

Emparejamiento selectivo y desigualdad en Argentina

Víctor Funes Leal¹

Tesis de Maestría

Maestría en Economía

Universidad Nacional de La Plata

Director: Leonardo Gasparini

Fecha de defensa: 19/05/2015

Códigos JEL: D31, C15, J12

¹El autor agradece los aportes del director de ésta tesis, así como los de Mariana Marchionni, Martín Cicowiez, Javier Alejo, Cecilia Velázquez, Noemí Katzkowicz, Adriana Romero, Victoria Orozco, Víctor Elías, Ana María Cerro, a los participantes de los seminarios de las Universidades Nacionales de La Plata y Tucumán, a los de la XLIX Reunión Anual de la Asociación Argentina de Economía Política y a dos evaluadores anónimos del VII Congreso Nacional de Estudiantes de Posgrado en Economía, cualquier error u omisión remanente es de exclusiva responsabilidad del autor.

Resumen

En este trabajo se busca evaluar cuál es el impacto del emparejamiento selectivo sobre dos magnitudes: la oferta laboral femenina y la desigualdad en la distribución del ingreso familiar. Con éste objetivo se utilizan tres métodos de descomposición, los de Cancian y Reed (1999), Bredemeier y Juessen (2013) y Greenwood y col. (2014) con el objeto de estudiar el impacto de una mayor selectividad en la formación de parejas sobre las horas de trabajo ofrecidas en el mercado y sobre la desigualdad medida a través del índice de Gini y el coeficiente de variación del ingreso total familiar.

Para ambos casos, se encuentra que el emparejamiento selectivo explica sólo un aparte de las variaciones, siendo el grueso de éstas atribuibles a otros factores (disminución de la fecundidad, menor brecha salarial y otros). Por otro lado, la selectividad no muestra variaciones significativas en el período estudiado, independientemente del indicador que se utilice.

Abstract

This thesis evaluates the impact of assortative mating over two very closely linked variables: female labor supply and family income inequality. To attain this objective, three methods are used, those of Cancian and Reed (1999), Bredemeier and Juessen (2013) and Greenwood et al. (2014) in order to decompose the effect of a greater selectivity in couple formation over working hours by women or family income inequality measured with the Gini index and the coefficient of variation of total family income.

In all three cases it is found that assortative mating plays a minor role in the determination of variations in both variables, because these can be attributed to other factors (lower fertility, reductions in the wage gap and others). Moreover, I show that variations in sorting are minimal in the period used, regardless of the indicator used.

Índice general

1. Aspectos generales	4
1.1. Introducción	4
1.2. Importancia del emparejamiento selectivo	5
1.3. Revisión de la literatura	6
1.4. Algunos hechos estilizados	8
2. Emparejamiento selectivo e ingresos familiares	14
2.1. Introducción	14
2.2. Modelo	14
2.3. Datos	16
2.4. Estadísticas descriptivas	17
2.5. Escenarios contrafácticos	21
2.6. Resultados	23
2.7. Conclusión	25
3. Emparejamiento selectivo en niveles educativos y desigualdad	27
3.1. Introducción	27
3.2. Método de simulación	28
3.2.1. Descripción de los grupos	29
3.2.2. Simulación	32
3.2.3. Coeficientes de regresión	33
3.2.4. Tau de Kendall	35
3.2.5. Suma de diagonales	37
3.2.6. Tablas de contingencia	37
3.3. Indicadores de desigualdad	39
3.3.1. Definición de los indicadores	39
3.4. Experimentos contrafácticos	42
3.5. Chequeos de robustez	47
3.6. Conclusión	49
4. Emparejamiento selectivo y oferta laboral femenina	51
4.1. Introducción	51
4.2. Modelo	52
4.2.1. Datos y fuentes	52
4.2.2. Análisis cuantitativo	53
4.2.3. Estimación de salario de mujeres	55

4.2.4. Imputación de salarios	56
4.2.5. Obtención de datos para el modelo	58
4.3. Simulación	61
4.3.1. Calibración	61
4.3.2. Datos de entrada	62
4.3.3. Resultados	63
4.4. Conclusión	66
4.5. Anexo: gráficos	71
4.6. Anexo: cuadros	74

Capítulo 1

Aspectos generales

1.1. Introducción

El mercado laboral en todos los países occidentales y, en Argentina particularmente, sufrió un cambio muy importante en los últimos 30 años debido al ingreso de las mujeres en el mercado laboral, este hecho tuvo un impacto muy fuerte tanto en la oferta como en la demanda de trabajo femenina y masculina, así también como en otras variables como ser la fecundidad, la distribución del ingreso intrahogar, tasas de nupcialidad y divorcios, etc. (Killingsworth y Heckman, 1986).

Dentro de la vasta literatura sobre oferta y demanda laboral femenina, existe un tema en particular que ha comenzado a atraer cada vez más atención en los últimos años, el “emparejamiento selectivo”¹, que hace referencia al hecho que la decisión de formar una pareja a largo plazo entre personas de sexo opuesto es una decisión racional que involucra una serie de factores relacionados con el grado de afinidad de ambas personas (nivel educativo, ingresos laborales, origen étnico, religión, entre otros).

Schwartz (2010) afirma que el ingreso de la mujer al mercado laboral tuvo como consecuencia un cambio radical en la naturaleza de la relación económica subyacente en los matrimonios. Hasta mediados del Siglo XX el matrimonio implicaba una relación de complementariedad entre las fuentes de ingresos, dónde el hombre ofrecía casi la totalidad de su tiempo en el mercado laboral, mientras que la mujer hacía todo lo contrario, dedicando la mayor parte de su tiempo a la producción de bienes públicos dentro del hogar. Éste esquema de división del trabajo fue cambiando a medida que mayor cantidad de mujeres se integraron al mundo laboral, haciendo que, en primer lugar, las normas sociales se volvieran más igualitarias y, en segundo lugar, se incremente la simetría entre las carac-

¹En este trabajo escogeremos esta traducción del término “assortative mating”, puesto que también puede traducirse como “apareamiento selectivo”, se escogió este término tras consultar la página web de Real Academia Española (www.rae.es), puesto que la segunda acepción refiere más a cuestiones reproductivas, mientras que la primera es más general. Véase también: http://es.wikipedia.org/wiki/Emparejamiento_selectivo

terísticas de los cónyuges, particularmente los ingresos y el nivel educativo de cada uno de los cónyuges.

1.2. Importancia del emparejamiento selectivo

Una forma útil de describir la importancia del emparejamiento selectivo a la hora de estudiar la desigualdad en los ingresos laborales es a través de una simulación de montecarlo. Para ello se simula una economía ficticia compuesta por 2000 personas, 1000 mujeres y 1000 varones, cuyos ingresos laborales se generan por medio de una distribución uniforme cuyos valores se encuentran entre entre 0 y 20000 ($U[0, 20000]$), luego se calcula el ingreso por decil de cada hombre, mujer y per cápita familiar (igual al promedio del ingreso de los respectivos cónyuges, ya que no hay hijos en este ejercicio simple) y el coeficiente de Gini de dicho ingreso per cápita familiar en base a cada uno de los 1000 individuos de la muestra simulada, luego se repite este ejercicio 1000 veces y se calcula el promedio de los ingresos medios por cada decil y los respectivos coeficientes de Gini bajo diferentes tipos de emparejamiento.

En el primer escenario, detallado en el cuadro 1.1 se simula la situación de emparejamiento aleatorio, es decir, cada persona forma una pareja con quien comparte fila en la distribución del sexo opuesto, luego se calcula el ingreso per cápita de la pareja, suponiendo que la distribución intrahogar es igualitaria, de modo que cada cónyuge se queda con la mitad de éste.

Cuadro 1.1: Patrones de emparejamiento-simulación de montecarlo

Decil	Promedios por decil del ingreso per cápita familiar		Emparejamiento		
	Salario hombre	Salario mujer	Aleatorio	Positivo	Negativo
1	1.009,55	1.009,27	2.994,89	1.009,41	9.263,34
2	3.009,68	3.009,19	5.453,70	3.009,43	9.530,20
3	5.002,30	5.004,36	7.057,66	5.003,33	9.694,74
4	6.995,74	7.004,91	8.360,88	7.000,32	9.827,00
5	8.996,03	9.001,50	9.480,47	8.998,77	9.944,32
6	10.995,39	10.999,09	10.518,56	10.997,24	10.056,74
7	12.996,51	12.999,44	11.641,37	12.997,98	10.174,02
8	14.996,36	14.999,96	12.938,57	14.998,16	10.305,72
9	16.994,03	16.997,38	14.545,29	16.995,71	10.469,94
10	18.988,67	18.995,28	17.010,92	18.991,97	10.736,29
Coeficiente de Gini-media			23,31	33,21	2,44
Coeficiente de Gini-desv. est.			0,0053	0,0053	0,0008

En el segundo caso, se plantea la existencia de emparejamiento selectivo positivo, esto es, cada persona se casa con quien le corresponde en su fila, previo ordenamiento de salarios de mayor a menor, esto equivale a que cada persona

forma una pareja con quien posea salarios iguales o superiores al propio, de modo tal que primero se casan las personas de mayores ingresos, luego las que le siguen y así hasta llegar a los de menores ingresos.

El esquema de la quinta columna del Cuadro 1.1 muestra que este tipo de emparejamiento selectivo incrementa el coeficiente de Gini del ingreso per cápita familiar en un 41 % con respecto a la situación inicial, con sólo cambiar el orden de la formación de parejas la desigualdad en la distribución del ingreso familiar se incrementó notablemente.

En el último caso, representado en la sexta columna del Cuadro 1.1 se modifica la regla de emparejamiento selectivo haciendo que ahora las mujeres más pobres se casen con los hombres más ricos, es decir, el escenario extremo opuesto al anterior con el objetivo de evaluar la variación de la desigualdad. Ahora, cambiando al regla de formación de parejas, la desigualdad medida por el coeficiente de Gini se redujo diez veces con respecto al escenario base, del mismo modo la variabilidad del coeficiente de Gini a lo largo de las 1000 muestras cae notablemente en relación a los demás casos. Se concluye que la estrategia de conformación de parejas es un factor importante para determinar la desigualdad en la distribución del ingreso.

1.3. Revisión de la literatura

El análisis económico de la familia ingresó a la corriente principal del análisis económico con los trabajos de Becker (1991)(pág 110) según el cual existe un nivel de emparejamiento óptimo, que tiene lugar cuando no es posible cambiar las combinaciones de personas de manera tal que se beneficie a alguien sin perjudicar a otro y, que el emparejamiento positivo tiene lugar en la medida que la productividad se maximice entre las diferentes combinaciones de parejas, mientras que si existen beneficios derivados de la especialización, dónde los hombres participan del mercado laboral y las mujeres dedican su tiempo a la provisión de bienes públicos en el hogar, tenderá a prevalecer el emparejamiento selectivo negativo.

Luego, Lam (1988), desarrolló un modelo teórico alternativo en el cual las ganancias del matrimonio provenían del consumo conjunto de bienes públicos, un aspecto no considerado por Becker, generando como consecuencia emparejamiento selectivo positivo.

