



ARTÍCULOS

## **Comparación de Ponderaciones en Regresiones Probit Simultáneas en un Modelo para la Estimación de la Participación Laboral**

Verónica Herrero y Mónica Bocco

Revista de Economía y Estadística, Cuarta Época, Vol. 45, No. 2 (2007), pp. 95-124.

<http://revistas.unc.edu.ar/index.php/REyE/article/view/3841>



La Revista de Economía y Estadística, se edita desde el año 1939. Es una publicación semestral del Instituto de Economía y Finanzas (IEF), Facultad de Ciencias Económicas, Universidad Nacional de Córdoba, Av. Valparaíso s/n, Ciudad Universitaria. X5000HRV, Córdoba, Argentina.

Teléfono: 00 - 54 - 351 - 4437300 interno 253.

Contacto: [rev\\_eco\\_estad@eco.unc.edu.ar](mailto:rev_eco_estad@eco.unc.edu.ar)

Dirección web <http://revistas.unc.edu.ar/index.php/REyE/index>

### Cómo citar este documento:

Herrero, V. y Bocco M. (2007). Comparación de Ponderaciones en Regresiones Probit Simultáneas en un Modelo para la Estimación de la Participación Laboral. *Revista de Economía y Estadística*, Cuarta Época, Vol. 45, No. 2 (2007), pp. 95-124.

Disponible en: <http://revistas.unc.edu.ar/index.php/REyE/article/view/3841>

El Portal de Revistas de la Universidad Nacional de Córdoba es un espacio destinado a la difusión de las investigaciones realizadas por los miembros de la Universidad y a los contenidos académicos y culturales desarrollados en las revistas electrónicas de la Universidad Nacional de Córdoba. Considerando que la Ciencia es un recurso público, es que la Universidad ofrece a toda la comunidad, el acceso libre de su producción científica, académica y cultural.

<http://revistas.unc.edu.ar/index.php/index>



REVISTAS  
de la Universidad  
Nacional de Córdoba



Universidad  
Nacional  
de Córdoba



FCE  
Facultad de Ciencias  
Económicas



1613 - 2013  
400  
AÑOS



Revista de Economía y Estadística - Vol. XLV - N° 2 - (2007) - pp. 95-124  
Instituto de Economía y Finanzas - Facultad de Ciencias Económicas  
Universidad Nacional de Córdoba - Argentina

## **Comparación de Ponderaciones en Regresiones Probit Simultáneas en un Modelo para la Estimación de la Participación Laboral**

**VERONICA HERRERO**

*Instituto de Economía y Finanzas - FCE  
Universidad Nacional de Córdoba  
vherrero@eco.unc.edu.ar*

**MONICA BOCCO**

*Departamento de Matemática - FCA  
Universidad Nacional de Córdoba  
mbocco@agro.uncor.edu*

### **Resumen:**

*En estimaciones de participación laboral, basadas en muestras de diseño complejo, generalmente no se discute el uso de ponderadores. Se considera implícitamente que los datos surgieron de un muestreo aleatorio simple.*

*Este trabajo estudia los efectos de aplicar tres alternativas de ponderación: i) no ponderar; ii) asignar pesos proporcionales al tamaño del aglomerado; iii) seleccionar una sub-muestra que permita una auto-representación. La aplicación se realiza para una onda de la Encuesta Permanente de Hogares en un modelo de ecuaciones probit simultáneas.*

*La evaluación de las estimaciones obtenidas permite concluir que no ponderar no produce sesgos relevantes y se constituye en un antecedente empírico para estimaciones con modelos no lineales para un sistema de ecuaciones con variables dependientes limitadas.*

**Palabras clave:** regresión con muestras complejas, probit simultáneos, efectos de diseño, participación laboral colectiva

**Clasificación JEL:** C13, C35, C52, J21, J22.

**Abstract:**

*Estimations of labor participation, based on samples of complex design, generally do not discuss the use of weights; implicitly considering that the data came from a simple random sampling.*

*This paper studies the effects of the application of three weighing alternatives : i) not to weigh ii) to assign weights which are proportional to agglomerate size iii) to select a sub-sample which allow self-representation. The application is carried out for a wave of the Permanent Household Survey (Encuesta Permanente de Hogares - EPH) on a model of simultaneous probit equations.*

*The evaluation of the estimates obtained suggests that not to weigh produces no significant biases and represents an antecedent to the empirical estimates with nonlinear models for a system of equations with limited dependent variables.*

Keywords: regression with complex samples, simultaneous probit equations, design effects, collective labor participation

JEL Classification: C13, C35, C52, J21, J22.

## **I. INTRODUCCION**

En estimaciones de participación laboral basadas en muestras de diseño complejo, generalmente no se discute el uso de ponderadores; considerando implícitamente que los datos surgieron de un muestreo aleatorio simple. En particular, los datos disponibles para la realización de estudios del mercado del trabajo provienen, casi exclusivamente, de muestras estratificadas de subdivisiones censales, en los aglomerados geográficos involucrados.

Al analizar datos provenientes de una encuesta con diseño muestral complejo una etapa crítica es la expansión de la información relevada en la muestra al universo de referencia. Por esto cobra importancia la selección y fundamentación de la metodología para ponderar los datos, ya que puede afectar la comparabilidad y conclusiones derivadas de los resultados.

Entre los problemas más comunes en las encuestas con distintos niveles de estratificación se pueden mencionar la introducción de sesgos en la etapa de estimación y la existencia de subgrupos de la población sub o sobre-representados. Dependiendo de la magnitud de los errores pueden registrarse sesgos pronunciados. Por esto, es importante conocer cuáles

son las limitaciones impuestas a los datos por el diseño de muestra que generó la información, a fin de evaluar si las hipótesis sobre las que se basan los métodos de ajuste y análisis se cumplen y, al mismo tiempo, determinar si las conclusiones tienen validez estadística (Medina, 1998).

En el caso de las investigaciones causales que emplean análisis de regresión, la elección del método de estimación y el procedimiento de ponderación implica una decisión substancial asociada con la calidad de resultados producidos. Por ejemplo, en los análisis que sólo utilizan el método de estimación de Mínimos Cuadrados Ordinarios (MCO), a partir de muestras complejas no se toma en cuenta el supuesto de independencia de los errores (Kish y Frankel, 1974). Si la selección está basada en valores de una variable dependiente o en una variable de diseño que se correlaciona con la dependiente pero no incluida como variable independiente, las estimaciones por MCO son en general asintóticamente sesgadas (Quesenberry *et al.*, 1986). En este contexto, distintos trabajos (Fuller, 1975; Holt *et al.*, 1980) proponen como alternativas la utilización del método de mínimos cuadrados generalizados (MCG) o de máxima verosimilitud (MV). En tales casos, el ponderador<sup>1</sup> es el inverso de la probabilidad de selección de la observación (expansores) o un factor ajustado por no respuesta y post-estratificación.

Para el caso de Argentina la principal fuente de datos oficial involucrada en estudios económicos y sociales corresponde precisamente a una muestra de diseño complejo como es la Encuesta Permanente de Hogares (EPH). El diseño de la muestra de la EPH está realizado de manera que permita representar a cada aglomerado, en términos adecuados, con el fin de garantizar un nivel aceptable de error muestral en la estimación de las tasas laborales, principal objetivos de esta encuesta.

El problema de la selección del procedimiento de ponderación es especialmente relevante debido a que la muestra del total de aglomerados no corresponde a una muestra autoponderada. Pizarro Briones (2002) advierte sobre esta restricción relacionada con muestras de diseños complejos: “no siempre es posible trabajar con diseños autoponderados y aún en este caso es necesario incorporar ponderadores para ajustar la muestra”, y destaca también tres razones por las que se recurre a estos diseños: (1) en el caso de la estratificación, con la finalidad de reducir las varianzas y obtener estimaciones a nivel de estratos; (2) la formación de conglomerados

<sup>1</sup> Un ponderador  $w_i$  para una observación  $i$  de la muestra significa que dicha observación representa a  $w_i$  elementos de la población.

dos a los efectos de reducir el costo y (3) la autoponderación, cuando se prioriza la simplicidad en la expansión de los diseños.

