	41.7	85.4	64.0

(*)El porcentaje de mujeres sin ingreso no es el complemento del porcentaje de participación debido a la existencia de un número pequeño de mujeres que declaran estar ocupadas pero sus ingresos son cero.

En la tabla se observa que los ingresos ceros se deben principalmente a que las mujeres son inactivas y no desocupadas.

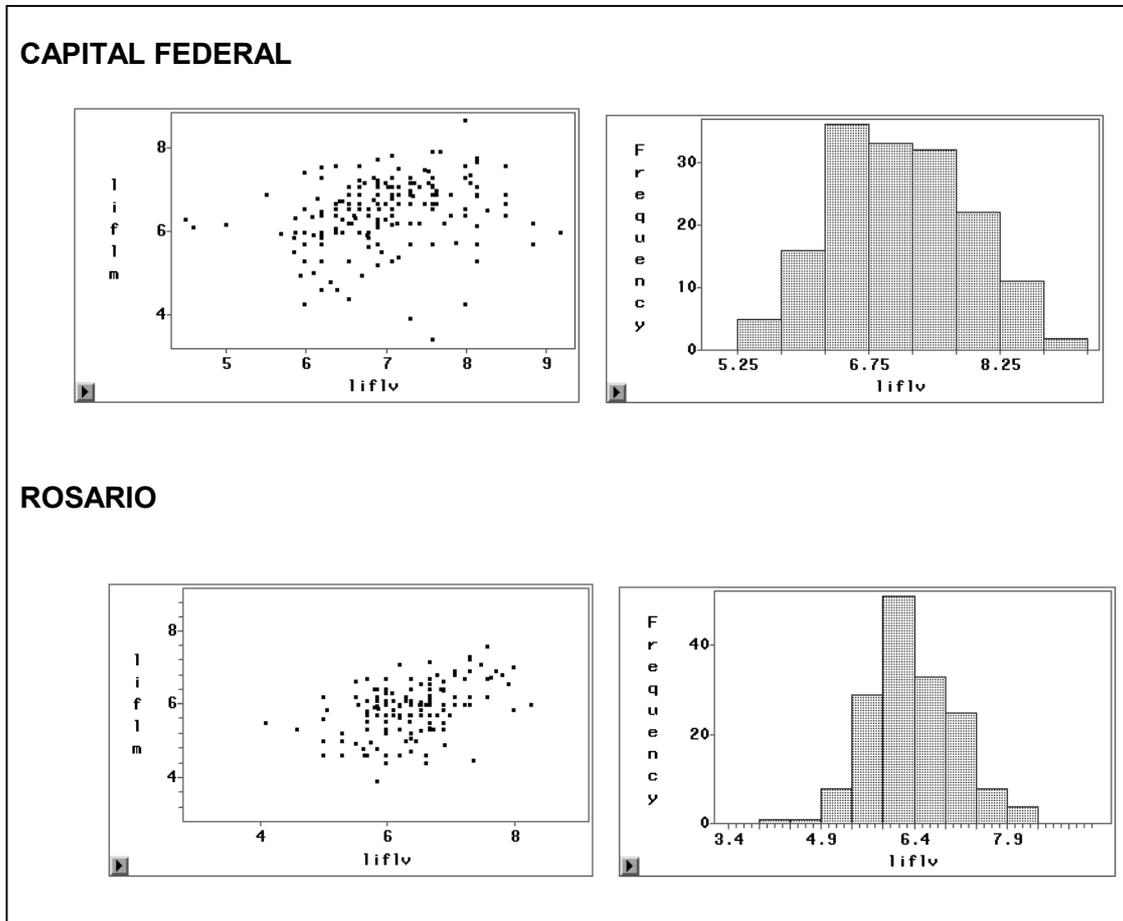
La mediana y el rango de ingresos de ambos cónyuges, excluidos los ingresos ceros se presentan en la Tabla IV.2. Se puede destacar que Capital Federal posee un ingreso medio para el varón que duplica el correspondiente a Rosario, mientras que para la mujer es un 75% más alto. Por otro lado, el rango de los ingresos del varón es el 70% mayor en Capital Federal y además, es el triple en el caso de la mujer. Esto se podría asociar con la participación femenina en el mercado de trabajo, ya que Capital Federal posee una participación laboral femenina aproximadamente 50% mayor.