A partir del trabajo pionero de Cancian y Reed (1999) la literatura sobre emparejamiento selectivo en economía ha utilizado como método de análisis el uso de micro-descomposiciones, esto es, el uso de distribuciones contrafácticas de las características de las poblaciones para estudiar el efecto de las reglas de formación de parejas.

A su vez, se han empleado mayoritariamente tres métodos de descomposición:

1. **Descomposiciones paramétricas:** en éste caso se utiliza como medida de la desigualdad de la distribución del ingreso al coeficiente de variación (razón entre el error estándar y la media de una variable continua), sin embargo éste es una medida inexacta de la desigualdad porque tiende a darle mayor ponderación a las transferencias entre individuos cuya diferencia de ingresos sea elevada Gasparini, Cicowiez y Sosa Escudero (2013) (pág 383), la justificación para su uso reside en que dicho indicador puede descomponerse linealmente en sus componentes, cómo se verá en la próxima sección. Una vez calculada la descomposición se plantean una serie de escenarios contrafácticos a partir de los cuales se calcula el coeficiente de variación y se mide el impacto de cada uno de ellos como la simple diferencia en los valores del estadístico, el ejemplo más reciente de su uso es el trabajo de Campos-Vázquez, Hicapié y Rojas-Valdéz (2012) que lo utiliza para estudiar la importancia del emparejamiento selectivo sobre la desigualdad de la distribución del ingreso familiar en México.
2. **Descomposiciones no paramétricas:** otra línea de investigación parte del uso del método de descomposición propuesto por DiNardo, Fortin y Lemieux (1996), éste método consiste en simular distribuciones contrafácticas en base a un conjunto de covariadas y comparar los resultados que surgen si se varía una de ellas, manteniendo al resto constantes en los valores de un año base previamente definido. Por ejemplo, Daly y Valletta (2006) descomponen una serie de indicadores² partiendo del ingreso familiar equivalente de la muestra de familias de la *Current Population Survey* de los Estados Unidos buscando evaluar el efecto de la creciente dispersión de los ingresos de los hombres sobre la distribución del ingreso familiar. Otro ejemplo es el de Eika, Mogstad y Zafar (2014), quienes comparan los resultados obtenidos para Estados Unidos con los de Noruega sobre niveles de escolaridad de individuos haciendo uso de una serie de microdatos para éste último país, dónde puede identificarse no sólo el nivel educativo sino también la profesión del individuo y su cónyuge.
3. **Descomposiciones semiparamétricas:** otro camino , más reciente, consiste en utilizar métodos combinados de micro-descomposición, un ejemplo de ellos es el trabajo de Greenwood y col. (2014) que utiliza una descomposición semiparamétrica del coeficiente de Gini utilizando un gran número de grupos, a los cuales se les calcula el ingreso y tamaño promedio y con éstos más la distribución empírica de los niveles de educación conjuntos de los cónyuges y de sus participaciones laborales se simulan escenarios contrafácticos en base a variaciones entre un año base y otro posterior, permitiéndoles variar a alguna de las mencionadas características de la población.

²Mediana, desviación estándar, coeficiente de variación, razones de deciles (90/10, 50/10, 90/50, 70/25 y 95/5), coeficiente de Gini, coeficiente de Theil, desviación logarítmica media y tasa de pobreza

1.4. Algunos hechos estilizados

El indicador de desigualdad de mayor uso, tanto en economía como en otras disciplinas sociales es el **Índice de Gini**, puesto que posee una serie de propiedades que lo hacen preferible a los demás índices (Gasparini, Cicowiez y Sosa Escudero (2013), pág 374). Éste índice se define de la siguiente manera:

$$G = -1 + 2 \int_0^{\infty} \frac{y}{\mu(F)} dF(y)$$

En la Figura 1.1 se muestra la variación del coeficiente de Gini del ingreso per cápita familiar real, aquí puede verse que existen dos períodos claramente diferenciados, por un lado la década de 1990, dónde la desigualdad creció fruto de una serie de factores (mayor brecha de retornos a la habilidad, apertura de la economía e incremento de la Población Económicamente Activa relativa a la masa de asalariados). Luego de la crisis de 2001-2002, se observa una reversión de la tendencia anterior, con una caída del índice, desde más de 0,5 a menos de 0,42 en una década; ésta reducción estuvo motivada por, en primer lugar, la recuperación económica de 2003-2008, la aparición de numerosas políticas distributivas progresivas y mayores tasas de empleo que en el decenio anterior³.

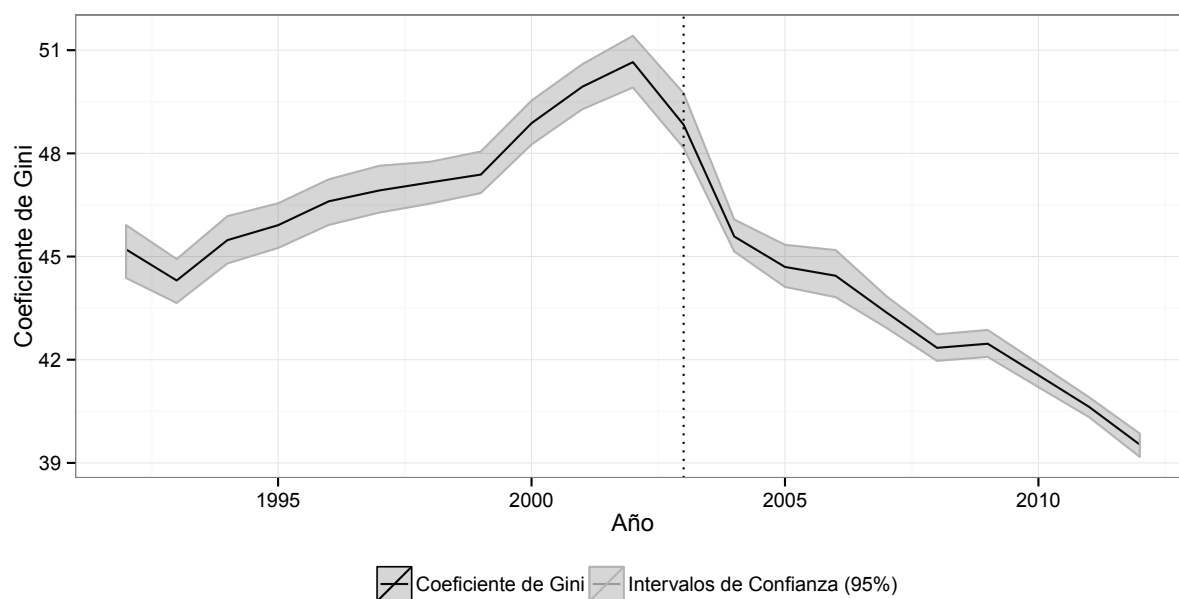


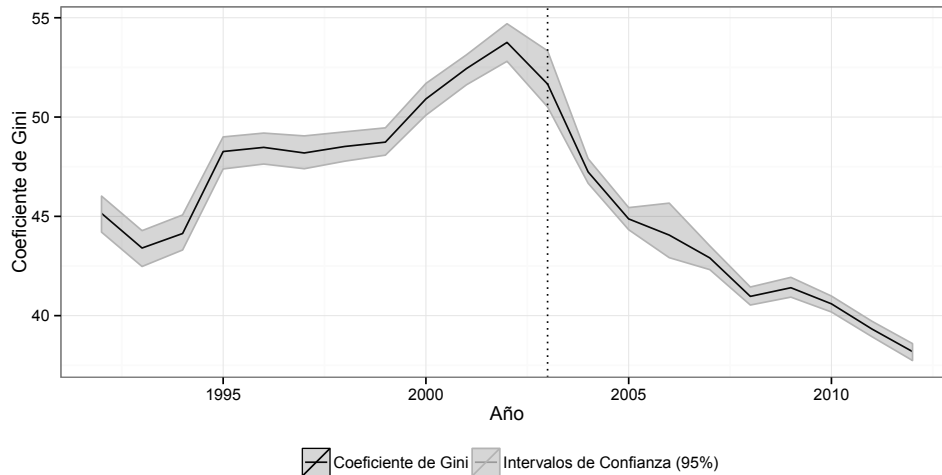
Figura 1.1: Coeficiente de Gini del Ingreso per cápita familiar real e intervalos de confianza estimados por bootstrap con 500 repeticiones

En el gráfico se incluyen también los intervalos de confianza dónde la varianza del estimador se calculó por medio de bootstrap (James y col., 2013), para resol-

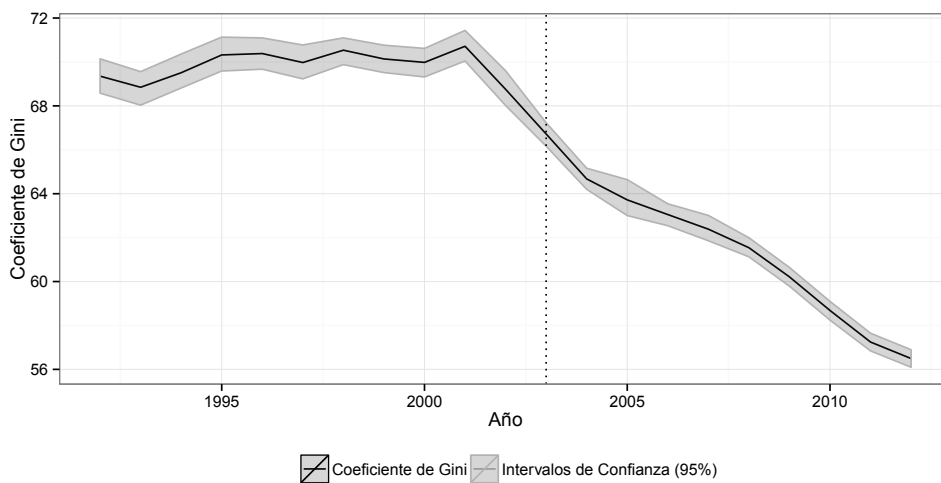
³Aunque, también tuvo lugar un crecimiento de similar magnitud en las tasas de subempleo e informalidad laboral

ver el problema que surge debido a que el coeficiente de Gini es un estimador puntual de la desigualdad y por si sólo no provee una medida de su variabilidad.

El método de bootstrap consiste en realizar un número finito (denotado como R) de muestras aleatorias de observaciones pertenecientes a la base de datos y calcular en cada una el estadístico en cuestión (en este caso el índice de Gini), una vez que se poseen estas estimaciones se calcula la media y la desviación estándar⁴, y es ésta última la que se utiliza para calcular los intervalos de confianza, en este caso del 95 %.