Las consideraciones anteriores implican que algunos aglomerados que participan con un mayor peso proporcional en la población total tienen un menor porcentaje de representación en la muestra, y viceversa. Tal como afirman Guillén *et al.* (2000), es fundamental considerar adecuadamente las características del ponderador aplicado. Es decir, en caso de no realizar ponderaciones adecuadas, los intervalos de confianza obtenidos podrían ser erróneos, debido a que las varianzas serían más elevadas que las reportadas en muestreo aleatorio simple, lo cual daría lugar a una subestimación del error muestral.

A fin de apreciar la magnitud de las posibles distorsiones, cuando se aplican ponderadores Ganninger *et al.* (2007) proponen utilizar el efecto de diseño, como medida del aumento de la varianza de un estimador causada por la utilización de un diseño complejo en relación con el esquema de muestreo aleatorio simple.

Entre distintas opciones de ponderación que han sido evaluadas, especialmente en el contexto de análisis de regresión, en el presente estudio serán consideradas tres alternativas, consistentes en estimaciones de regresiones:

- i) sin considerar ningún tipo de ponderación
- ii) utilizando una ponderación proporcional al tamaño del aglomerado
- iii) reduciendo el tamaño de la muestra a fin de su autorepresentación.

El modelo que se empleó para comparar las alternativas de ponderación considera como variable de estudio la decisión de participación laboral conjunta de los miembros de parejas. El modelo fue desarrollado por Herrero (2007) y en este caso será aplicado a las observaciones muestrales provenientes del segundo semestre de 2005 de la EPH. En este modelo se consideran simultáneamente dos ecuaciones que dan cuenta de dichas decisiones.

El objetivo de este estudio es evaluar las tres opciones de ponderación consideradas, a fin de analizar las ventajas y costos de cada una en términos de complejidad, posibles sesgos y eficiencia, para un caso donde se incorpora, como elemento particular, el análisis de los resultados de un modelo de ecuaciones probit simultáneas.

El trabajo se organiza de la siguiente manera: la sección II contiene una descripción del modelo considerado, los datos empleados y las alternativas de tratamiento de los datos que se comparan; incluye también la descripción de los procedimientos a través de los cuales se propone realizar la confrontación de alternativas. En la sección III se da cuenta de los resultados generados con cada alternativa evaluada, con un análisis de los mismos. En la siguiente sección IV, se procede a la comparación de resultados y su discusión con respecto a la bibliografía existente. Finalmente, se resumen los principales hallazgos y conclusiones del trabajo.

## II. MATERIALES Y MÉTODOS

### II.1. Modelo estadístico

Cuando se analiza la participación efectiva o no en el mercado laboral, es posible observar y registrar el comportamiento individual. Ahora bien las utilidades de dicha participación son constructos teóricos no observables, por lo cual para su cuantificación necesitan estimarse a partir de modelos de variables latentes. Si con una variable dicotómica  $y_{ji}$  representamos la concreción de la mayor o menor utilidad de participar ( $U_{ji}$ ) contra la de no participar ( $\bar{U}_{ji}$ ) en el mercado de trabajo, la decisión de participación de cada cónyuge está dada por:

$$y_{ji} = \begin{cases} 1 & \text{si } U_{ji} - \bar{U}_{ji} > 0 \\ 0 & \text{si } U_{ji} - \bar{U}_{ji} \leq 0 \end{cases} \quad (1)$$

y la decisión de participación simultánea de ambos cónyuges, define el sistema:

$$\begin{cases} y_{1i}^* = X_i' \pi_1 + v_{1i}; & y_{1i} = 1 \quad \text{si } y_{1i}^* > 0, \quad 0 \text{ en caso contrario} \\ y_{2i}^* = X_i' \pi_2 + v_{2i}; & y_{2i} = 1 \quad \text{si } y_{2i}^* > 0, \quad 0 \text{ en caso contrario} \end{cases}$$

$$E(v_1) = E(v_2) = 0$$

$$V(v_1) = V(v_2) = 1$$

$$\text{cov}(v_1, v_2) = \rho \quad (2)$$

En este sistema,  $\pi_j$  corresponde al vector de coeficientes asociados a los respectivos determinantes (matriz  $X$  de variables explicativas) en la ecuación de participación de cada miembro de la pareja. El término  $v_{ji}$  corresponde a una variable aleatoria, cuya magnitud muestra la diferencia entre el valor esperado de la participación para cada individuo  $j$  y para cada pareja  $i$  de la población estudiada y la verdadera participación del individuo respectivo. El supuesto de que los términos aleatorios de las ecuaciones están correlacionados permite considerar que ambas decisiones tienen aspectos de interinfluencia en el caso de la participación laboral de los cónyuges.

Mallar (1977) desarrolla el criterio de identificación de sistemas de ecuaciones simultáneas en modelos de probabilidad, básicamente idéntico al correspondiente a aquellos para sistemas de ecuaciones lineales. A fin de posibilitar la identificación de los parámetros del modelo se requiere que, al menos una variable de una de las ecuaciones no sea explicativa de la otra ecuación. En el caso particular que se presenta, hay más de una variable que influye en la participación de uno de los cónyuges que, en principio, no afectaría la participación del otro.

En ambas ecuaciones (2) la variable dependiente es dicotómica, con valor 1 en el caso que el respectivo cónyuge participe y 0 si no participa en el mercado laboral.

En el trabajo de Douguet y Simonnet (2004), las estimaciones de este sistema de ecuaciones se desarrollan a través del método de mínimos cuadrados asintóticos. En este trabajo se utilizó la estimación por máxima verosimilitud provista por el procedimiento biprobit disponible en STATA.

## II.2. Variables del modelo

A partir de las especificaciones empleadas en diferentes estudios sobre la decisión de participación laboral de los cónyuges, se seleccionaron distintas variables para la estimación.

Las variables incluidas como explicativas de la participación laboral de la cónyuge mujer fueron:

- *Edad de la cónyuge*. Se espera que la edad presente una relación cuadrática con respecto a la participación, alcanzando su máxima participación a una edad  $x_0$ , luego de la cual comienza a decrecer (Wong y Levine, 1992). Para capturar este efecto se incorporan dos representaciones de la variable (edad y edad al cuadrado),

esperando signo positivo para la primera y negativo para la segunda.

- *Tener algún hijo menor de un año*: Haber tenido un bebé en el último año disminuye la probabilidad de participación de las esposas en todos los estudios que dan cuenta de la participación laboral femenina.
- *Número de hijos de 1 a 5 años*: La relación negativa entre el número de hijos en edad pre-escolar y la probabilidad de participación de las esposas también es un resultado coincidente en diferentes investigaciones. Ejemplos, a partir de diferentes abordajes técnicos de la cuestión se encuentran en Douguet y Simmonet (2004), Sasaki (2002), Shaw (1994), entre otros. Es de esperar que las tareas de cuidado y supervisión de los hijos en este intervalo etario demanden una carga importante de tiempo, ya que son generalmente desarrolladas por las madres dado el esquema tradicional de división de roles por géneros en los hogares.
- *Número de hijos de 6 a 12 años*: La cantidad de hijos en edad escolar primaria se espera que afecte negativamente la participación de las esposas.
- *Número de hijos de 13 a 17 años*: En el caso de este intervalo etario de los hijos, puede esperarse que algunos contextos favorezcan la participación de las cónyuges y otros las retraigan.
- *Presencia de mujer inactiva de 18 años y más en el hogar*: Diversos estudios incorporan la presencia de alguna potencial ayuda para tareas domésticas por parte de otra persona residente en el hogar, como un facilitador de la participación laboral de las mujeres (Wong y Levine, 1992; Orcutt Duleep y Sanders, 1993). Sin embargo, debe considerarse que la presencia de personas mayores inactivas también podría estar relacionada con la convivencia con ancianos que, en caso de requerir cuidados y/o atención especial, suelen afectar en mayor medida y en sentido negativo la participación de la esposa (Ettner, 1995).
- *Participación laboral del esposo*: El hecho de que el esposo participe en el mercado laboral se podría incidir positivamente en la probabilidad de participación de la esposa. La empleabilidad de los miembros del hogar podría ser potenciada entre sí a través de la disposición de información más precisa sobre condiciones y opor-

tunidades en el mercado, o bien como viabilizadora de redes de contactos. No obstante, si bien no surge de las aplicaciones empíricas recientes consideradas, es plausible una relación negativa, en tanto el efecto ingreso de la participación del cónyuge varón, determinara una menor propensión a la participación de la mujer.