Tabla IV.2: Mediana y rango del ingreso laboral de ambos cónyuges por aglomerado

Aglomerado	liflv		liflm(sin ceros)	
	Mediana	Rango	Mediana	Rango
Capital Federal	1050	9910	700	5970
Rosario	500	4960	400	1985

La asociación entre los ingresos cuando ambos cónyuges perciben remuneración es diferente en ambos aglomerados, en Capital Federal no parece existir asociación entre los ingresos de ambos cónyuges, en cambio en Rosario se insinúa una asociación positiva. La distribución de los ingresos de los varones cuyas mujeres no tienen ingresos es bastante distinta en ambos aglomerados, en el caso de Capital Federal la distribución es más uniforme presentándose similar cantidad de varones en varios intervalos de ingresos, mientras que Rosario presenta un pico pronunciado para los varones de ingreso medio (Gráfico IV.1).

Gráfico IV.1: Diagramas de dispersión: liflv vs.liflm (para ambos positivos) e histogramas de liflv dado liflm=0 para los dos aglomerados.



Como una primera aproximación se realiza una prueba de independencia (χ^2) para medir si existe asociación entre el ingreso laboral de la mujer respecto de cada una de las siguientes variables: escolaridad, edad, número de hijos menores de 6 años y número de hijos entre 6 y 18 años. Se categorizaron las variables de este análisis de la siguiente manera: ingreso laboral en positivo y cero, escolaridad en sin escolaridad, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, mayor que secundaria y edad, en los intervalos: 10-24, 25-34, 35-44, 45-54, >55.

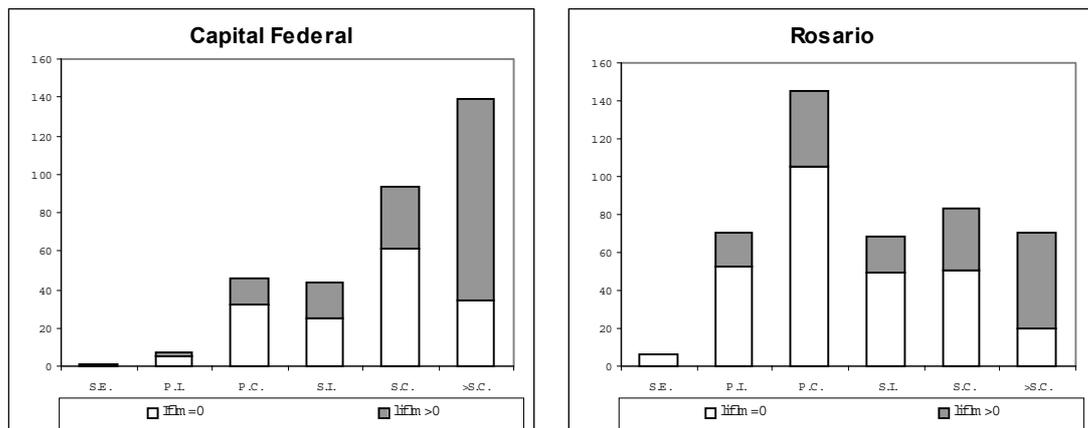
Los resultados se presentan en la Tabla IV.3. Se puede destacar que en los dos aglomerados el comportamiento de la variable escolaridad difiere según la mujer tenga ingreso cero o no, en cambio para la variable edad existe asociación en Capital Federal pero no en Rosario. Por otro lado, no existe diferencia para el número de hijos menores de seis años en ninguno de los dos aglomerados, pero sí existe asociación para los hijos entre 6 y 18 años en Rosario.

Tabla IV. 3: Asociación entre mujeres según ingreso cero o no y variables exógenas

Aglomerado	medescom	Edad	cont6	cont18
Capital Federal	Si	Si	No	No
Rosario	Si	No	No	Si

La escolaridad de la mujer también es muy distinta en ambos aglomerados, en Capital Federal predominan las mujeres con educación mayor que secundaria (42%), mientras que en Rosario el porcentaje más alto se presenta para aquéllas con primaria completa (33%). Esto también puede ser un motivo de la menor participación laboral de las mujeres en Rosario. En cuanto a las mujeres que no participan en el mercado laboral, en los dos aglomerados la mayoría pertenece a los niveles más bajo de educación (Gráfico IV.2).