(a) Coeficiente de Gini del Ingreso laboral masculino real



(b) Coeficiente de Gini del Ingreso laboral femenino real

Figura 1.2: Descomposiciones del ingreso total familiar

En las Figuras 1.2(a) y 1.2(b) se muestran los coeficientes de Gini y sus respectivos intervalos de confianza calculados con las mismas especificaciones que

⁴Sea \hat{G}_i cada una de las estimaciones y $\bar{G} = \frac{1}{R} \sum_{i=1}^R \hat{G}_i$ su media, la desviación estándar del coeficiente de Gini estimada por éste método es: $SE(\hat{G}) = \sqrt{\frac{1}{R-1} \sum_{i=1}^R (\hat{G}_i - \bar{G})^2}$

en el gráfico anterior para los ingresos totales de hombres y mujeres. El ingreso total se define de la misma forma que en la base **SEDLAC**⁵, esto es, como la suma de los ingresos laborales y no laborales para todos los individuos. A su vez, los ingresos laborales son iguales a la suma de los montos percibidos por todas las ocupaciones (principal y secundaria). Los ingresos no laborales, por otro lado, son iguales a la suma de los ingresos por jubilaciones y pensiones, ingresos de capital y transferencias (seguro de desempleo, becas, cuotas de alimentos, etcétera, así como también cualquier otro ingreso cuyo origen no sea la o las ocupaciones del individuo).

La desigualdad del ingreso total masculino se comporta casi de forma idéntica a la del ingreso per cápita familiar, debido a que continúa siendo su principal componente, sin embargo la evolución de la desigualdad del ingreso total femenino es muy diferente, si bien exhibe la misma tendencia que el ingreso familiar (creciente en los 90 y decreciente en los 2000), su nivel es muy superior al de los anteriores, esto refleja la existencia de un número significativo de mujeres que deciden no participar del mercado laboral relativo a los hombres (y también, aunque en menor medida, una tasa de desempleo superior relativa a ellos).

Otra característica notoria de la desigualdad del ingreso total femenino es que su reducción en la década de 2000 es más pronunciada que la de las demás, reflejando la creciente participación laboral de éstas. Finalmente, cabe señalar que el punto de inflexión en la desigualdad de éstas coincide con el año 2003, cuando se modificó la metodología de muestreo de la Encuesta Permanente de Hogares, lo que lleva a preguntarse en qué medida el cambio metodológico afecta a ésta caída, ya que tanto el ingreso familiar como el ingreso laboral masculino comenzaron a reducirse antes del cambio metodológico⁶.

Otra forma de mostrar la evolución de la desigualdad es a través de una estimación no paramétrica de la distribución del ingreso por medio del método de “kernels”, éstos son histogramas suavizados de las variables de ingreso, en este caso, dado que el ingreso ya sea individual o familiar, se caracteriza por tener una distribución altamente asimétrica debido a que existe un gran número de familias con ingresos muy bajos o cercanos a cero y un pequeño número de éstas que poseen ingresos muy elevados. Para resolver éste inconveniente se utiliza una transformación logarítmica, por ser una transformación monótona creciente y, gracias a ello, preserva el orden de cada observación.

El método de estimadores kernels de Rosenblatt y Parzen (Pagan y Ullah, 1999) busca aproximar la distribución del ingreso, cualquiera fuese la definición de éste, $f(y)$ a través de la siguiente función:

$$\bar{f}(y_0) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} K \left[\frac{y_i - y_0}{h} \right]$$

⁵<http://sedlac.econo.unlp.edu.ar>

⁶Los valores de 2003 se calcularon utilizando la EPH “puntual” para el primer semestre y la “continua” para el segundo, de manera tal que el valor de las variables para éste año es una suerte de híbrido entre ambos tipos de encuestas.

La función $K(\cdot)$ es el denominado “kernel” y mide que tan cerca está cualquier observación y_i de y_0 , existen numerosas opciones para estimar la función kernel, en los estudios de distribución del ingreso se utilizan mayoritariamente dos tipos de kernels, el normal y el Epanechnikov (Cowell y Flachaire, 2015):

$$K\left(\frac{y_i - y_0}{h}\right) = \begin{cases} \frac{3}{4} \left[1 - \left(\frac{y_i - y_0}{h}\right)^2 \right] & \text{si } \left| \frac{y_i - y_0}{h} \right| < 1 \\ 0 & \text{si } \left| \frac{y_i - y_0}{h} \right| \geq 1 \end{cases}$$

Para estimar las funciones de distribución se utilizará el último kernel, además de ésto, es necesario determinar el parámetro h , denominado “ancho de banda”, según Gasparini, Cicowicz y Sosa Escudero (2013) la elección de éste tiene más importancia que la función de kernel misma. Aquí se utilizó el método de Scott, según el cual el ancho de banda óptimo (h^*) es igual a:

$$h^* = 1,06\sigma n^{-1/5}$$

Dónde n es el tamaño de la muestra y σ es un estimador de la desviación estándar de $\log(y)$. En la figura 1.3 se muestra la distribución del logaritmo del ingreso total familiar real (a precios constantes de 2012) corregido por adulto equivalente en los dos años extremos de la muestra (1992 y 2012) y un año intermedio (2002), elegido especialmente por ser el año en el cual la crisis de 2001-2002 tuvo su apogeo.

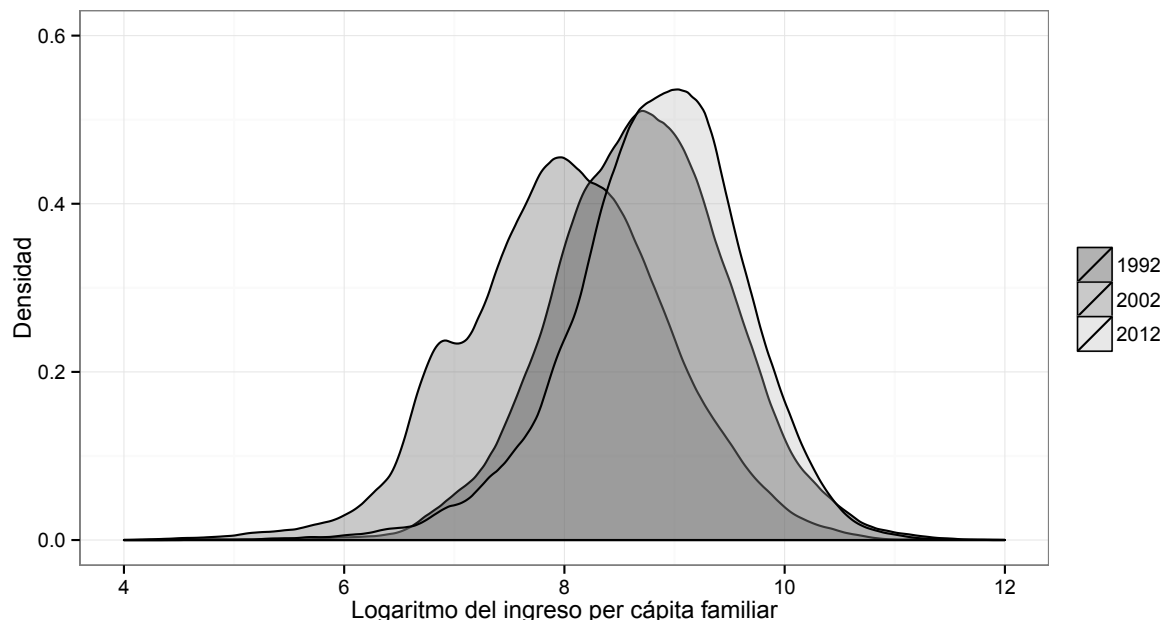
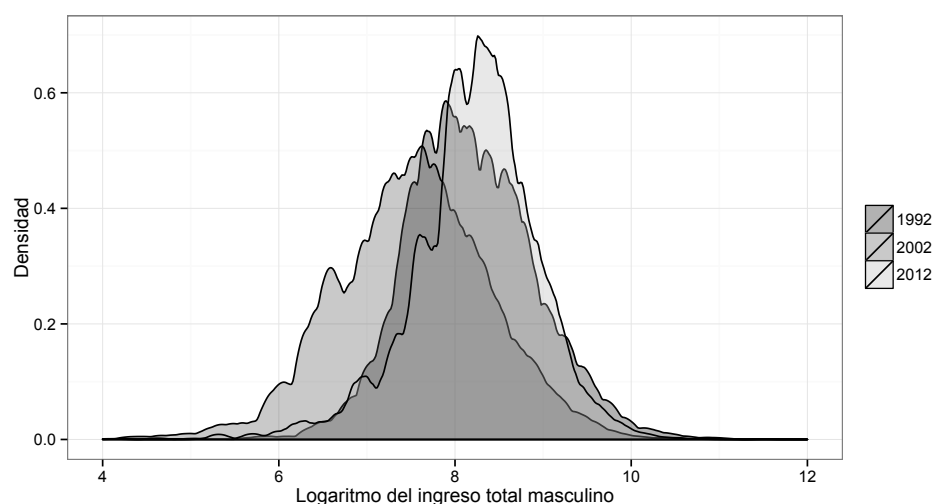


Figura 1.3: Distribución del logaritmo del ingreso total familiar real

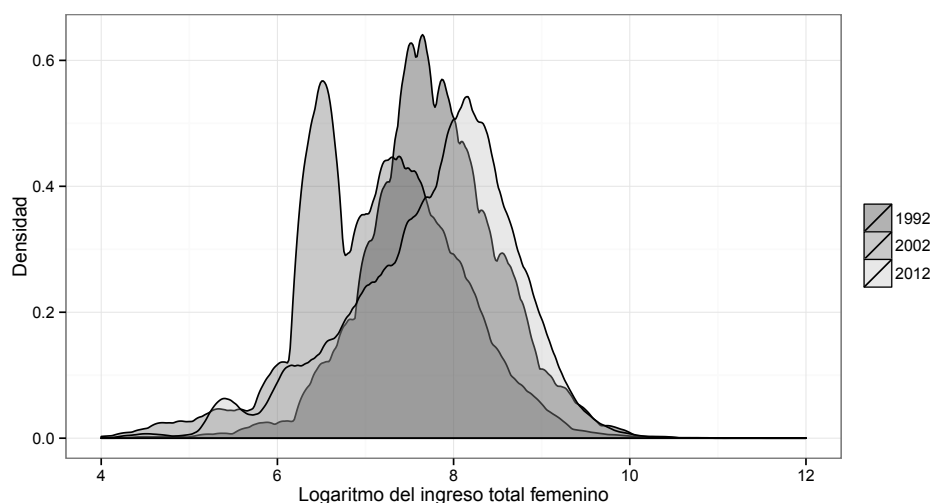
Las distribuciones superpuestas muestran, en primer lugar, el efecto de la

crisis de 2001-2002, ya que los ingreso medio y mediano se redujeron considerablemente con relación a 1992, mientras que se recuperaron en la década posterior. La escala logarítmica impuesta por la transformación no permite apreciar con claridad el hecho que el ingreso total familiar promedio cayó de \$ 7806,78 en 1992 a \$ 4208,77 en 2002, para volver a crecer luego y llegar a los \$ 9062,95 en 2012, esto es, se duplicó con relación a su valor en el punto más bajo de la crisis, mientras que creció un 16 % en relación a su valor de 1992.

En la Figura 1.4 se muestran las distribuciones de los ingresos totales masculinos y femeninos de manera similar al ingreso total utilizando la misma función de kernel y ancho de banda óptimo.