- *Participación laboral femenina en el aglomerado*: Mientras más aceptada es la actividad de las mujeres en el trabajo extra-doméstico, mayor es la probabilidad de que una esposa en particular participe.
- *Esposo desocupado*: Como postula la hipótesis del trabajador adicional, la situación de desempleo del esposo se asocia con una probabilidad mayor de participación de la esposa (ya sea que comienza a participar por esa situación coyuntural, o porque es menos propensa a abandonar un empleo).
- *Estudios de la esposa*: Poseer mayor nivel de estudios se espera que aumente la probabilidad de participación de la esposa (tal como explica detalladamente Larrañaga, 2006). El efecto puede considerarse una consecuencia del costo de oportunidad de las personas con mayor inversión en educación, tal como explica Wainerman (1979). Para su análisis se consideran en el modelo los distintos niveles de educación alcanzados.
- *Poseer más escolaridad que el cónyuge*: La variable trata de captar la relación de igualdad entre géneros al interior de la pareja. La situación de escolaridad de la cónyuge respecto de su pareja es mencionada como un indicador de mayor o menor equidad de género (Pinelli, 2004): se espera que aquellas con mayor nivel educativo que su cónyuge tomen decisiones de manera más igualitaria en sus hogares y sean probablemente más independientes y/o más propensas a participar laboralmente.
- *Tasa de desempleo del aglomerado*: Se podrían esperar dos efectos contrapuestos, y el signo sería la consecuencia de la mayor incidencia de uno sobre otro en la decisión de participación laboral. El signo de la relación sería positivo si debido a la situación de desempleo elevado, las cónyuges tratan de conseguir trabajo o no abandonan el que poseen, como respuesta a una condición de mayor inseguridad económica relativa. El signo sería el opuesto, negativo, si debido a las peores condiciones para conseguir o man-

tener un empleo en el aglomerado de referencia, las cónyuges dejan de buscar trabajo (efecto desaliento), tal como encuentran Leibowitz y Klerman (1995).

- *Región en la que residen*: Se incorpora una variable dicotómica, que señala con valor 1 (uno) a los hogares residentes en el Área Metropolitana del país (Ciudad de Buenos Aires y Partidos de Gran Buenos Aires) y 0 (cero) a los hogares residentes en el resto del país. Esta variable permite captar la diferencia en la probabilidad de participación, ya controlados los demás factores, (algunos de los cuales son las causas de las diferencias entre regiones, como mayor escolaridad de la población, mayor incidencia del desempleo o diferente proporción de mujeres activas laboralmente, conformación del grupo familiar, entre otras).
- *Estado conyugal*: La variable dicotómica identifica con valor 1 a las parejas en unión libre, con respecto del resto que corresponden a parejas casadas. Estudios previos dan cuenta de una relación positiva entre estar en unión libre y participación de la cónyuge, por un lado por la posible mayor inestabilidad de estas uniones, y en otros contextos por una mayor autonomía de las mujeres en tal condición. Como advierte Castro Martín (1997) es probable que la composición de las parejas en unión libre se caracterice por dos segmentos, uno con mujeres en una condición de desigualdad muy marcada en su pareja y otro con mujeres en situación de gran independencia.
- *Son propietarios de la vivienda*: Puede esperarse signo positivo o negativo, según cuál sea el efecto que en definitiva prevalezca. Puede ser negativo si implica una menor necesidad de recursos corrientes con respecto a quienes deben enfrentar el pago mensual de un alquiler; puede, por el contrario, ser positivo, si lo consideramos un indicador de mejor situación económica relativa, y esperamos que las esposas de hogares más favorecidos sean las que tienen más participación laboral.
- *Ocupan la vivienda de manera irregular*: El grupo de hogares que ocupa la vivienda en la que reside, con respecto a quienes alquilan, tiene menos necesidad de recursos corrientes para cumplir con un pago de cierto monto fijo en una fecha estipulada, por lo que el efecto podría ser negativo. No obstante, debe considerarse que quienes ocupan la vivienda irregularmente pueden ser relativa-

mente tan pobres que la necesidad de ingresos determinen una mayor participación laboral de las esposas.

Las variables incluidas como explicativas de la participación laboral del cónyuge varón son:

- *Edad del cónyuge y edad al cuadrado*: El efecto de la edad sobre la participación laboral del cónyuge varón es análoga a la descrita para la relación correspondiente para la participación de la cónyuge (relación cuadrática).
- *Hogar no nuclear*: El efecto esperado es positivo. Este tipo de hogares incluye o bien la presencia de padres y/o suegros, o de hijos con sus respectivas familias. Gong y van Soest (2002) consideran especialmente las implicaciones sobre la participación laboral de estas estructuras familiares.
- *Poseen hijos menores de un año*: El efecto que se espera es positivo. Una mayor necesidad de recursos económicos, especialmente en el período en que muchas esposas dejan de trabajar para cuidar a los hijos pequeños, pueden impulsar a los esposos a una mayor participación laboral.
- *Número de hijos de 1 hasta 5 años, número de hijos de 5 hasta 12 años y número de hijos de 12 hasta 17 años*: Se espera que la probabilidad de participación laboral sea mayor entre los que tienen más hijos en estos intervalos de edades, debido a las mayores necesidades de ingresos.
- *Vive mujer inactiva mayor de 18 años en el hogar*: Se espera que no tenga efecto significativo para el hombre.
- *Participación laboral de la esposa*: Trata de captar en qué medida la decisión del esposo es independiente o no de la decisión de la cónyuge mujer de participar o no.
- *Nivel educativo del esposo*: Como indicador indirecto de una mejor remuneración relativa del trabajo, se espera que la participación sea mayor entre los que poseen una mayor escolarización, si bien podría esperarse un efecto no tan elevado como para la mujeres, debido en principio, a la marcada división de roles al interior de los hogares.
- *Región en la que residen*: Con esta variable se puede captar, como

en el caso de la participación laboral de las cónyuges, el efecto de otros aspectos no considerados y asociados puramente con diferencias regionales.

- *Son propietarios de la vivienda y ocupan la vivienda de manera irregular*: Se espera una relación similar a la descrita para la participación de la cónyuge.

### II.3. Muestra analizada

La muestra de trabajo parte de los datos proporcionados en las Bases Usuarías de la EPH. Para esta aplicación, sólo se incluyeron las parejas (casadas o unidas) conformadas por el Jefe de hogar y su cónyuge, ya que son las únicas identificables en la EPH.

La medición del segundo semestre de 2005 de la EPH permite realizar estimaciones sin pérdida de datos por falta de respuesta en todas las variables consideradas. El tamaño de la muestra considerada en la base fue de 10.980 parejas y el total de parejas residentes en los principales aglomerados urbanos del país, representadas una vez expandidos los casos fue de 3.313.304 parejas.

El diseño de la muestra de la EPH (INDEC, 2003) puede considerarse complejo porque en su construcción es multietápica, contempla tres aspectos básicos de dimensiones distintas: el dominio geográfico a estudiar (aglomerado), radios censales (unidades muestrales seleccionadas en una primera etapa) y estratificación de los radios según características de la población (datos sobre la vivienda, el nivel de instrucción de las personas, etc.), tal como se detalla en Messere y Hoszowski (2000).