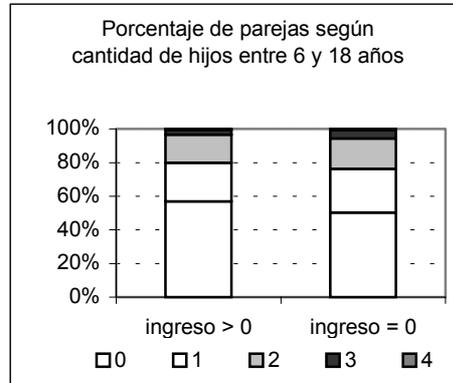
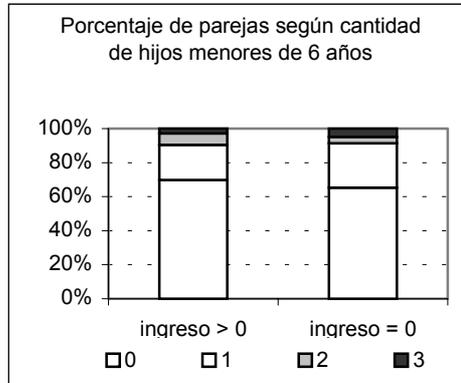
Gráfico IV.2: Escolaridad de la mujer según si percibe ingreso o no por aglomerado



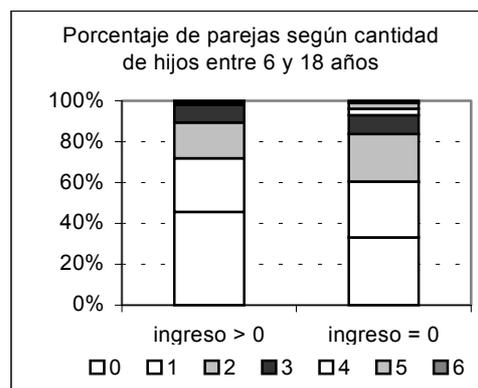
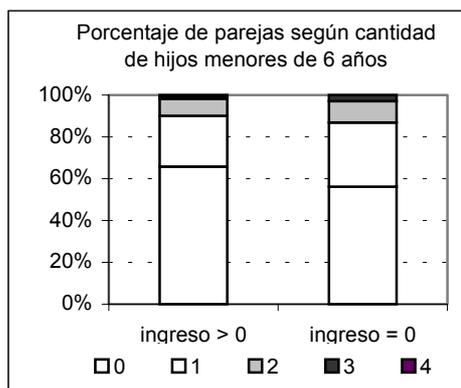
En cuanto a la cantidad de hijos (menores de 6 años y entre 6 y 18 años), si se realiza una comparación dentro de cada aglomerado, existe en general una proporción más grande de parejas con un número mayor de hijos cuando la mujer no participa del mercado laboral, siendo este hecho más pronunciado en Rosario. Además la proporción de parejas que tienen varios hijos es superior en Rosario que en Capital. Este puede ser un motivo adicional para explicar la mayor tasa de participación de Capital Federal (Gráfico IV.3).

Gráfico IV.3:

Capital Federal



Rosario



El breve análisis descriptivo realizado sugiere comportamientos bastante disímiles entre los aglomerados analizados, a esto se podría agregar que pueden existir pautas sociales y culturales distintas entre los mismos. Por estos motivos, es de esperar que la estimación de los modelos pueda diferir entre ellos.

V. Modelos empíricos

Para modelar el ingreso laboral de la pareja conyugal se considera el modelo (I.9)-(I.10a)-(I.10b). Se realizan las estimaciones por los tres métodos enunciados en la sección II, utilizando el módulo IML de SAS. La nomenclatura presentada por el modelo nombrado, para esta aplicación, es la siguiente :

Y_1 : vector de la variable endógena para el varón (liflv),

Y_2 : vector de la variable endógena para la mujer (liflm),

X_1 : matriz de características medibles (variables exógenas) que afectan a Y_1 , (edadv, edad2v, mescov, lifnlv, mescom, lifnlm, cont6, cont18),



X_2 : matriz de características medibles (variables exógenas) que afectan a Y_2 , (edad_m, edad_{2m}, mescom, lifnlm, mescov, lifnlv, cont6, cont18).