(a) Distribución del logaritmo del ingreso total masculino



(b) Distribución del logaritmo del ingreso total femenino

Figura 1.4: Distribuciones del ingreso total por género a precios constantes de 2012

En las Figuras 1.4(a) y 1.4(b) se muestran las distribuciones de los ingresos totales de hombres y mujeres para los mismos años que el gráfico anterior, aquí puede verse con claridad lo siguiente:

- Ambas distribuciones no difieren en gran medida en su forma, en gran parte debido a que la transformación logarítmica elimina a las personas con ingresos cero, esto es, reduce considerablemente la brecha salarial.
- La crisis de 2002 desencadenó una serie de medidas de protección social que tuvieron un mayor impacto en la población femenina, lo que se evidencia en la distribución bimodal que tuvieron sus ingresos totales en dicho año.
- Si bien la brecha salarial se ha reducido notoriamente, la distribución de ingresos femeninos continúa siendo más asimétrica que la de los masculinos, puesto que acumula mayor masa en los valores inferiores a su moda.

Capítulo 2

Emparejamiento selectivo e ingresos familiares

2.1. Introducción

En el capítulo previo se mencionó que uno de los primeros intentos de medir el impacto del emparejamiento selectivo sobre la desigualdad en la distribución del ingreso familiar fue el trabajo de Cancian y Reed (1999). El objetivo de la presente sección consiste en estudiar cual es el papel que cumplen los ingresos laborales femeninos en la distribución del ingreso total familiar, para el análisis se dividirá la muestra, proveniente de las familias encuestadas en la **Encuesta Permanente de Hogares** durante el período comprendido entre 1992 y 2012.

Con éste objetivo se divide a la muestra en dos grupos: hogares compuestos por parejas casadas y hogares compuestos por personas solteras, separadas o divorciadas, luego se descompone el **Coefficiente de variación** del ingreso total familiar (corregido por adulto equivalente) en las contribuciones de cada elemento que lo compone: ingreso laboral de cada cónyuge, ingreso residual (no imputable al trabajo de ninguno de los anteriores) y a los coeficientes de correlación entre cada uno de ellos.

2.2. Modelo

El coeficiente de variación se utiliza como medida de desigualdad a fines de utilizar su propiedad más importante, el poder ser descompuesto linealmente en cada uno de sus elementos: si el ingreso familiar total del hogar h es Y_{ht} , entonces éste puede definirse de la siguiente forma: $Y_{ht} = Y_{hm} + Y_{hf} + Y_{hr}$, donde Y_{hm} es el ingreso laboral de los hombres, Y_{hf} es el de las mujeres y Y_{hr} es el ingreso residual, entonces:

$$CV_A^2 = S_m^2 CV_m^2 + S_f^2 CV_f^2 + S_r^2 CV_r^2 + 2\rho_{mf} S_m S_f CV_m CV_f + 2\rho_{mr} S_m S_r CV_m CV_r + 2\rho_{fr} S_r S_f CV_f CV_r \quad (2.1)$$

Dónde:

S_i : Participación de la fuente i en el ingreso familiar total.

C_i : Coeficiente de variación de la fuente i ($CV_i = \sigma_i/\mu_i$)

ρ_{ij} : Coeficiente de correlación entre las fuentes i y j

μ_i : Ingreso medio de la fuente i , con i igual a t (ingreso familiar), f (ingreso de la mujer), m (ingreso del marido) o r (ingreso de otras fuentes).

Ésta ecuación es válida sólo para el subconjunto de las parejas casadas, para estudiar el impacto de los contrafácticos sobre la totalidad del ingreso, es necesario realizar una descomposición adicional que las incluya. Entonces, si el subíndice A denota a las parejas casadas y B al resto, entonces el coeficiente de variación total (CV) puede descomponerse como:

$$CV^2 = \mu_A \left(\frac{\bar{Y}_A}{\bar{Y}} \right) CV_A^2 + \mu_B \left(\frac{\bar{Y}_B}{\bar{Y}} \right) CV_B^2 + \frac{\left[\mu_A \left(\frac{\bar{Y}_A}{\bar{Y}} \right) + \mu_B \left(\frac{\bar{Y}_B}{\bar{Y}} \right) \right]}{\bar{Y}} \quad (2.2)$$

Los autores utilizan como medida de la desigualdad al coeficiente de variación, a pesar de ser una medida inexacta de la desigualdad porque tiende a darle mayor ponderación a las transferencias entre individuos cuya diferencia de ingresos sea elevada (Gasparini, Cicowiez y Sosa Escudero, 2013) (pág 383).

Una vez efectuada la descomposición, se realizan cuatro experimentos contrafácticos para evaluar los efectos sobre la distribución del ingreso familiar, cada uno de ellos consiste en fijar los valores de los parámetros en los del año base (1992), vale decir, suponiendo que no hubo ninguna variación en éstos, mientras que se les permite variar a algunos parámetros, pero sólo uno a la vez:

1. El primer escenario muestra que le hubiese sucedido a la desigualdad del ingreso familiar si todas las variables en la ecuación 2.1 se mantuvieran constantes en sus valores del año inicial (1992), excepto la desigualdad del ingreso laboral de los hombres (CV_m^2). Con éste contrafáctico se obtiene la contribución de los hombres casados a la desigualdad del ingreso familiar total.
2. El segundo escenario es igual al primer, pero agregando la variación en la desigualdad del ingreso laboral femenino, ésto implica permitirle variar a S_f^2 o a CV_f^2 , si se le permite variar al primero y se deja al segundo constante en sus valores de 1992, implica simular que habría sucedido si las mismas

mujeres que trabajaban en el año base percibieron un porcentaje creciente del ingreso familiar. En el segundo caso, cuando S_f^2 se mantiene constante en su valor del año inicial, mientras que varía CV_f^2 , se obtiene el efecto de la estructura salarial femenina sobre la desigualdad del ingreso familiar, el impacto de la desigualdad del ingreso femenino es igual a la combinación de ambos elementos, por lo que se los estudiará en conjunto.

3. Por último, el efecto del emparejamiento selectivo en este modelo se obtiene vía la correlación entre los ingresos laborales de ambos cónyuges (ρ_{mf}). Si éste es un elemento decisivo para determinar la desigualdad del ingreso total de las parejas casadas, debería observarse una diferencia notable en la desigualdad con este contrafáctico vis-à-vis la desigualdad observada, si el emparejamiento selectivo es mayor en 2012 que en 1992, entonces la desigualdad debería ser menor cuando se mantiene fijo este parámetro que la observada.
4. Finalmente, se estudia el impacto de la tasa de nupcialidad sobre la desigualdad del ingreso total con un contrafáctico que se basa en mantener constante la proporción de parejas casadas en sus valores de 1992 en la ecuación 2.2.

2.3. Datos

Los datos para realizar la descomposición provienen de la Encuesta Permanente de Hogares (EPH), en la cual se seleccionó a los hombres y mujeres mayores de 25 años y menores de 65, se optó por escoger esos años debido que a partir de 1992 se incorporaron la mayoría de los aglomerados urbanos que componen actualmente a dicha encuesta, en segundo lugar existe una serie de problemas relacionados con la calidad de la medición de los datos de ingresos durante el episodio hiperinflacionario de 1989-1990 que generan problemas de comparabilidad de los datos entre ambas décadas.

Para sortear los problemas mencionados pueden usarse dos estrategias, una consiste en acotar la cobertura geográfica del análisis usando sólo los datos de Capital Federal (Ciudad Autónoma de Buenos Aires a partir de 1997) y el Gran Buenos Aires, mientras que la segunda, consiste en acotar el período de tiempo, siendo ésta la adoptada aquí, el motivo reside en que no queda claro cual sería la contribución de usar un período de tiempo mayor si los datos poseen el citado problema de medición de variables de ingresos.

En esta sección se define a los ingresos totales masculinos y femeninos como la suma de los ingresos laborales y no laborales atribuibles a cada cónyuge. Los ingresos laborales se definen como la suma de los ingresos provenientes de la ocupación principal y los de la ocupación secundaria, de existir éstos. Los ingresos no laborales, por su parte, se definen como la suma de ingresos provenientes de: jubilaciones y pensiones, rentas de capital (intereses o rentas por

plazos fijos, alquileres de inmuebles o acciones de sociedades anónimas), transferencias monetarias y en especie (seguros de desempleo, becas estudiantiles, ayudas sociales, etcétera) y por otros ingresos no laborales no especificados.

La comparación del ingreso total entre familias se realizó por medio de escalas de equivalencia que permiten corregir por la cantidad de hijos en cada hogar, para ello se utilizó la escala de equivalencia elaborada por el INDEC y cuyos detalles se mencionan en la sección “Emparejamiento selectivo y desigualdad”.

Para calcular el ingreso familiar total equivalente, debe primero calcularse el tamaño del hogar equivalente en función del número de adultos (definidos como personas mayores a 18 años de edad), niños (personas menores a 10 años de edad) o adolescentes y su género. Si cada adulto equivalente j en el hogar h se denota como a_{jh} , entonces el número total de adultos equivalentes será de:

$$aeq = \sum_{j \in h} a_{jh} \quad (2.3)$$

De manera tal que, si Y_h es el ingreso total familiar, entonces el ingreso familiar por adulto equivalente del hogar h (y_{ih}) será igual a:

$$y_{ih} = \frac{Y_h}{\sum_{j \in h} a_{jh}} \quad \forall i \in h \quad (2.4)$$

El coeficiente de variación utilizado como medida de desigualdad posee un serio problema: no es robusto a la presencia de “outliers”, esto implica que la presencia de al menos una persona con ingresos extraordinariamente altos posee un efecto notable en la desigualdad, a fines de evitar éstos inconvenientes, en primer lugar, se expresa a todo los ingresos en pesos constantes de 2012 y, luego, se descartan los valores de ingresos totales masculinos y femeninos que se ubiquen por encima del percentil 95 de su distribución, al igual que con los ingresos residuales.

2.4. Estadísticas descriptivas

A pesar de contar con datos para una cantidad de períodos relativamente acotada, se pueden observar algunas tendencias relacionadas con la mayor participación de las mujeres en el mercado laboral. En primer lugar, en la Figura 2.1 se muestra que los salarios (a precios constantes) de las mujeres evolucionaron de manera casi idéntica a los de los hombres en el período bajo análisis, mientras que su nivel se mantuvo siempre por debajo del de los hombres, sin una tendencia clara a reducirse o aumentar, aunque la brecha salarial entre los dos grupos parecería ser levemente anticíclica puesto que se reduce en las recesiones y se expande en los períodos de crecimiento.

En todas las figuras a continuación, se agrega una línea vertical coincidente con el año 2003, motivada en el hecho que en el mencionado año tuvo lugar una importante modificación en la Encuesta Permanente de Hogares, modificándose la periodicidad de la realización de ésta (semestral a trimestral) y la metodología de muestreo (de allí que la EPH pasó a denominarse “continua”), estos cambios representan un quiebre metodológico exógeno que puede estar reflejado en los datos, motivo por el cual se señala su ubicación en los gráficos para evitar interpretaciones erróneas.