### II.4. Alternativas de ponderación

#### *i) Estimación sin ponderar*

Una primera estimación del modelo se aplicó sin utilizar ningún tipo de ponderador. Como enfatiza Sanchez Carrión (2000) la decisión de no ponderar la muestra con ningún tipo de ajuste por el problema muestral, es en sí misma una elección de ponderación particular. Esta primera alternativa consideró un diseño muestral autoponderado, ya que todos los hogares tienen el mismo ponderador. Se trabajaron los datos en la estimación como si en la muestra de la EPH todos los hogares tuvieran la misma

probabilidad de selección. De esta manera, la probabilidad de selección sería la misma para todos al interior de cada estrato. Por otra parte, se procedió como si la fracción de muestreo  $f$  correspondiente al estrato  $h$  fuera la misma para todos los estratos de la EPH, esto es, todos los elementos de la muestra comparten el mismo ponderador de selección, el cual viene dado por la siguiente expresión:

$$w_{ij} = \frac{1}{f} = \frac{\text{total de personas del aglomerado}}{\text{total de personas de la muestra}} \quad (3)$$

*ii) Estimación con ponderadores que reflejan el peso relativo de las muestras por aglomerados*

La segunda opción consistió en ponderar cada caso con los pesos relativos (no absolutos como se produciría a través de los expansores) que les corresponderían en una muestra auto-ponderada. Los mismos pueden expresarse como:

$$\text{con } \hat{p}_{ij} = \frac{\sum_{k \in K} w_k y_{kij}}{\sum_{k \in K} w_k} \quad \text{con } y_{kij} = \begin{cases} 1 & \text{si corresponde a la pareja } i \\ & \text{del aglomerado } j \\ 0 & \text{en caso contrario} \end{cases} \quad (4)$$

donde:

$$\hat{p}_{ij} = \frac{\text{Suma de pesos para la pareja } i \text{ del aglomerado } j}{\text{Suma de pesos para todas las parejas de la muestra}} \quad (5)$$

La estructura de la matriz que contiene los respectivos ponderadores se representa de la siguiente manera:

$$\begin{array}{cccc|c} & & & & \text{Total} \\ \begin{matrix} \hat{p}_{11} & \hat{p}_{12} & L & \hat{p}_{1n} \\ \hat{p}_{21} & \hat{p}_{22} & L & \hat{p}_{2n} \\ M & M & O & M \\ \hat{p}_{m1} & \hat{p}_{m2} & L & \hat{p}_{mn} \end{matrix} & \begin{matrix} \hat{p}_{1+} \\ \hat{p}_{2+} \\ M \\ \hat{p}_{m+} \end{matrix} & \\ \text{Total} & \begin{matrix} \hat{p}_{+1} & \hat{p}_{+2} & L & \hat{p}_{+n} \end{matrix} & \begin{matrix} | & 1 & | \end{matrix} & & \end{array} \quad (6)$$

donde  $n$  el número de parejas en cada uno los  $m$  aglomerados de aplicación.

*iii) Estimación basada en la máxima muestra autoponderada posible de los datos*

La tercer alternativa, que evita el incremento artificial del tamaño de la muestra, implicaba tomar como referencia, en primer lugar al aglomerado que tenga la peor relación de sub-representación (es decir, aquel en el cual se deberían haber tomado más casos en una muestra autoponderada y finalmente participa con un menor porcentaje relativo en la muestra recogida). Luego, se procedió a considerar el tamaño de la muestra total si se utilizaran todos los casos de ese aglomerado, y a partir de ese valor se determinó la participación relativa de los demás en la muestra total.

En los restantes aglomerados se tomaron muestras al azar, de manera que sólo quedaron tantos casos como hubieran correspondido en una muestra autoponderada.

## **II.5. Comparación de las alternativas de ponderación**

### *Bondad de ajuste*

La evaluación de la bondad de ajuste requiere considerar una medida resumen que señale el grado de adecuación con que una alternativa del modelo especificado se aproxima los datos observados. Para modelos con variable dependiente limitada, Maddala (1990) sugiere que la precisión puede ser juzgada o bien en términos del ajuste entre las probabilidades calculadas y la frecuencia de respuestas observadas, o en términos de las respuestas observadas predichas por el modelo. En este estudio se seleccionó, entre varias opciones analizadas, comparar las diferencias que podrían aparecer en el porcentaje de casos correctamente predichos por el modelo estimado considerando cada alternativa de ponderación.

### *Efecto de diseño e ineficiencia*

El indicador introducido por Kish (1995) para medir las consecuencias de utilizar un diseño de muestra distinto del aleatorio simple en las estimaciones de parámetros, ha sido ampliamente extendido y aplicado, en particular, a los casos asociados con la estimación de regresiones. Con énfasis en los procedimientos vinculados con la estimación de las varianzas correspondientes, Esquivel Bocanegra *et al.* (2003) retoman la definición clásica de Efecto de Diseño (ED). Para un estimador  $\beta$  de un coeficiente  $\beta$ , el ED se define como:

$$ED_w(\hat{\beta}) = \frac{V_w(\hat{\beta})}{V_{SP}(\hat{\beta})} \quad (5)$$

donde  $V$  indica la varianza del respectivo estimador,  $w$  se refiere al diseño de muestreo complejo (o el procedimiento con ponderación correspondiente en el caso de los incorporados en este estudio); mientras que,  $SP$  se refiere a un esquema de muestreo aleatorio simple (el procedimiento i) de estimación sin ponderar). A los efectos del cálculo, los autores mencionados proponen un estimador  $ed$ :

$$ed_w(\hat{\beta}) = \frac{\hat{V}_w(\hat{\beta})}{\hat{V}_{SP}(\hat{\beta})} \quad (7)$$

La utilización de este indicador permite evaluar la conveniencia de aplicar cierto procedimiento de ponderación a los datos muestrales, frente a la opción de utilizarlos suponiendo que cada observación proviene de una selección con igual probabilidad. Basados en los valores surgidos del  $ed$ , Medina (1998) sugiere una interpretación para valorar los resultados alternativos:

$$ed \begin{cases} <1 & \text{el interalo de confianza es muy grande} \\ =1 & \text{el interalo de confianza es correcto} \\ >1 & \text{el interalo de confianza es muy pequeño} \end{cases} \quad (8)$$

Un concepto complementario, corresponde a la idea de Ineficiencia ( $If$ ) reportada en Korn y Graubard (1995); éstos autores hacen énfasis en analizar más bien la pérdida de eficiencia provocada por utilizar una regresión ponderada frente a la sin ponderar, que de hecho es insesgada (Fuller, 1984 y Skinner, 1986).

$$If = 1 - \frac{\hat{V}_{SP}(\hat{\beta})}{\hat{V}_w(\hat{\beta})} \quad (9)$$

En este sentido sólo recomiendan utilizar las alternativas ponderadas si la ineficiencia es baja (ya sea en términos absolutos como relativos a los intervalos de confianza obtenidos para los coeficientes de regresión estimados). Korn y Graubard (1995) destacan que “una ineficiencia baja es un precio pequeño a pagar por una estimación potencialmente más robusta de los errores estándar”. El cálculo de la ineficiencia adquiere sentido cuando el efecto de diseño indica que resulta más conveniente el uso de ponderadores para la estimación.

### III. RESULTADOS

#### *i) Estimación sin ponderar*

Con este procedimiento, las variables que no resultaron significativas, para el caso de la probabilidad de participación de la cónyuge fueron: edad al cuadrado, reside en el hogar otra mujer (inactiva), la cónyuge tiene estudios primarios completos o secundario incompleto, tasa de desocupación del aglomerado y situación de tenencia de la vivienda.

En tanto que para la probabilidad de participación del hombre, no resultaron significativas: hogar no nuclear, poseer hijos menores de 1 año o hijos de 13 a 17 años, reside en el hogar otra mujer (inactiva). Para el cónyuge varón, el nivel educativo alcanzado y la situación de tenencia de la vivienda, no diferencian su probabilidad de participación laboral.

*Ventajas y desventajas del procedimiento:* Los resultados obtenidos, muestran el inconveniente de no considerar la importancia poblacional relativa de los datos provenientes de los diferentes aglomerados. Este procedimiento deriva en coeficientes más influenciados (sesgados) por aquellos aglomerados sobrerrepresentados. Pese a esta limitación, su principal beneficio es que este procedimiento por ser el más utilizado en los diferentes estudios que toman la EPH como base de análisis, permite resultados comparables.

#### *ii) Estimación con ponderadores que reflejan el peso relativo de las muestras por aglomerados*

Con este procedimiento, las variables que no resultaron significativas, para el caso de la probabilidad de participación de la cónyuge fueron: reside en el hogar otra mujer (inactiva), tasa de participación femenina y de desocupación general del aglomerado, forma de tenencia de la vivienda y si reside en el Área Metropolitana.