Tabla V.1: Estimación de los coeficientes para ambos cónyuges por aglomerado según distintos métodos.

Variable	Rosario		Capital Federal		Rosario		Capital Federal	
	MC3E	Nel-Ols	MC3E	Nel-Ols	MC3E	Nel-Ols	MC3E	Nel-Ols
	VARON				MUJER			
Intercepto	1.9736*** (0.4383)	4.0954*** (0.6075)	4.6507*** (0.7752)	4.3294*** (0.0685)	1.4548 (0.8026)	38.593 (20.4515)	3.8215*** (1.0137)	141.896* (81.940)
Liflcony ⁽¹⁾	0.4383*** (0.0625)	0.0259 (0.0271)	0.0138 (0.0644)	-0.0087 (0.0440)	0.5943*** (0.0862)	-14.13*** (5.2839)	0.0211 (0.0916)	-34.867* (18.553)
Edad	0.03244 (0.0280)	0.0548** (0.0266)	0.0001 (0.0351)	0.0401 (0.0326)	0.0346 (0.0369)	0.967*** (0.3375)	0.0096 (0.0197)	3.0061* (1.6677)
Edad2	-0.0003 (0.0003)	-0.0005 (0.0003)	0.0002 (0.0004)	-0.0003 (0.0004)	-0.0004 (0.0004)	-0.0267 (0.4859)	0.0372 (0.0550)	-2.631* (1.5836)
Medescov	0.0633*** (0.0147)	0.0587*** (0.0100)	0.0888*** (0.0154)	0.0907*** (0.0132)	-0.0169 (0.0183)	1.599*** (0.4731)	0.0607 (0.0450)	1.764** (0.7152)
Lifnlv	-0.0290 (0.0563)	-0.0177 (0.0448)	-0.0860* (0.0487)	-0.0805* (0.0428)	0.0298 (0.0665)	-0.018*** (0.0054)	-0.0007 (0.0005)	-0.0136** (0.0046)
Medescom	0.0197 (0.0141)	0.0291 (0.0150)	0.0460*** (0.0174)	0.0251 (0.0325)	0.0189 (0.0169)	0.946*** (0.2294)	0.0983*** (0.0196)	1.433*** (0.4329)
Lifnlm	0.0251 (0.0674)	0.0128 (0.0372)	-0.0690 (0.0500)	-0.0074 (0.0397)	0.0089 (0.0789)	-0.6072 (0.4416)	-0.1019 (0.0627)	-0.846** (0.3304)
Cont6	0.0762 (0.0673)	0.0325 (0.0595)	0.0603 (0.0869)	0.0680 (0.1078)	0.0022 (0.0789)	-1.639*** (0.5548)	-0.290*** (0.1005)	2.348 (2.3428)
Cont18	0.0671 (0.0406)	0.0493 (0.0324)	0.0740 (0.0608)	0.0235 (0.0456)	-0.1230** (0.0479)	-0.6435 (0.3295)	-0.1665** (0.0694)	0.510 (0.2976)

* significativo al 10%, ** significativo al 5%, *** significativo al 1%. Los valores entre paréntesis corresponden a los desvíos estándares de los estimadores. ⁽¹⁾liflm para la ecuación del varón y liflv para la de la mujer.

La Tabla V.1 presenta las estimaciones realizadas por el método propuesto por Nelson y Olson (1978), también para realizar comparaciones se presentan los resultados de aplicar mínimos cuadrados tres etapas (MC3E) al mismo modelo, considerando sólo las parejas en que ambos cónyuges tienen ingreso positivo. Este último grupo es el que generalmente se analiza en los estudios sobre el tema en nuestro país.

Los coeficientes estimados muestran que la censura afecta especialmente los resultados que se encuentran para las mujeres, y a su vez que el comportamiento es distinto para cada aglomerado.



Para los varones sólo son significativas en todos los casos, el intercepto y su escolaridad. Además en Rosario, si se consideran sólo las parejas donde ambos cónyuges tienen ingreso positivo, resulta significativo con signo positivo el ingreso del otro cónyuge, en cambio para Capital Federal ese mismo grupo presenta coeficiente significativo positivo para la escolaridad de la mujer y negativo para el ingreso de fuente no laboral propio.