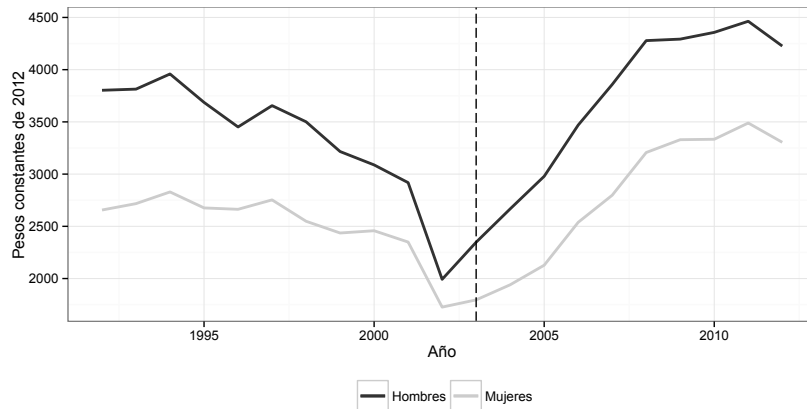


Figura 2.1: Ingresos laborales promedio por género y por año

Por otro lado, la tasa de participación en el mercado laboral por género muestra una evolución diferente, aquí si puede apreciarse una creciente participación de las mujeres en el mercado laboral, a pesar de ser todavía muy inferior a la de los hombres, creció de forma sostenida en el período, salvo por una leve reducción durante la crisis de 2002. Aquí también se observa que la tendencia de ambas series es idéntica, difiriendo sólo entre niveles según el respectivo género, por otro lado, la curva de participación laboral de mujeres casadas parecería llegar a un “tope” en el 50 % de participación.

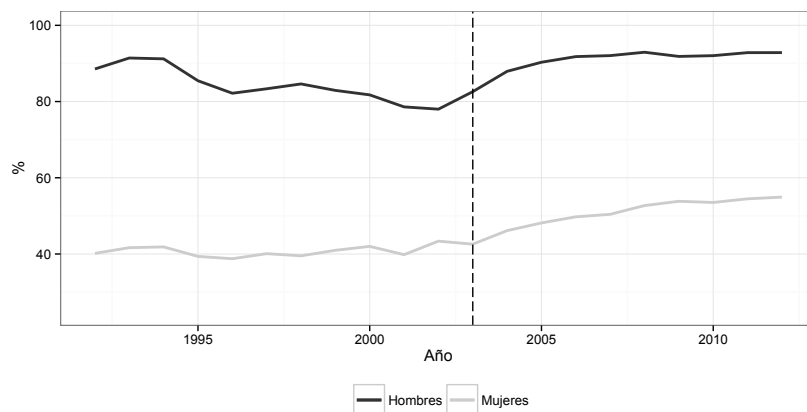


Figura 2.2: Participación laboral por género y por año

Las Figuras 2.3(a) and 2.3(b), replican al descomposición del Coeficiente de variación del ingreso total familiar desarrollada por Cancian y Reed (1999), definido como la suma de los ingresos laborales de cada uno de los cónyuges más un ingreso “residual” no atribuible a ninguna de las fuentes anteriores¹.

En la Figura 2.3(a) se muestra la evolución de los coeficientes de variación de cada uno de los componentes del ingreso familiar, aquí se observa que la desigualdad, medida por medio de éste estadístico, evoluciona de manera casi idéntica a la del ingreso laboral del marido, mostrando un crecimiento durante la década del 90, llegando a un máximo durante la crisis de 2001-2002, para luego reducirse durante la década siguiente.

El coeficiente de variación del ingreso laboral femenino se comporta de forma similar, pero mostrando una mayor desigualdad (igual a poco menos del doble que el de sus maridos), pero mostrando también una tendencia decreciente, no ocurre así con la desigualdad del ingreso residual, cuyo comportamiento es mucho más errático.

La contribución a la desigualdad total de cada componente depende de la participación de la media de cada uno sobre la media del total: $S_K = \mu_K / (\mu_H + \mu_W + \mu_R)$, en la Figura 2.3(b) se evidencia que el ingreso laboral de los hombres es el principal componente del ingreso familiar, si bien muestra una leve reducción en la última década, no aparenta aún perder peso relativo al de la mujer.

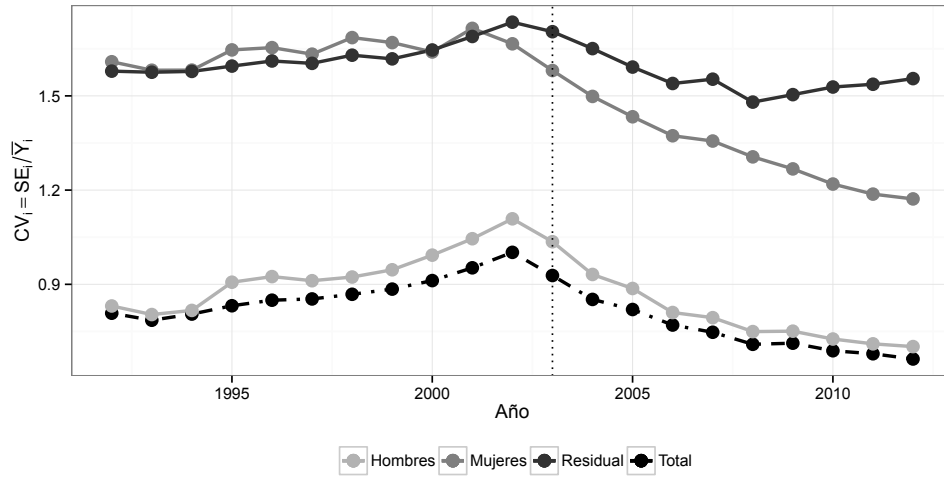
Los componentes restantes son los coeficientes de correlación entre los niveles de ingresos de los cónyuges en una pareja, y entre cada uno de éstos y el ingreso residual. Claramente, cuanto más alta (y positiva) sea la correlación entre los ingresos de los cónyuges, mayor será el grado de selectividad en la formación de parejas, esto es, mayor será el emparejamiento selectivo positivo, casi contrario, será negativo.

La Figura 2.3(c) muestra que el coeficiente de correlación no ha tenido grandes variaciones en el período, a pesar de tener una leve tendencia a crecer durante la primera mitad de la década de 1990, lo que, en un principio sería evidencia a favor de selectividad positiva².

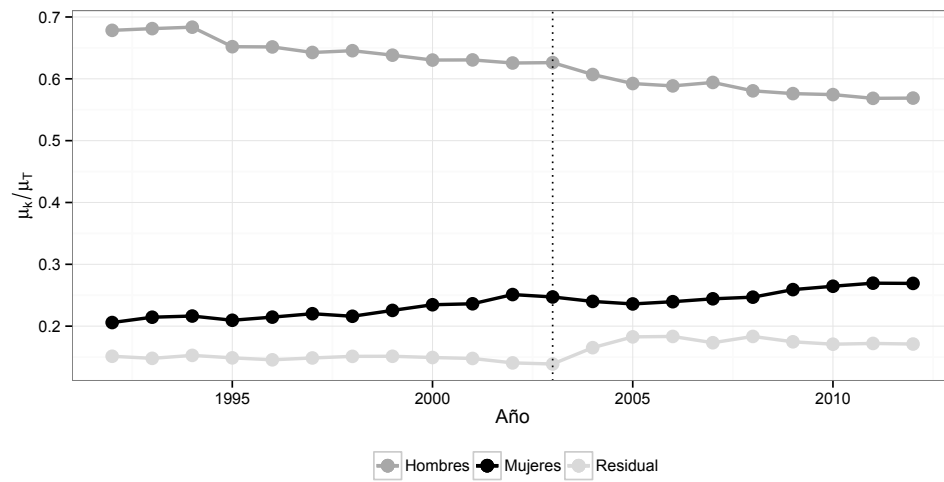
Una mayor correlación positiva implica mayor desigualdad medida por el coeficiente de variación, lo mismo ocurre con las correlaciones entre los otros dos componentes del ingreso, cabe mencionar que la magnitud de las correlaciones es relativamente baja para cualquiera de los tres casos, por lo que su contribución a la desigualdad total es bastante acotada (menor al 10%).

¹Todas las magnitudes fueron calculadas a precios constantes de 2012

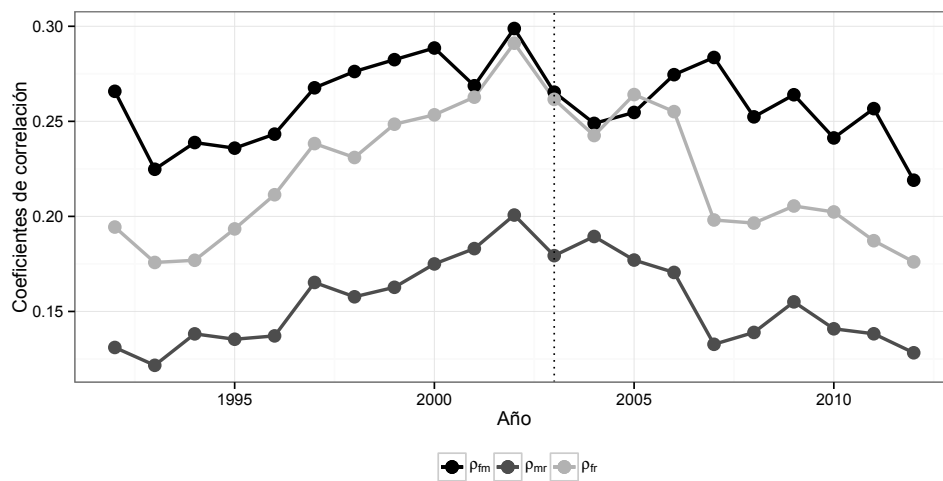
²Más adelante se verá que éste fenómeno es una posible causa del menor ajuste del método a los datos de Argentina que a los de Estados Unidos.



(a) Coeficiente de variación



(b) Participación en el ingreso total familiar



(c) Coeficiente de correlación entre los ingresos de los cónyuges y el ingreso residual

Figura 2.3: Descomposiciones del ingreso total familiar

2.5. Escenarios contrafácticos

Dada la descomposición detallada en la sección anterior, es posible crear un gran número de escenarios contrafácticos en función a cuales componentes se mantengan constantes y cuáles se les permita variar. Al igual que Cancian y Reed (1999), en cada escenario se mantendrá constantes a todos los componentes de la ecuación 2.1, salvo a los que se detallaron en la sección previa. Una característica del método de descomposición utilizada es que el resultado final dependerá del orden en el que se utilicen las variables para la descomposición, esta propiedad se denomina “path-dependence” o “dependencia del sendero” Fortin, Lemieux y Firpo (2011), ya que el resultado final depende de la variable de resultado (el ingreso familiar equivalente en este caso) condicional a cada una de las variables utilizadas para descomponerlo.