En tanto que para la probabilidad de participación del hombre, no resultaron significativas: tipo de hogar, poseer hijos menores de 1 año o en grupo etario de 13 a 17 años, si reside en el hogar otra mujer inactiva. Para el cónyuge varón, el nivel educativo alcanzado y la situación de tenencia de la vivienda, no diferencian su probabilidad de participación laboral.

*Ventajas y desventajas del procedimiento:* El beneficio de este procedimiento es que al considerar ponderadores corrige el sobrepeso de las observaciones provenientes de aglomerados sobrerrepresentados, y de la reducida presencia de casos de los restantes. Por contrapartida, una de las limitaciones, es que de manera ficticia se están replicando las observaciones que

corresponden a los aglomerados con mayor importancia real que la reflejada en la muestra. A partir de la repetición de las observaciones se reduce, posiblemente, la variabilidad intrínseca de las relaciones que se intenta modelar.

*iii) Estimación basada en la máxima muestra autoponderada posible de los datos*

Con este procedimiento de ponderación, y a fin de que la muestra de la EPH sea autoponderada, se debe reducir la misma. En la siguiente Tabla 1 se presenta la reducción del tamaño muestral necesaria para aplicar esta alternativa.

**Tabla 1**  
**Tamaño del aglomerado en la muestra original y en la muestra reducida para otorgar autorepresentación**

Aglomerado	Cantidad de observaciones de la muestra original	% de la muestra original	Cantidad de observaciones de la muestra reducida	% de la muestra reducida	Pérdida de casos con información	Pérdida en % de la muestra disponible
Gran La Plata	546	5.0	175	3.5	371	67.9
Bahía Blanca	387	3.5	80	1.6	307	79.3
Gran Rosario	791	7.2	295	6.0	496	62.7
Santa Fe y S. Tomé	406	3.7	105	2.1	301	74.0
Gran Paraná	364	3.3	59	1.2	305	83.7
Com. Rivadavia	284	2.6	34	0.7	250	88.0
Gran Mendoza	699	6.4	193	3.9	506	72.3
Corrientes	379	3.5	64	1.3	315	83.0
Gran Córdoba	834	7.6	296	6.0	538	64.5
Formosa	293	2.7	45	0.9	248	84.8
Neuquén y Plottier	248	2.3	60	1.2	188	75.7
S. S. Jujuy y Palpalá	347	3.2	51	1.0	296	85.2
Río Gallegos	300	2.7	23	0.5	277	92.4
Gran Catamarca	260	2.4	37	0.7	223	85.9
Salta	402	3.7	87	1.8	315	78.3
La Rioja	277	2.5	31	0.6	246	88.7
San Luis-El Chorrillo	264	2.4	42	0.8	222	84.1
Gran San Juan	482	4.4	102	2.1	380	78.9
Santa Rosa y Toay	262	2.4	31	0.6	231	88.1
Ciudad de Buenos Aires	766	7.0	746	15.1	20	2.6
Partidos de Gran Buenos Aires	2389	21.8	2389	48.3	0	0
Total	10980	100.0	4945	100.0	6035	

Fuente: Elaboración propia con datos de INDEC (BUA - EPH 2º semestre de 2005)

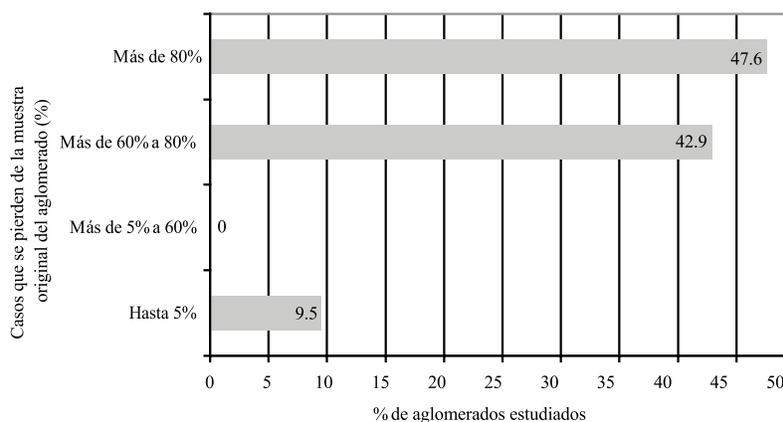
Las variables que no resultaron significativas, para un aumento de la probabilidad de participación de la cónyuge fueron: edad al cuadrado, residencia en el hogar otra mujer (inactiva), la cónyuge tiene estudios primarios completos o secundario incompleto, la cónyuge mujer tiene más escolaridad que su cónyuge, tasa de participación laboral femenina y tasa de desocupación del aglomerado, así como el tipo de ocupación de la vivienda.

En tanto que para la probabilidad de participación del hombre, las variables relacionadas con la edad del cónyuge, el hogar no nuclear, poseer hijos menores de 1 año, o en los grupos etéreos de 6 a 17 años, la residencia en el hogar de otra mujer (inactiva), el nivel educativo alcanzado y la situación de tenencia de la vivienda, no resultaron significativas

*Ventajas y desventajas del procedimiento:* Los defectos a primera vista de esta alternativa son evidentes: en primer lugar se pierden una cantidad importante de casos, lo cual atenta contra el principio de aprovechar la mayor cantidad de información muestral disponible; Como se muestra en el Gráfico 1, exceptuando sólo dos aglomerados (Ciudad de Buenos Aires y Partidos del Gran Buenos Aires) el porcentaje de pérdida de casos en la muestra disponible se ubica por encima del 60% y para ocho aglomerados supera el 85%.

Otra desventaja, la constituye la reducción de la variabilidad, debida al menor tamaño muestral, que para el análisis de algunas variables puede ser relevante.

**Gráfico 1**  
**Aglomerados de la muestra original (en %) según representatividad que pierden para otorgar autorepresentación**



Fuente: Elaboración propia

#### IV. COMPARACIÓN Y DISCUSIÓN DE RESULTADOS

En general, los signos y magnitudes que resultaron después de aplicar los tres procedimientos de ponderación son consistentes con los obtenidos en diversos trabajos empíricos. En modelos de participación laboral los resultados comunes para diferentes países, muestran que un mayor nivel educativo aumenta la probabilidad de participación, un aumento del ingreso de otro miembro de la familia la reduce, al igual que la presencia de hijos económicamente dependientes reduce la participación en el caso de las mujeres mientras que para los hombres hay una mayor variabilidad (Hernández *et al.*, 2005; Benven *et al.*, 2007).

Cabe señalar que en este caso, la fuerte vinculación entre las variables independientes de cada ecuación se presumiría que podía asociarse con multicolinealidad, y las implicancias en la calidad de las estimaciones (en particular de los correspondientes errores estándar). Sólo evidenciaron algún grado de correlación lineal (en el primer caso negativa y en el segundo positiva) entre sí, las siguientes variables en la ecuación de la cónyuge.

- Edad de los cónyuges y cantidad de hijos en las primeras edades
- Residencia en Área Metropolitana de la cónyuge y Participación laboral femenina en el aglomerado

##### IV.1. Coeficientes significativos en cada procedimiento de ponderación

En la Tabla 2 se resumen las estimaciones de la probabilidad de participación de cada cónyuge, para las alternativas evaluadas. Para cada variable que resultó significativa, en al menos uno de los procedimientos, se presenta el coeficiente estimado que indica el efecto del cambio de la misma sobre la probabilidad del cónyuge respectivo analizado, considerando las demás constantes, así como su valor de probabilidad  $p$ .