Por otro lado, en Rosario cuando se consideran las mujeres con ingreso positivo, sólo el ingreso del cónyuge y la cantidad de hijos entre 6 y 18 años son significativos, mientras que en Capital Federal lo son el intercepto, su propia escolaridad, la cantidad de hijos menores de 6 años y la cantidad de hijos entre 6 y 18 años. Tanto en Capital Federal como en Rosario, al considerar la muestra completa aumentan las variables que explican la respuesta. Para el primer caso con signo positivo el intercepto, edad, escolaridad propia y del cónyuge y con signo negativo el ingreso del cónyuge, la edad al cuadrado, el ingreso de fuente no laboral propio y del cónyuge. Para el segundo, con signo positivo la edad, su escolaridad y la del cónyuge, y con signo negativo el ingreso del cónyuge, el ingreso de fuente no laboral del esposo y la cantidad de hijos menores de seis años.

En las estimaciones realizadas con los otros métodos que tienen en cuenta en la segunda etapa de la estimación la covariancia entre los errores (AGLS (Amemiya, 1979) y E-M), se presenta el problema que la estimación de dicha covariancia resulta ser un valor extremadamente grande, lo que invalida todas las estimaciones posteriores. Entre las posibles causas de dichos valores estaría el alto porcentaje de censura (47 y 64 % de ingresos de la mujer es igual a cero).

Si se analiza el estimador consistente de la covariancia de los errores propuesto por Amemiya,

$$\hat{\sigma}_{12} = T^{-1} \sum (Y_{2t} \hat{v}_{1t} \hat{F}_t^{-1}),$$

donde \hat{v}_{1t} es el residuo de aplicar MCO a la primera ecuación en la primera etapa y

$$\hat{F}_t = \int_{-\infty}^{\hat{\pi}_2' x_t} (2\pi\sigma_2^2)^{1/2} \exp\left[-\frac{(\tau)^2}{2\sigma_2^2}\right] d\tau.$$

se puede ver que en la suma intervienen los ingresos de la mujer distintos de cero y en forma inversa la distribución acumulada de los valores estimados de dichos ingresos. Este último término es el que afecta el crecimiento desmesurado de la variancia para los casos en que el valor estimado es un valor del extremo izquierdo de la distribución, lo que repercute en un valor prácticamente cero de probabilidad.



Se realizan dos tipos de análisis exploratorios para investigar el problema planteado:

i) Se re muestrea la población en estudio con distintos porcentajes de censura. Para ambos aglomerados, el valor de la covariancia se hace extremadamente grande cuando la censura es del 35% aproximadamente, el valor de la constante se hace negativa con valores pronunciadamente decrecientes y el peso relativo del coeficiente de su escolaridad, en el caso de la mujer, va aumentando considerablemente a medida que se aumenta el porcentaje de censura, como así también los coeficientes de los hijos con signo negativo (Tabla V.2). El valor de la constante está directamente relacionado con el porcentaje de censura, a medida que ésta aumenta desplaza la distribución hacia la izquierda. Por otro lado, el aumento de censura prioriza la influencia relativa de la educación y la cantidad de hijos.

Tabla V.2. Estimación de los coeficientes de la constante y escolaridad para la mujer para distintos porcentajes de censura

Porcentaje de censura		0	10	20	30	40	50*	64
Capital Federal	Intercepto	3.1228	-0.3978	-4.772	-7.847	-6.201	-9.479	
	Medescom	0.1014	0.2793	0.3632	0.5128	0.5428	0.6517	
	Cont6	-0.290	-0.913	-0.890	-1.007	-1.418	-1.823	
	Cont18	-0.166	-0.334	0.086	-0.591	-0.414	-0.620	
Rosario	Intercepto	4.391	-2.5046	4.3519	-0.6489	-8.4252	-11.3785	-15.152
	Medescom	0.042	0.0579	0.1751	0.2421	0.2837	0.3733	0.4399
	Cont6	0.054	-0.456	0.106	0.750	-1.010	-1.279	-1.497
	Cont18	-0.099	-0.254	-0.516	-0.698	-0.772	-0.774	-0.877

* El porcentaje de censura máximo en Capital Federal es de 47% y es el que se utilizó para el cálculo.

ii) Se analizan las características de las mujeres que presentan valores extremos, especialmente su escolaridad que es la variable que siempre resulta significativa cualquiera sea el porcentaje de censura, tanto en la primera etapa de estimación Tobit como en la segunda etapa del método de Nelson y Olson (1978). Además la magnitud del coeficiente es mayor cuando mayor es la censura.