A modo de ejemplo, sea y la variable de resultado y x_1, x_2, x_3 un conjunto de características observables de la población, entonces, los métodos de descomposición buscan responder la pregunta “¿qué sucedería con la distribución de y condicional a x_j $j = 1, 2, 3$ si alguno de estos se modifica?”, en general no hay un orden pre especificado para descomponer el resultado por cada una de las características, por lo que sería posible utilizar las variables en cualquier orden, pero el resultado no será el mismo debido a que las distribuciones condicionales son diferentes.

Si, en primer lugar se descompone las variaciones de y en cada una de las características observables en el orden x_1, x_2, x_3 , la descomposición será diferente que si se utiliza otro orden, por ejemplo, el inverso x_3, x_2, x_1 :

$$f(y, x_1, x_2, x_3) = f(y/x_1, x_2, x_3) \times g_1(x_1/x_2, x_3) \times g_2(x_2/x_3) \times g_3(x_3) \quad (2.5)$$

$$f(y, x_1, x_2, x_3) = f(y/x_1, x_2, x_3) \times g_{1'}(x_3/x_1, x_2) \times g_{2'}(x_2/x_1) \times g_{3'}(x_1) \quad (2.6)$$

En ambos se descompone la misma distribución condicional, pero por diferentes “rutas”, pero $g_i \neq g_{i'}$, con $i = 1, 2, 3$, las distribuciones condicionales son diferentes, por lo que los resultados de las descomposiciones serán distintos, a pesar que la distribución original es la misma. Nótese además que el número posible de formas de descomponer un resultado es función de la cantidad de características observadas que se utilicen (en este caso son 3, por lo que la cantidad total será de $3! = 6$, en la sección posterior se mostrará como se encara ésta problemática en la micro-descomposición utilizada.

Ésta dependencia del sendero tiene un impacto clave para cualquier estadístico calculado en base a escenarios contrafácticos, continuando con el ejemplo anterior, sean $t = 1$ el período inicial y $t = 2$ el final, además, sea G el estadístico calculado en base a la distribución de y condicional a sus características, entonces, su valor calculado en base a la distribución observada de y en ambos períodos será:

$$G[f(y_1, x_{11}, x_{21}, x_{31})]$$

$$G[f(y_2, x_{12}, x_{22}, x_{32})]$$

Si lo que se busca medir es el impacto de la característica x_3 , manteniéndola constante en su valor de $t = 1$, pero con los valores de x_1 y x_2 correspondientes a $t = 2$, ésto puede calcularse de dos formas diferentes, siendo los valores obtenidos para cada una también diferentes, si G^* es el valor del estadístico en éstos escenarios contrafácticos, se obtendrá entonces:

$$G_1^*[f(y_2, x_{12}, x_{22}, x_{31})] = G^*[f(y/x_{12}, x_{22}, x_{32})g_1(x_{12}/x_{22}, x_{32})g_2(x_{22}/x_{32})g_3(x_{31})]$$

$$G_2^*[f(y_2, x_{12}, x_{22}, x_{31})] = G^*[f(y/x_{12}, x_{22}, x_{31})g_{1'}(x_{22}/x_{12}, x_{32})g_{2'}(x_{12}/x_{32})g_{3'}(x_{31})]$$

Claramente $G_1^* \neq G_2^*$, por lo que el resultado de la microdescomposición depende del sendero, es decir, del orden en el cual se descompongan las variables.

En este caso se utilizará en primer lugar el ordenamiento sugerido por Campos-Vázquez, Hicapié y Rojas-Valdéz (2012):

Escenario 1 : en el primer escenario se analiza que hubiera sucedido con la desigualdad del ingreso familiar si todas las variables de la ecuación 2.1 se hubiesen mantenido constantes en sus valores de 1992, lo que implica fijar los valores de todos los parámetros en los del mencionado año, pero permitiendo variaciones en CV_m^2 . Con este contrafáctico se obtiene la contribución de la desigualdad de los hombres casados al total.

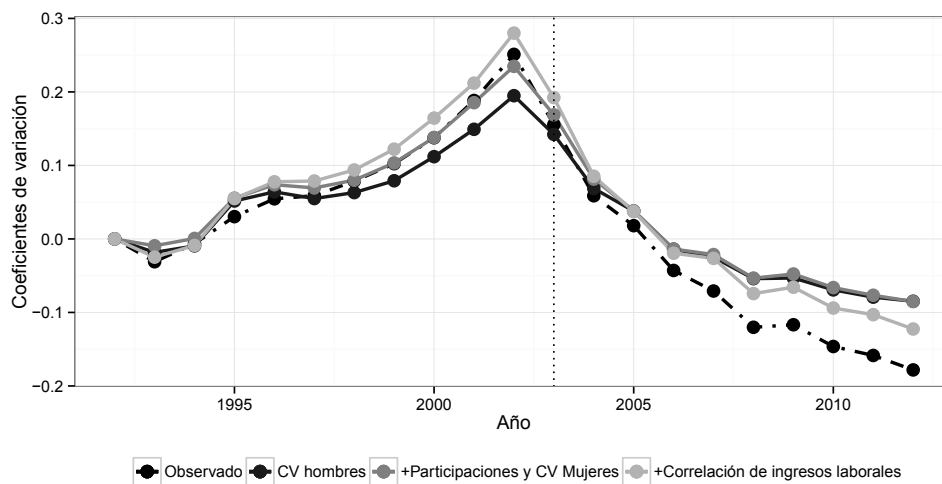
Escenario 2 : al igual que en el caso anterior cabe hacerse la misma pregunta, pero para el caso de las mujeres casadas, siguiendo a los autores, esto puede hacerse por dos vías: a través de variaciones en S_f^2 o en CV_f^2 , por un lado es posible mantener todas las variables constantes y permitirle variar sólo a S_f^2 , lo que permitiría obtener el efecto de un mayor salario para el mismo tipo de mujeres que estaban trabajando en 1992, pero no un incremento en la oferta laboral de éstas. Por otro lado, si se le permite variar a CV_f^2 , se estaría calculando el efecto de la estructura salarial femenina en la desigualdad de la distribución del ingreso familiar. Debido a que los cambios en la participación laboral femenina afectan a ambos, se calcula la contribución de éstos dos en conjunto.

Escenario 3 : el emparejamiento selectivo en este modelo opera a través de variaciones en $\rho_{m,f}$, si la selectividad es más importante como determinante de la desigualdad del ingreso familiar, debería haber una importante diferencia entre los valores obtenidos con éste versus los observados en la realidad.

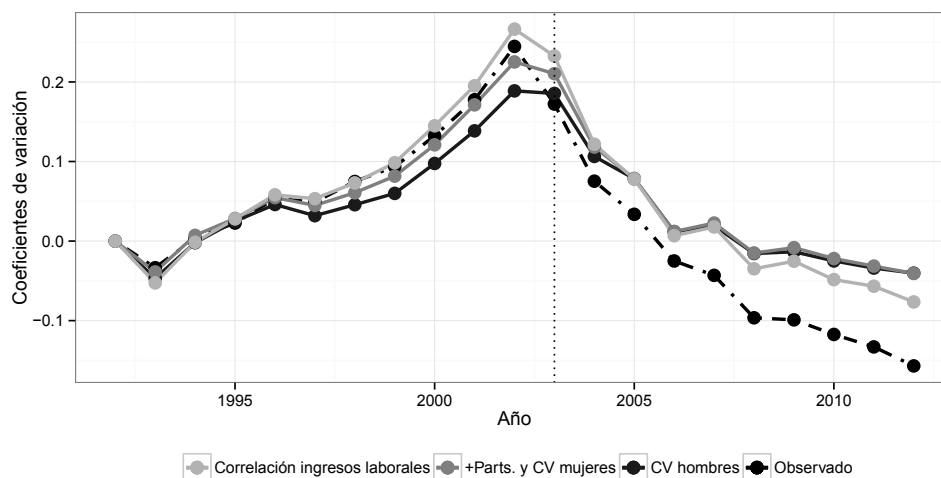
Escenario 4 : el último escenario centra su atención en el total de hogares de la economía, no sólo en los casados buscando responder la pregunta: ¿qué hubiese sucedido con la desigualdad si la tasa de nupcialidad se hubiese mantenido constante en su valor de 1992? Para responder esta pregunta se utiliza la ecuación 2.2, dónde se mantiene constante al parámetro μ_A .

2.6. Resultados

En la Figura 2.4 se muestran los resultados de realizar la descomposición del coeficiente de variación del ingreso familiar equivalente para las parejas casadas (Figura 2.4(a)) y para el total de parejas (Figura 2.4(b)) utilizando el orden de descomposiciones descrito en la sección anterior.



(a) Parejas casadas



(b) Todos los hogares

Figura 2.4: Distribuciones contrafácticas por tipo de familia (1992-2012)

Las figuras muestran que los efectos de las descomposiciones son heterogéneos relativos al coeficiente de variación observado tanto para las parejas casadas como para el total de la población. En la figura 2.4(a) se muestra el coeficiente de variación del ingreso total familiar observado (línea discontinua) y los coeficientes para cada año estimados en cada escenario. La desigualdad del ingreso total masculino, medida por medio de su coeficiente de variación, es el principal componente de la desigualdad del ingreso familiar, además de ser el principal componente de éste (ya que el valor de S_m es superior al de S_f y S_r en todos los años de la muestra).

En el segundo escenario, a ésta variación se le adicional el efecto conjunto de S_f y CV_f , la desigualdad se aproxima casi en su totalidad a la observada, pero la contribución conjunta de éstos es secundaria, en comparación a la desigualdad de los ingresos masculinos.

Para el último escenario se agrega a los anteriores la variación de ρ_{mf} , la correlación entre ingresos de los cónyuges y la medida de emparejamiento selectivo utilizada aquí, el efecto que posee es incluso menor que el de la desigualdad femenina, si bien incrementa la desigualdad levemente, casi todo su impacto es neutralizado por su baja ponderación relativa a las demás fuentes de la desigualdad. Cabe mencionar que a partir del año 2005 todas las simulaciones pronostican una caída de la desigualdad inferior a la observada en la realidad, esto se debe a que las ponderaciones de todos los componentes (salvo la de los ingresos femeninos) se mantienen constantes en sus valores de 1992, tanto la participación de los ingresos masculinos sobre el total como los coeficientes de correlación entre los ingresos de ambos cónyuges y el residuo cayeron en el último quinquenio de la década pasada en relación a sus valores en los años anteriores.

La figura 2.4(b) muestra las mismas descomposiciones, pero utilizando la ecuación 2.2 para analizar sus efectos sobre la distribución del ingreso familiar para la totalidad de individuos, pero fijando la tasa de nupcialidad en su valor de 1992. Allí puede verse que los resultados son muy similares a los observados para las parejas casadas.