**Tabla 2**  
**Comparación de coeficientes para las tres**  
**alternativas de ponderación**

Variables	Sin ponderar		Ponderación proporcional al tamaño		Reducción de tamaño de muestra	
	Coefficiente estimado	Valor <i>p</i>	Coefficiente estimado	Valor <i>p</i>	Coefficiente estimado	Valor <i>p</i>
<b>Probabilidad de participación de la cónyuge mujer</b>						
Edad de la cónyuge	<b>-0.186</b>	0.043	<b>-0.358</b>	0.001	<b>-0.331</b>	0.071
Edad de la cónyuge al cuadrado	0.085	0.300	<b>0.194</b>	0.047	0.193	0.265
Posee hijos menores de 1 año	<b>-0.391</b>	0.000	<b>-0.345</b>	0.000	<b>-0.296</b>	0.001
Número de hijos de 1 a 5 años	<b>-0.114</b>	0.000	<b>-0.201</b>	0.000	<b>-0.195</b>	0.000
Número de hijos de 6 a 12 años	<b>-0.102</b>	0.000	<b>-0.103</b>	0.000	<b>-0.088</b>	0.000
Número de hijos de 13 a 17 años	<b>-0.041</b>	0.001	<b>-0.034</b>	0.040	<b>-0.050</b>	0.011
El cónyuge varón participa en el mercado laboral	<b>1.697</b>	0.000	<b>1.724</b>	0.000	<b>1.616</b>	0.000
La cónyuge tiene estudios primarios completos o secundario incompleto	0.002	0.958	<b>-0.078</b>	0.060	-0.059	0.365
La cónyuge mujer tiene estudios secundarios completos o superiores incompletos	<b>0.097</b>	0.026	<b>0.099</b>	0.055	<b>0.126</b>	0.095
La cónyuge tiene estudios superiores completos	<b>0.918</b>	0.000	<b>0.858</b>	0.000	<b>0.977</b>	0.000
La cónyuge mujer tiene más escolaridad que su cónyuge	<b>0.076</b>	0.006	<b>0.104</b>	0.001	0.064	0.119
Tasa de participación laboral femenina en el aglomerado	<b>0.034</b>	0.012	0.011	0.368	0.028	0.234
El cónyuge varón está desocupado	<b>0.265</b>	0.000	<b>0.332</b>	0.000	<b>0.322</b>	0.000
Estado conyugal = unido	<b>0.199</b>	0.000	<b>0.230</b>	0.000	<b>0.181</b>	0.000
Residen en el área metropolitana del país	<b>-0.077</b>	0.022	-0.028	0.411	<b>-0.099</b>	0.068
Constante	<b>-1.622</b>	0.000	<b>-1.644</b>	0.000	<b>-1.487</b>	0.000
<b>Probabilidad de participación del cónyuge varón</b>						
Edad del cónyuge	<b>0.188</b>	0.001	<b>0.264</b>	0.000	-0.069	0.501
Edad del cónyuge al cuadrado	<b>-0.736</b>	0.000	<b>-0.950</b>	0.000	-0.142	0.595
Número de hijos de 1 a 5 años	<b>0.169</b>	0.000	<b>0.179</b>	0.000	<b>0.119</b>	0.032
Número de hijos de 6 a 12 años	<b>0.080</b>	0.004	<b>0.123</b>	0.000	0.055	0.196
La cónyuge mujer participa en el mercado laboral	<b>1.774</b>	0.000	<b>1.783</b>	0.000	<b>1.683</b>	0.000
Residen en el área metropolitana del país	<b>0.152</b>	0.000	<b>0.184</b>	0.000	<b>0.148</b>	0.009
Constante	<b>0.287</b>	0.002	<b>0.221</b>	0.010	<b>0.327</b>	0.002

*Negrita: Coeficientes distintos de cero al 10% de significación. Todas las variables cuantitativas han sido estandarizadas.*

*Fuente: Elaboración propia con datos de INDEC (BUA - EPH 2° semestre de 2005)*

### **a. Determinantes de la participación laboral de la cónyuge mujer**

En las diferentes alternativas de ponderación se capturan las relaciones más importantes reportadas en las referencias sobre la temática.

Entre las variables que resultan significativas y aumentan la probabilidad de participación, se destacan las que tienen que ver con el nivel educativo alcanzado, con la participación o no del cónyuge varón en el mercado del trabajo y con el estado conyugal de la mujer. Estos resultados concuerdan con lo que expresan, entre otros, Cerrutti y Zenteno (2000) quienes consideran el capital humano (educación formal, no formal e informal, experiencia laboral previa, etc.) y el ciclo de vida individual y familiar (edad, estado civil, posición en el hogar, número y edad de los hijos, estructura familiar, etc.) como variables de atención en los estudios sobre determinantes de la participación femenina en los mercados de trabajo.

Los coeficientes estimados para la variable que da cuenta de la participación del cónyuge varón, presentan valores similares en los tres tipos de ponderaciones aplicadas, que señalan efecto positivo en la participación de la mujer. El efecto de que el cónyuge varón se encuentre desocupado es análogo en las tres metodologías, si bien el valor absoluto del mismo es levemente menor en la alternativa sin ponderación. Estos resultados reflejan la mayor importancia que cobran en la muestra los aglomerados con menos porcentaje de la población nacional (sobrerrepresentados).

El estado conyugal “unido” incide positivamente en la participación femenina con efecto similar, cualquiera sea la alternativa de ponderación analizada. Este efecto podría deberse o bien al reflejo de la mayor inestabilidad percibida frente a “estar casado”, lo cual impulsaría a las mujeres al mercado laboral, como resguardo ante una disolución, o alternativamente, deberse al propósito de mantener cierta independencia económica individual.

Las variables que disminuyen la probabilidad de participación de la mujer, siempre considerando los tres tipos de ponderaciones propuestas, son: edad de la cónyuge; poseer hijos menores de 1 año y el mayor número de hijos en diferentes edades. Es decir, la estructura familiar tiene un impacto importante y negativo sobre la probabilidad de participación de la

cónyuge.

En este caso vemos que la presencia de hijos en el hogar es un fuerte condicionante, (detectada con signo negativo) de la probabilidad de participación femenina, decreciente a medida que mayor sea la edad de los hijos (valor absoluto de los coeficientes cada vez menores).

Por otro lado, la edad de la mujer presenta una menor incidencia (menor valor absoluto) cuando se consideran los datos provenientes de la EPH en el formato que denominamos sin ponderar.

Las variables que no son cuantificadas como estadísticamente válidas en los tres procedimientos analizados, fueron:

- La cónyuge tiene estudios primarios completos o secundario incompleto, en este caso sólo la segunda alternativa de ponderación da cuenta de una influencia en sentido negativo; debe observarse que esta variable tiene, de todas maneras, un efecto cuantitativo muy reducido y puede deberse a que las observaciones repetidas de esta alternativa replican más unidades que las realmente observadas en algunos aglomerados.
- La cónyuge mujer tiene más escolaridad que su cónyuge, esta variable da cuenta de la consolidación de la sociedad basada en el conocimiento; en la actualidad, la mujer ocupa cada vez más espacios en el ámbito social y especialmente en el laboral, por tanto, sería esperable que en los tres tipos de ponderación se observaran probabilidades que inciden en forma positiva. Aquí seguramente la alternativa tres, como consecuencia de una pérdida en la cantidad de datos que considera para el análisis o por la reducción de la variabilidad de la misma, no tiene poder de discriminación.
- Tasa de participación laboral femenina en el aglomerado, esta variable sólo es detectada y está positivamente relacionada con un aumento de la participación de la mujer en el primer procedimiento; aunque el impacto de esta variable es muy reducido sobre la variable dependiente estudiada.
- Que la residencia de la mujer sea el área metropolitana del país afecta negativamente su probabilidad de participación en el mercado ocupacional; esta influencia es sólo informada como estadísticamente significativa cuando se consideran las primera y

tercera alternativas. En relación con el signo del coeficiente, puede estar vinculado con la influencia de otras variables. Este efecto puede también deberse a lo planteado por Quesenberry *et al.* (1986) acerca de los sesgos debidos a la interrelación entre selección y variable independiente.

### **b. Determinantes de la participación laboral del cónyuge varón**

Para las tres formas de ponderaciones propuestas, el modelo utilizado da cuenta de las siguientes variables que inciden en un aumento de la probabilidad de participación del hombre: número de hijos de 1 a 5 años, la cónyuge mujer participa en el mercado laboral y residencia en el área metropolitana del país.

Como se observa en la Tabla 2, para el caso del cónyuge varón, la participación de su compañera en el mercado del trabajo es la razón de mayor peso en su propia participación.

En cuanto al lugar de residencia, vivir en el área metropolitana del país incide en la participación masculina, aunque con valores sensiblemente menores a los que dan cuenta de la participación de la mujer que conforma la pareja.