El nivel de escolaridad de las mujeres que trabajan y generan valores extremos es similar al de las mujeres que no trabajan (Gráfico V.1).

Si se consideran además las variables cantidad de hijos menores de 6 años y cantidad de hijos entre 6 y 18 años, se puede observar que las mujeres empleadas que generan valores extremos poseen en general mayor cantidad de hijos en ambos grupos (Gráfico IV.2 y Gráfico IV.3).

VI.- Consideraciones Finales

Los resultados empíricos encontrados en este estudio ayudan a comprender principalmente tres aspectos a tener en cuenta al realizar trabajos que traten de explicar la conformación del ingreso individual de los miembros de la pareja conyugal en la Argentina.

Primero, a través del breve análisis descriptivo realizado en la sección III se puede apreciar el comportamiento disímil entre aglomerados, tanto por el porcentaje de mujeres que no participan en el mercado laboral, como por las características diferentes de estas mujeres según sea el aglomerado. Por supuesto que este hecho, luego se verá reflejado en la estimación del modelo que intenta explicar la formación del ingreso.

Gráfico V.1. Distribución de las mujeres según ingreso, escolaridad y problemas de estimación

Capital Federal



Rosario





Segundo, si se trabaja sólo con las parejas en que ambos tienen ingreso positivo (como es habitual en nuestro país), los resultados y por ende las interpretaciones varían fundamentalmente, especialmente en el caso de las mujeres.

Tercero, el alto porcentaje de censura en los datos (47% y 64%), hace que se tengan dificultades en la estimación de los parámetros aún cuando se utilicen métodos recomendados tanto en la literatura estadística como econométrica. En este trabajo se presentan los estimadores de Nelson y Olson (1978) porque de los métodos probados es el único que permite obtener una estimación a pesar del alto porcentaje de censura, especialmente en cuanto a la variancia de los estimadores, no obstante se debería seguir investigando sobre la real eficiencia de los mismos. Los otros métodos utilizados son aquellos que tienen en cuenta la covariancia entre los errores de las dos ecuaciones, uno es el AGLS de Amemiya (1979) y otro el propuesto por Huang et al.(1987) que maximiza la función de verosimilitud de la normal bivariada truncada por medio del algoritmo E-M. En el primer caso la estimación de la covariancia entre los errores de las dos ecuaciones en la segunda etapa se hace extremadamente grande, lo que anula los cálculos de los pasos posteriores, y en el segundo, los valores de la variancia de los estimadores se vuelven demasiado grandes generando estimaciones carentes de significado.

Del breve análisis exploratorio sobre las causas de la falta de eficiencia de los dos últimos métodos, es interesante notar que al considerar el nivel de la escolaridad de las mujeres que trabajan y generan valores extremos, se puede apreciar que es similar al de las mujeres que no trabajan (Gráfico V.1). Este resultado parece ser lógico ya que el gran porcentaje de censura afecta los coeficientes del modelo Tobit haciendo que las características de las mujeres que no trabajan influyan de manera tal que el ingreso estimado sea menor si un individuo presenta características similares a las del grupo que no participa del mercado laboral. Por lo tanto, aquellas mujeres que trabajan pero que tienen características similares a las que no trabajan, generan valores ubicados a la izquierda de la distribución del ingreso, produciendo valores muy pequeños en la función de distribución. De acuerdo a la fórmula sugerida por el método de Amemiya, estos valores participan en el cálculo de la covariancia de los errores y su inclusión genera un aumento desmedido de la estimación que invalida el cálculo de la segunda etapa.