En la sección anterior se mencionó que la principal debilidad del método de descomposición es que el orden en el que se toman los componentes determina el resultado final de la descomposición, para sortear este inconveniente se implementó la solución descrita en Campos-Vázquez, Hicapié y Rojas-Valdéz (2012). Ésta solución consiste en calcular las $6 \times 3 = 18$ combinaciones posibles de ordenes de descomposición, y calcular la contribución promedio de cada uno de los contrafácticos en los seis escenarios en los que aparece y graficar su variación relativa al año base, el resultado de ésta puede verse en la figura 2.5:

Aquí se observa que el tomar promedios de cada uno de los componentes equipara las contribuciones de éstos a la variación total, manteniéndose la conclusión previa que el principal elemento que explica la desigualdad del ingreso familiar es la desigualdad de ingresos masculinos, mientras que la de los ingresos

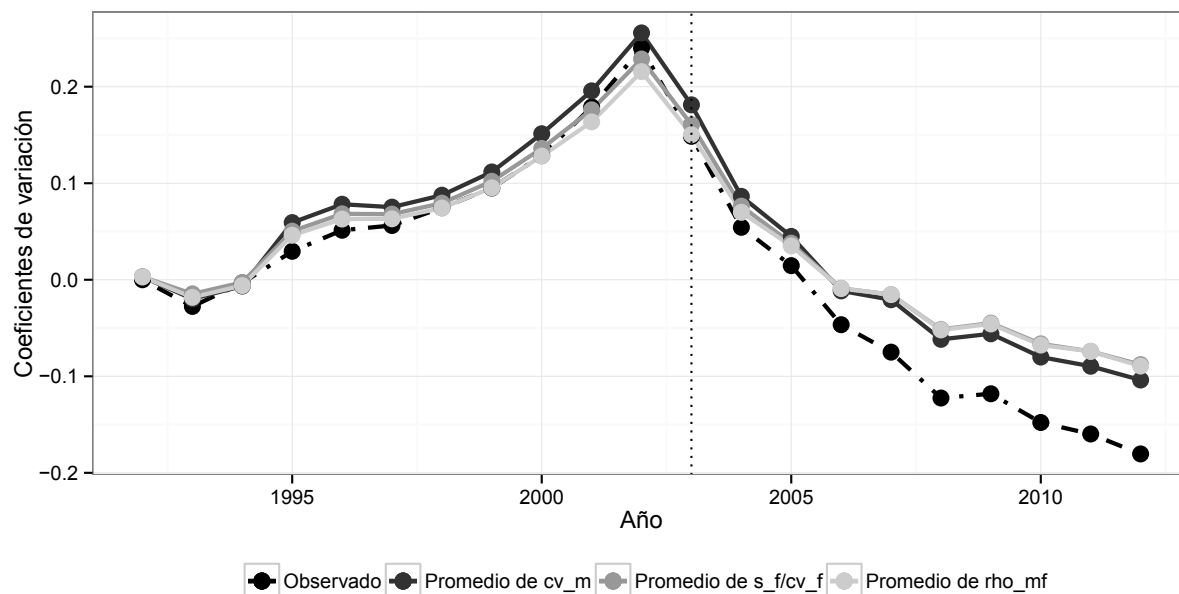


Figura 2.5: Contribuciones promedio de cada escenario contrafáctico

femeninos y la correlación entre ambos ocupan un rol relativamente secundario en su determinación. En otras palabras, la conclusión es robusta al ordenamiento de los factores y que, si la participación laboral femenina no hubiese incrementado, la desigualdad sería aún mayor, mientras que el emparejamiento selectivo sólo cumple un rol marginal en la determinación de la desigualdad del ingreso de las parejas casadas.

2.7. Conclusión

El objetivo de la presente sección consistió en estudiar cuál era la evidencia existente acerca del rol que cumple el emparejamiento selectivo en la determinación de la desigualdad en la distribución del ingreso total familiar. Si bien se utilizó un método en extremo sencillo y ya casi en desuso, su simplicidad resultó extremadamente útil para un objetivo también muy sencillo.

El método utilizado consistió en descomponer el coeficiente de variación del ingreso total familiar en un conjunto de determinantes:

- Coeficiente de variación de los ingresos totales (ingresos laborales y no laborales) de hombres y mujeres.
- Coeficiente de variación del ingreso residual.
- Participaciones de cada una de las categorías mencionadas sobre el total.
- Coeficientes de correlación de Pearson entre cada uno de los tipos de ingreso.

Se mostró que el emparejamiento selectivo cumple un rol muy secundario entre los determinantes de la desigualdad, tanto de las parejas casadas como de la totalidad de los individuos, debido a dos motivos. En primer lugar, su participación es relativamente baja (igual al producto entre las participaciones de hombres y mujeres y los coeficientes de variación de ambos) y, en segundo lugar, su valor apenas cambió en los 21 años utilizados para comparar su evolución, por lo que su contribución a la desigualdad fue casi constante.

Este resultado resulta muy importante por las consecuencias de política que acarrea, puesto que la magnitud del emparejamiento selectivo es una variable de difícil modificación por las políticas públicas, mientras que su contribución es marginal, implica que existe aún espacio para implementar mecanismos para mejorar la distribución del ingreso.

De acuerdo con estos resultados, una política de mejora de la distribución del ingreso debería focalizarse en la desigualdad de la distribución del ingreso masculino, el femenino y la participación laboral de éstas. Con respecto al último punto, en Gasparini y col. (2015) se muestra que en Argentina y, en gran parte de América Latina existe evidencia de una desaceleración en el crecimiento de la participación laboral femenina a partir de la segunda mitad de la década pasada, en particular la de las mujeres casadas, y como consecuencia de ello, su contribución a la reducción de la desigualdad tenderá a reducirse de continuar esta tendencia.

Capítulo 3

Emparejamiento selectivo en niveles educativos y desigualdad

3.1. Introducción

En esta capítulo se avanzará sobre el estudio del impacto del emparejamiento selectivo sobre la distribución del ingreso de una forma directa, ya que se utilizará la metodología de Greenwood y col. (2014), la cual consiste en un ejercicio de microdescomposición que permite estudiar el impacto de los cambios en el patrón de selectividad sobre un indicador de desigualdad, puntualmente el índice de Gini (Gasparini, Cicowiez y Sosa Escudero (2013), pág 374).

El espíritu del método de simulación es similar al empleado por Bredemeier y Juessen (2013), ya que se trata de una microdescomposición semiparamétrica que busca estudiar el efecto de cambios en el grado de asociación de los niveles educativos de las parejas casadas entre dos períodos de tiempo, pero en este caso no realiza ningún supuesto sobre la forma funcional de las preferencias. Éste modelo deriva de uno utilizado previamente por los mismos autores Greenwood y col. (2015), dónde se modela las decisiones de los individuos en función a la decisión de formar de parejas y sus respectivas ofertas laborales y niveles educativos, así mismo se derivan condiciones de equilibrio y de emparejamiento estable, las cuales sirven como base para otro método de simulación distinto al que se utilizará aquí.

En el mencionado artículo, los autores afirman que, si se aleatoriza la formación de parejas, el coeficiente de Gini para Estados Unidos en 2005 caería de 0,43 a 0,34, utilizando datos de la American Community Survey a través de IPUMS. Sin embargo, Eika, Mogstad y Zafar (2014) utilizando datos de la CPS (Current Population Survey), también para EEUU, encuentran que aleatorizar el emparejamiento genera una caída más modesta de la desigualdad, de 0,403 a 0,384, medida también por el índice de Gini. Por último, el trabajo de Hryshko, Juhn y McCue (2014) muestra una caída aún más modesta en la desigualdad de

0,295 a 0,290 en 2004 utilizando datos de Seguridad Social¹ y de 0,273 a 0,264 con datos del Panel Study of Income Dynamics (PSID).

3.2. Método de simulación

En esta sección se detallará el método de simulación utilizado por los autores y que se replicará aquí, éste método de simulación semiparamétrico parte de las tablas de contingencia que describen la distribución conjunta de los niveles educativos de los cónyuges en parejas casadas para simular que ocurriría con éstas si el emparejamiento fuera aleatorio. Es útil aquí seguir la taxonomía descrita en Harmenberg (2014), según la cual hay al menos dos enfoques contrapuestos para aleatorizar el emparejamiento.

A fines de evaluar cual es el efecto del emparejamiento selectivo, se comparan los resultados observados en la realidad con un contrafáctico, es decir cómo se comportarían las variables de ser aleatoria la formación de parejas. Los mencionados métodos difieren en qué variable mantienen constante a la hora de aleatorizar, en el primero se mantienen constantes los ingresos de los individuos, mientras que en el segundo son los ingresos del hogar los que se mantienen constantes.

El primero, denominado “enfoque de adición”, computa un ingreso de cada pseudo-hogar sumando el ingreso de los cónyuges, mientras que el segundo, llamado aquí “enfoque de imputación” imputa el pseudo-ingreso suponiendo que se distribuye de la misma forma que el ingreso real para hogares de similares características.

Sean:

x_m, x_f : características individuales de hombres y mujeres (en este caso, nivel educativo y participación laboral).

y_m, y_f : ingreso individual de cada uno de los cónyuges, de manera tal que en el ingreso total familiar del hogar i se define como: $y_i = y_{mi} + y_{fi}$.

$f(z_i)$: distribución efectiva de la variable z_i

Bajo el enfoque de adición, los atributos x_m, x_f se mantienen constantes en cada pareja, vale decir, se calcula la distribución:

$$(f(y_m|x_m) * f(y_f|x_f)) f(x_m, x_f) \quad (3.1)$$

Lo que equivale a suponer que los individuos cuyas características $m(f)$ se distribuyen aleatoriamente, de modo que $y_m|x_m$ e $y_f|x_f$ son independientes.

¹Los datos provienen de la Survey of Income and Program Participation panels (SIPP-SSA)

Bajo aleatorización no condicional, los atributos se mantienen constantes, por lo que la expresión anterior se reduce a:

$$f(y_m) * f(y_f) \quad (3.2)$$

Por lo que el enfoque de imputación computa la siguiente expresión: $f(y|x_m, x_f) * f(y_m) * f(y_f)$ y el ingreso del hogar se calcula a partir de sus características observables.

Cada enfoque tiene sus ventajas y desventajas, el primero tiene como ventaja no perder información al aleatorizar la población existente, pero considera a la oferta laboral y al ingreso como exógenos a la formación de hogares. El segundo, en cambio, posee la ventaja de tener en cuenta la oferta laboral y los ingresos a la hora de formar hogares, pero pierde información ya que los atributos no permiten realizar una imputación perfecta.

Este trabajo utiliza el método de imputación, puesto que aproxima la distribución del ingreso por quintiles, observa la distribución del ingreso para cada combinación de niveles educativos, y computa la distribución contrafáctica aleatoria para cada combinación de niveles educativos de las parejas. Luego se utiliza la mencionada distribución de niveles educativos como ponderadores y se suma la distribución del ingreso por quintiles a lo largo de los niveles de educación para obtener la nueva distribución del ingreso por hogares.

3.2.1. Descripción de los grupos

El ejercicio de simulación que utiliza el modelo requiere que se describan correctamente los grupos en los cuales se clasificará a los hogares, en primer lugar, las parejas casadas se indexan en base la educación del marido, de la mujer, la participación laboral de cada uno de ellos y el número de hijos (5 índices)²:

\mathcal{M}_{E_H} : índice de educación del marido, $E_H \in \{P-, P, S-, S, U-, U\}$ dónde cada categoría representa primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, superior incompleta y superior completa respectivamente.