No se registran, para los tres tipos de ponderaciones propuestas, ninguna variable (estadísticamente significativa) que disminuyan la probabilidad de participación del cónyuge varón.

Análogamente y para la participación del hombre integrante de la pareja se destacan, por ser captadas de distinta manera a partir de los tres procedimientos analizados, las variables:

- Edad y edad al cuadrado del cónyuge: Resultaron significativas y con los signos esperados para las alternativas primera y segunda. Seguramente la razón por la cual el tercer procedimiento no captura este efecto corresponde a la ya mencionada para el caso de la participación laboral femenina (reducción del tamaño muestral).
- Se observa el mismo resultado para la variable número de hijos de 6 a 12 años.

## IV.2. Comparación de los procedimientos de ponderación

### *Bondad de ajuste del modelo*

Al medir la bondad de ajuste del modelo, considerando la proporción de casos totales bien clasificados o de predicciones correctas<sup>2</sup>, encontramos que los tres procedimientos muestran las mismas potencialidades predictivas en general, aunque en una mayor cantidad de configuraciones, la opción sin ponderar resulta un mejor predictor. La tabla 3 muestra los porcentajes correctos de predicción.

**Tabla 3:**  
**Porcentaje correctamente predicho del total de casos de la muestra que reviste tal característica**

Configuración de participación de la pareja	Alternativas de ponderación		
	Sin ponderar	Ponderación proporcional	Reducción de casos
Ambos cónyuges participan	84%	80%	79%
Ninguno de los cónyuges participa	73%	66%	54%
Sólo el cónyuge participa	9%	24%	27%
Sólo la cónyuge participa	0%	0%	0%

Fuente: Elaboración propia

Las tres alternativas presentan idéntica relación entre configuraciones que son predichas correctamente en mayor proporción que otras. De acuerdo con la Tabla 3, las predicciones con mayor bondad de ajuste se pre-

<sup>2</sup> El indicador de Bondad de ajuste surge de considerar de la siguiente manera las predicciones correctas simultáneas para ambos cónyuges:

$$\hat{p}_i = x_i' \hat{\beta}$$

$$\hat{y}_i = 1 \quad \text{si } \hat{p}_i > k; \quad k = 0,5$$

$$y_i^c = 1 \quad \text{si } \hat{y}_i = y_i$$

Luego, la proporción de predicciones correctas es:  $H = \frac{\sum_{i=1}^n y_i^c}{n}$

sentan en los casos de parejas en los que ambos cónyuges participan, y con porcentajes menores en los casos correspondientes a parejas donde ninguno participa, en ambos casos es mayor la bondad de la opción sin ponderar.

En el caso de la configuración de participación exclusiva del hombre, aplicar algún procedimiento de ponderación mejora la bondad de ajuste, aunque en los tres casos con porcentajes correctamente predichos considerablemente más bajos. Ninguna de las alternativas permite identificar correctamente a parejas donde sólo la cónyuge participa, es decir, como destaca el último renglón de la Tabla 3, no se identifican casos para los cuales la predicción del modelo coincide con la participación observada. Esta baja capacidad predictiva es esperable atendiendo a la reducida cantidad de observaciones en la muestra original en las cuales sólo la mujer participa.

#### *Efecto de diseño e ineficiencia*

En la Tabla 4 se presentan los valores obtenidos para el efecto de diseño y la ineficiencia, que resultan de contrastar ambas alternativas de ponderación con la opción sin ponderar; notar que por definición estos valores no se calculan en el caso en que la ponderación proporcional del efecto de diseño son mayores que 1 (aparecen sin valor en la Tabla 4). Cabe destacar que si bien la alternativa iii) no implica el uso de ponderadores, se comparan las varianzas de los coeficientes respectivos a fin de disponer de una medida de evaluación de la eficiencia relativa en términos de la alternativa i) consistente en no ponderar.

En las comparaciones realizadas para ambas alternativas de ponderación con respecto a la opción de no ponderar, la menor eficiencia de esta última opción se verifica en prácticamente todos los coeficientes de regresión estimados. En la alternativa de ponderación ii) las variables con coeficientes de  $ef$  mayor que 1 (es decir, resulta más eficiente no ponderar que hacerlo con este procedimiento), corresponden, tanto en el caso de la mujer como del hombre, a aquellas con fuertes diferencias entre aglomerados, situación que determina los valores de los ponderadores.

Para la mayoría de las variables se observa que la alternativa de ponderación iii) presenta valores del efecto de diseño más distantes de 1, esto muestra nuevamente que la reducción de variabilidad conduce a no detectarlas, en este caso, como significativas.

**Tabla 4**  
**Efecto de diseño e Índice de Ineficiencia correspondientes a la aplicación de ponderación a la muestra (alternativas ii) y iii)).**

Variables Explicativas del Modelo	Efecto de diseño		Ineficiencia
	Ponderación proporcional (ii)	Reducción de muestra (iii)	Ponderación proporcional
<b>Probabilidad de participación de la cónyuge</b>			
Edad de la cónyuge	0.702	0.253	-
Edad de la cónyuge al cuadrado	0.699	0.224	-
Posee hijos menores de 1 año	0.559	0.432	-
Número de hijos de 1 a 5 años	0.619	0.364	-
Número de hijos de 6 a 12 años	0.571	0.403	-
Número de hijos de 13 a 17 años	0.517	0.370	-
El cónyuge varón participa en el mercado laboral	0.589	0.289	-
La cónyuge tiene estudios primarios completos o secundario incompleto	0.874	0.359	-
La cónyuge mujer tiene estudios secundarios completos o superiores incompletos	0.706	0.331	-
La cónyuge tiene estudios superiores completos	0.595	0.348	-
La cónyuge mujer tiene más escolaridad que su cónyuge	0.744	0.451	-
Tasa de participación laboral femenina en el aglomerado	1.276	0.344	0.216
El cónyuge varón está desocupado	1.248	0.346	0.199
Estado conyugal = unido	0.621	0.439	-
Residen en el área metropolitana del país	0.976	0.390	-
Constante	0.651	0.321	-
<b>Probabilidad de participación del cónyuge</b>			
Edad del cónyuge	0.983	0.285	-
Edad del cónyuge al cuadrado	0.927	0.331	-
Número de hijos de 1 a 5 años	0.732	0.509	-
Número de hijos de 6 a 12 años	0.707	0.433	-
La cónyuge mujer participa en el mercado laboral	0.944	0.331	-
Residen en el área metropolitana del país	1.120	0.527	0.107
Constante	1.165	0.768	0.141

*Fuente: Elaboración propia*

Cuando la opción sin ponderar produce un valor de  $ef > 1$  (situaciones poco frecuentes) se evaluó el grado de ineficiencia que presentan. Tales situaciones refuerzan la necesidad de aplicar un sistema de ponderación frente a no considerarlas, ya que el mayor esfuerzo que denota la ponderación se ve compensado con estimaciones más precisas de los intervalos de confianza correspondientes.

## V. CONCLUSIONES

En la actualidad, las estadísticas oficiales, y dentro de éstas las que provienen de estudios que parten de diseños muestrales, son una de las fuentes más comúnmente empleadas en un número creciente de investigaciones económicas, demográficas y/o sociales. Por ello cobra importancia la consideración de aspectos metodológicos para analizar, y posteriormente comparar, adecuadamente los resultados derivados de aplicaciones teóricas o empíricas a las mismas (estimaciones, regresiones, entre otras).

Los diseños muestrales más habituales corresponden a esquemas multietápicos, por lo cual es importante considerar que, en general, en el análisis de los resultados la subestimación del error de muestreo puede generar intervalos de confianza erróneos. Las conclusiones derivadas de estos cálculos implicarían decisiones erróneas respecto a los estadísticos que se obtienen para contrastar la significancia estadística de variables a partir del ajuste de un modelo particular.

Evaluar comparativamente los resultados surgidos de tratamientos alternativos de los datos provenientes de las fuentes de información permite dar cuenta, no sólo de los cambios que presentan las variables estudiadas, sino principalmente identificar la metodología más conveniente de trabajo específico para cada una de las mismas.