El mismo fenómeno considerado desde otro punto de vista, muestra que los valores observados en escolaridad de la mujer (variable con coeficiente significativo y positivo en la primera etapa Tobit) para aquellas que trabajan y generan valores extremos, corresponden a los valores más bajos de la distribución correspondiente (Gráfico V.1). Si se consideran



además, las variables cantidad de hijos menores de 6 años y cantidad de hijos entre 6 y 18 años, cuyos coeficientes son también significativos pero negativos, se puede observar que las mujeres empleadas que generan valores extremos poseen en general mayor cantidad de hijos en ambos grupos. Como se indicó anteriormente, a medida que aumenta la censura disminuye marcadamente el valor del intercepto, el cual es negativo, mientras que el coeficiente correspondiente a escolaridad de la mujer aumenta y es positivo. De manera similar, pero en sentido contrario, los coeficientes correspondientes a las dos variables referidas a la cantidad de hijos son negativos y disminuyen a mayor censura. Sin embargo, en términos relativos, todos los coeficientes son pequeños en comparación al valor del intercepto. Por lo tanto, para compensar esta diferencia es necesario contar con valores altos en educación de la mujer y valores bajos en la cantidad de hijos para obtener un ingreso estimado que no se ubique a la izquierda de la distribución. En consecuencia, al observar las características de las mujeres cuyos valores hacen crecer extremadamente la covariancia, estos presentan en su mayoría los valores más bajos en cuanto a escolaridad y los más altos en cuanto a cantidad de hijos, que además, como se indicó previamente, son las características de la mayoría de las mujeres que no trabajan y que son más de la mitad de las mujeres en Rosario y casi la mitad en Capital Federal.

Por todo lo expuesto, se considera que puede ser de interés seguir explorando sobre la falta de adecuación de algunos métodos de estimación a casos como el que nos ocupa.

VII. REFERENCIAS BIBLIOGRÁFICAS

Akerlof, G (1997). "Social distance and social decisions" *Econometrica*, 65, 1005-1027.

Amemiya, T. (1974). "Multivariate regression and simultaneous equation models with dependent variables are truncated normal". *Econometrica*, 42, 999-1012.

---, T. (1978). "The estimation of a simultaneous-equation generalized probit model". *Econometrica*, 46, 1193-1205.

---, T. (1979). "The estimation of a simultaneous-equation Tobit model" *International Economic Review*, 20, nº1, 169-181.

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

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



- Blaconá, M.T., Garcia, M.del C., Borgognone, M.G., Bussi, J. y Pellegrini, J.L. (2001). "Consideraciones metodológicas sobre la estimación econométrica de las ecuaciones de ingresos de los integrantes de la pareja conyugal". Sextas jornadas de investigaciones en la Facultad de Ciencias Económicas y Estadística. ISSN:1666-3543- CD-ROM pág. 152-165.
- Boulier, Bryan L. and Rosenzweig, Mark (1984). "Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior", *Journal of Political Economy*, Vol.92m nº 4.
- Dempster, A., Laird, N. y Rubin, D. (1977). "Maximun Likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society, B*, 39, 1-22.
- Griliches, Z., (1977). "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, 45(1), pp 1-22.
- Heckman, J. (1974). "Shadow price, market wages and labor supply". *Econometrica*, 42, 679-694.
- Huang, C., Sloan, F. y Adamache, K. (1987). "Estimation of seemingly unrelated Tobit regressions via EM algorithm". *Journal of Business & Economic Statistics*, 5, nº 3, 425-430.
- Jacobsen, J. y Rayack, W. (1996). "Do Men Whose Wives Work Really Earn Less?", *American Economic Review*, Vol 86(2), May, pp 269-273.
- Lam, D. (1988). "Marriage Markets and Assortative Mating with Household Public Goods", *Theoretical Results and Empirical Implications*, *The Journal of Human Resources*, XXIII. 4, pp. 462-487.
- Lee, L. F. (1993). "Multivariate Tobit models in econometrics". Estimation of limited dependent variable models under rational expectations. Maddala, S. Rao, C. R. y Vinod, H. Elsevier Science Publishers.
- Mincer, J. (1970). "The distribution of labor income: a survey; With special reference to human capital approach", *Journal of Economic Literature*, Vol. VIII, Nº1.
- Nelson, F. y Olson, L. (1978). "Specification and estimation of a simultaneous-equation model with limited dependent variables". *International Economic Review*, 19,nº3, 695-709.
- Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, pp. 24-36.