La aplicación de un modelo de participación laboral de los cónyuges a datos de la EPH, para comparar tres alternativas de tratamiento de los mismos:

- i) no ponderar los datos
- ii) ponderar de acuerdo al peso proporcional en la muestra total
- iii) reducir la muestra de manera que resulte auto-representada

permite concluir que, sin importar la alternativa de ponderación, las variables que resultan significativas consistentemente son las relacionadas con el nivel educativo, los cambios en la estructura etaria de los niños del hogar, y la situación del cónyuge (laboral y/o de estado civil). Las variaciones en los coeficientes estimados son reducidas, y el indicador de efecto de diseño no evidencia, en general, que resulte necesaria la ponderación al tener en cuenta las varianzas estimadas en los diferentes procedimientos.

Si bien en este caso la utilización de los datos muestrales, tal como son provistos en la EPH y para modelos como el considerado, no

produce sesgos relevantes en las estimaciones, cabe señalar que hasta el momento no están suficientemente difundidos, resultados que avalen de manera general los procedimientos de ponderación vinculados con las estimaciones a partir de modelos no lineales de estimación de probabilidad para un sistema de ecuaciones con variables dependiente limitada, a partir de datos muestrales.

## VI. REFERENCIAS

- Benvin, E. y Perticará, M. (2007): "Análisis de los cambios en la participación laboral femenina en Chile". Paper provisto por Ilades-Georgetown University. *Economics* N° 180. Consultado en <http://www.economia.uahurtado.cl/pdf/publicaciones/inv180.pdf>
- Castro Martin, Teresa (1997): "Marriages without papers in Latin America". *Internacional Population Conference, IUSSP, Pekín*, Vol. 2: pp 941-960.
- Cerrutti, M. y Zenteno, R. (2000): "Cambios en el papel económico de las mujeres entre las parejas mexicanas". *Estudios Demográficos y Urbanos* 43, Vol. 15(1): pp. 65-95.
- Duguet E. y Simonnet V. (2004): "The participation of couples in the labor market: an econometric analysis". *Labor and Demography* 0411005, EconWPA. Consultado en <http://ideas.repec.org/p/wpa/wuwpla/0411005.html>.
- Esquivel Bocanegra, F., Méndez Ramírez, I. y Ramírez Valverde, G. (2003): "Algunas alternativas para estimar la matriz de covarianza de un estimador en un modelo de regresión lineal". *Revista Investigación Operacional*. Vol . 24, No. 3.
- Ettner, S. (1995): "The Impact of 'Parent Care' on Female Labor Supply Decisions" *Demography* Vol 32 (1): pp. 63-80.
- Fuller, W. (1975): "Regression analysis for sample surveys". *Sankhya C*, Vol 37: pp. 117-132
- Ganninger, M., Häder S. y Gabler, S. (2007): "Design Effects and Interviewer Effects in the European Social Survey: Where are we now and where do we want to go tomorrow?". Consultado en

[www.europeansocialsurvey.org/index.php?option=com\\_docman&task=doc\\_view&gid=197](http://www.europeansocialsurvey.org/index.php?option=com_docman&task=doc_view&gid=197)

- Greene, W (2003) *Econometric analysis*. 5th. Ed. Prentice Hall. New Jersey.
- Gong, X. y van Soest, A. (2002): "Family Structure and Female Labor Supply in México City". *The Journal of Human Resources*. Vol. 37(1): pp. 163-191.
- Guillén, M., Juncà, S., Rué, M. y Aragay, J. M. (2000): "Efecto del diseño muestral en el análisis de encuestas de diseño complejo. Aplicación a la encuesta de salud de Catalunya". *Revista española de Salud Pública*. Vol. 14(5): pp. 399-402.
- Herrero, V. (2007): "Configuraciones de participación laboral conjunta (1985-2005)". Colección de tesis en Biblioteca Centro de Estudios Avanzados, UNC.
- Hernández, P. J. y Méndez, I. (2005): "La corrección del sesgo de selección en los análisis de corte transversal de discriminación salarial por sexo: estudio comparativo en los países de la Unión Europea". *Estadística Española* Vol. 47(158): pp. 179 a 214
- Holt, D., Smith, M.F. y Winter, P.D. (1980) "Regression Analysis of Data from Complex Surveys". *Journal of the Royal Statistical Society*. Series A, Vol. 143(4): pp. 474-487.
- INDEC. (2003) La nueva Encuesta Permanente de hogares de Argentina. Consultado en [http://www.indec.mecon.ar/nuevaweb/cuadros/4/Metodologia\\_EPHContinua.pdf](http://www.indec.mecon.ar/nuevaweb/cuadros/4/Metodologia_EPHContinua.pdf)
- Kish, L. (1995). "Methods for Design Effects". *Journal of Official Statistics*, Vol. 11, N° 1, pp. 55-77.
- Kish, L. y Frankel, M. (1974): "Inference from Complex Samples," En *Journal of the Royal Statistical Society* B(36), pp: 1-37.
- Korn, E. y Graubard, B. (1995). "Analysis of Large Health Surveys: Accounting for the Sampling Design" *Journal of the Royal Statistical Society*. Series A (Statistics in Society), Vol. 158, No. 2, pp. 263-295.
- Larrañaga, O. (2006): "Participación laboral de la mujer. Chile 1958-

- 2003". En J Samuel Valenzuela, Eugenio Tironi y Timothy Scully (eds): *Familia, modernización y bienestar en Chile*. Taurus, Santiago.
- Leibowitz, A. y Klerman, J. (1995): "Explaining Changes in Married Mothers' Employment over Time." *Demography* Vol. 32(3): pp. 365-378.
- Maddala, G. (1990): *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press. New York.
- Mallar, C. (1977): "The estimation of simultaneous probability models", *Econometrica*, 45(7): 1717 – 1722.
- Medina, F. (1998): "Los errores de muestreo en las encuestas complejas: usos y abusos de la información". Reunión de la Asociación Internacional de Estadísticos de Encuestas y la Asociación Internacional de Estadísticas Oficiales, ISI, Aguascalientes, México, septiembre de 1998.
- Orcutt Duleep, H. y Sanders, S. (1993): "The Decision to Work by Married Immigrant Women" *Industrial and Labor Relations Review* Vol. 46(4): pp. 677-690.
- Pinelli, A. (2004) "Gênero e família nos países desenvolvidos". En *Gênero nos estudos de população*. Campinas, ABEP. Vol. 2: pp. 55-78.
- Pizarro Briones, M. (2002) "Introducción al análisis de muestras complejas". *10º Taller Regional del MECOVI*. Buenos Aires. Argentina.
- Quesenberry, C. y Jewell, N. (1986) "Regression Analysis Based on Stratified Samples". *Biometrika*, Vol. 73(3): pp. 605-614.
- Rodríguez de Messere, M. y Hozzowski, A. (2000) "Encuesta Permanente de Hogares de Argentina: su reformulación muestral". Documento presentado en el XXVIII Coloquio de la Sociedad Argentina de Estadística, Posadas.
- Sánchez-Carrión, J. (2000): *La bondad de la encuesta: el caso de la no respuesta*. Ed. Alianza. Madrid – España.
- Sasaki, M. (2002): "The Causal Effect of Family Structure on Labor Force Participation among Japanese Married Women". 2002. *The Journal of Human Resources* Vol. 37(2): pp. 429-440.
- Shaw, K. (1994) "The Persistence of Female Labor Supply: Empirical

Evidence and Implications." *The Journal of Human Resources*  
Vol. 29(2): pp. 348-378.

Skinner C. J. (1986): "Design Effects of Two-Stage Sampling". *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 48, No. 1. pp. 89-99.

Wainerman, C. (1979): "Educación, familia y participación económica femenina en la Argentina." *Desarrollo Económico*, Vol. 18(72): pp. 511-537.

Wong, R. y Levine, R. (1992): "The Effect of Household Structure on Women's Economic Activity and Fertility: Evidence from Recent Mothers in Urban Mexico". *Economic Development and Cultural Change*, Vol. 41(1): pp. 89-102.

## **VII. AGRADECIMIENTOS**

Agradecemos los aportes y las contribuciones del Dr. Roberto Giuliadori, en las discusiones iniciales que motivaron este trabajo.