\mathcal{M}_{E_W} : índice de educación de la mujer, $E_W \in \{P-, P, S-, S, U-, U\}$.

\mathcal{M}_{LFP_H} : índice de participación laboral del marido: $LFP_H \in \{WORK_H, \sim WORK_H\}$, donde la primer categoría refiere a los que participan del mercado laboral y la segunda a los que no.

\mathcal{M}_{LFP_W} : índice de participación laboral de la mujer: $LFP_W \in \{WORK_W, \sim WORK_W\}$.

²Se utiliza la misma nomenclatura que Greenwood y col. (2014) para facilitar la comparación.

\mathcal{M}_{KIDS} : índice de número de hijos: $KIDS \in \{0, 1, 2, 2+\}$, dónde la última categoría refiere a quienes tienen más de dos hijos.

El conjunto total de índices para parejas casadas es, entonces:

$$\mathcal{M} = \bigcup_{E_H, E_W, LFP_H, LFP_W, KIDS} (\mathcal{M}_{E_H} \cap \mathcal{M}_{E_W} \cap \mathcal{M}_{LFP_H} \cap \mathcal{M}_{LFP_W} \cap \mathcal{M}_{KIDS}) \quad (3.3)$$

De modo que el total de grupos para las parejas casadas será de: $6 \times 6 \times 2 \times 2 \times 4 = 576$

Para los hogares no casados, en cambio, las categorías son las siguientes:

$\mathcal{M}_{MARSTAT}$: índice de estado civil del individuo,

$MARSTAT \in \{NMAR_M, NMAR_F, DIV_M, DIV_F\}$, dónde cada categoría refiere a hombres solteros, mujeres solteras, hombres divorciados y mujeres divorciadas.

\mathcal{M}_E : índice de nivel educativo, $E \in \{P-, P, S-, S, U-, U\}$

\mathcal{M}_{LFP} : índice de participación laboral del individuo: $LFP \in \{WORK, \sim WORK\}$.

\mathcal{M}_{KIDS} : índice de número de hijos: $KIDS \in \{0, 1, 2, 2+\}$.

El conjunto de índices para personas no casadas es, entonces:

$$\mathcal{M} = \bigcup_{MARSTAT, E, LFP, KIDS} (\mathcal{M}_{MARSTAT} \cap \mathcal{M}_E \cap \mathcal{M}_{LFP} \cap \mathcal{M}_{KIDS}) \quad (3.4)$$

El número total de grupos para los solteros y divorciados es: $4 \times 6 \times 2 \times 4 = 192$, de aquí se tiene que el número total de grupos asciende a $576 + 192 = 768$.

Descripción de los datos

Los datos para realizar el ejercicio de simulación provienen también de la **Encuesta Permanente de Hogares** (EPH) de Argentina para los años 1992 a 2012, a diferencia de Greenwood y col. (2014) no se utilizaron datos de IPUMS³ por dos motivos: el primero reside en que éstos se originan en los datos de los censos de población y, en el caso de Argentina, los censos no relevan datos de ingresos; el segundo motivo reside en que al momento de escribir este artículo los datos del censo de población de 2010 no se encuentran cargados en su totalidad, no

³Integrated Public Use Microdata Samples, una base de datos de uso libre cuyo objetivo consiste en armonizar datos de censos de diferentes países del mundo a través del uso de muestras aleatorias de datos provistos por las agencias de estadísticas de cada país.

pudiéndose aún siquiera calcular las matrices de asociación entre cónyuges para dicho año.

Una cuestión accesoria al uso de las Encuestas Permanentes de Hogares que debe tenerse en cuenta es que en la mitad del período bajo análisis hubo un cambio metodológico de gran magnitud en éstas, el cual consistió principalmente en (Beccaria y Groisman, 2008):

- El relevamiento pasó de ser puntual con dos ondas (mayo y octubre) a ser continuo con cuatro ondas (una por trimestre).
- Se modificó el cuestionario en preguntas clave, las cuales limitaron la comparabilidad de datos en algunas características de la población relevada.
- Se incrementó el número de aglomerados relevados por la encuesta, incrementando su grado de representatividad de la población urbana.

Éstos cambios afectan a los resultados del estudio, aunque de una forma tangencial, ya que sólo se trabaja con ingresos agregados a nivel individual y familiar, así como también con características como situación ocupacional, estado civil, edad, etc. que no fueron modificados por los cambios metodológicos; sin embargo, el incremento de la frecuencia de muestro impacta positivamente en el número de datos totales por año, ya que se éstos duplican a partir de 2003.

En primer lugar, se eliminan de la muestra los hogares secundarios y a todos los individuos que no correspondan a las categorías “Jefe/a”, “Cónyuge/Pareja” o “Hijo/Hijastro/a”, así mismo se utilizará sólo a las personas cuyas edades estén comprendidas entre los 24 y 55 años, nuevamente con el objeto de captar a la menor cantidad posible de estudiantes y jubilados. Los hogares se clasifican según las cinco categorías de estado civil mencionadas y también según el género de sus miembros.

Una vez hecho ésto se generan los niveles educativos, a diferencia del artículo de Greenwood aquí se utilizan seis niveles de educación, puesto que si se utilizara la clasificación de éstos⁴ habría muy pocas observaciones en las últimas dos categorías, motivo por el cual también se escogió unir las categorías de educación superior no universitaria con la universitaria en una única categoría.

Luego, la simulación requiere el cálculo de un ingreso total familiar equivalente, en este caso se optó por no utilizar la escala de equivalencias de la OCDE que utiliza el trabajo original, la cual asigna una ponderación de 1 al hombre adulto, 0,5 a la mujer adulta y 0,3 a cada uno de los niños que pueda tener la pareja, sino que se utilizó la escala de adulto equivalente del INDEC:

⁴ *menos que secundario completo, secundario completo, universitario incompleto, universitario completo y más que universitario completo*

Cuadro 3.1: Tabla de adulto equivalente-INDEC

Edad	Sexo	Valor Adulto equivalente
Menor a un año	Ambos	0,33
1 año		0,43
2 años		0,50
3 años		0,56
4 a 6 años		0,63
7 a 9 años		0,72
10 a 12 años	Varones	0,83
13 a 15 años		0,96
16 a 17 años		1,05
10 a 12 años	Mujeres	0,73
13 a 15 años		0,79
16 a 17 años		0,79
18 a 29 años	Varones	1,06
30 a 59 años		1,00
60 y más años		0,82
18 a 29 años	Mujeres	0,74
30 a 59 años		0,74
60 y más años		0,64

Fuente: Bergés (2011).

El siguiente paso consiste en identificar a hombres y mujeres que formen parte de la **Población Económicamente Activa** (PEA), para tal caso se genera una variable dummy igual a uno si la persona forma parte de la PEA o a cero en caso contrario, en el caso de las parejas casadas se generan dos, una para cada cónyuge; por último, se genera la variable categórica de número de hijos.

3.2.2. Simulación

Para realizar los cálculos que el método requiere se utilizó el código provisto por los autores⁵, éste se compone de dos rutinas, la primera, es un programa de **Stata** que utiliza como insumos los archivos de censos de **IPUMS** y calcula las regresiones detalladas en la sección siguiente, así como también los índices por grupos, las matrices de asociación de cada período, la participación laboral femenina por cuantil, etc. La segunda son un conjunto de programas en **Matlab** que realizan las simulaciones que se detallan posteriormente.

Puesto que la estructura de los datos de la **Encuesta Permanente de Hogares** es muy diferente de la de IPUMS no es posible utilizar directamente el código para generar grupos y, como dificultad adicional en su aplicación, el programa requiere el cálculo del coeficiente de correlación por grupos, denominado

⁵Disponible en <http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.5.348>

“ τ de Kendall” cuyo cálculo en **Stata** resulta demasiado lento en términos de tiempo de cómputo.

A fines de evitar estos inconvenientes se replicó el análisis por medio de una serie de programas en **R**⁶, el primero para filtrar las variables necesarias en cada onda de las EPH, generar las observaciones por hogar y repetir el análisis para todas las ondas de la encuesta. El segundo programa calcula los estadísticos por grupo, tablas de contingencia y demás estadísticos, así como también la τ de Kendall de forma más eficiente⁷.

Una vez obtenidos los datos de salida, se corre el programa de **Matlab**, previa adaptación al hecho que se trabajará con 5 quintiles y seis categorías de nivel educativo, lo que implicó modificar gran parte del código para que el cómputo y la presentación de los resultados se adapte a la base de datos que se creó previamente.

Indicadores de emparejamiento selectivo

El modelo busca responder dos preguntas fundamentales:

1. ¿Hubo un incremento en el emparejamiento selectivo en los hogares durante el período que se estudia?
2. ¿En qué medida contribuye el emparejamiento selectivo a la desigualdad del ingreso de los hogares?

Para responder la primera pregunta se verá en esta sección en qué medida éstos fueron incrementándose en el tiempo, mientras que la segunda requiere un análisis más riguroso, motivo por el cual se utilizará el mencionado método.

En el artículo de Greenwood y col. (2014) se introducen una serie de indicadores de evolución del emparejamiento selectivo en el tiempo.

3.2.3. Coeficientes de regresión

La primer medida parte de la estimación de la ecuación:

⁶Un objetivo adicional perseguido durante la elaboración de éste trabajo consistió en optimizar el tiempo de ejecución del código, ya que el original de Greenwood y col. (2014) requiere poco mas de 6 horas para correr por completo (Con procesador Intel Celeron de núcleo dual con 4Gb de capacidad @ 1,50 GHz), mientras que los programas utilizados aquí lo hacen en aproximadamente 20 minutos, en gran parte gracias al uso de las funciones de los paquetes `dplyr` y `tidyr` descritas en Wickham (2011) y Wickham (2014) las cuales permiten reducir al mínimo el uso de bucles y con ello, mejorar la eficiencia de cómputo.

⁷Utilizando la función `cor.fk` de la biblioteca de programas “pcaPP” (Principal Components Analysis by Projection Pursuit), la razón de la ganancia en eficiencia de cómputo reside en la utilización del algoritmo Christensen-Abrevaya (Abrevaya, 1999), en vez del implementado en Stata.

$$EDU_{it}^f = \alpha + \beta EDU_{it}^m + \sum_{t \in T} \gamma_t * EDU_{it}^m * \text{Año}_{st} + \sum_{t \in T} \theta_t * \text{Año}_{st} + \epsilon_{it} \quad \epsilon_{it} \sim N(0, \sigma) \quad (3.5)$$

Dónde:

EDU_{it}^f, EDU_{it}^m : son los niveles educativos de hombres (m) y mujeres (f) que forman la pareja i en el año t .

Año_{st} : son dummies por año $t = 1992, \dots, 2012$ tales que $\text{Año}_{st} = 1$ si $s = t$ y $\text{Año}_{st} = 0$ en caso contrario.

El coeficiente β mide el grado de asociación entre los niveles educativos de los cónyuges en el año base (1992), mientras que los coeficientes γ_t miden la contribución relativa de la educación del marido con respecto a su mujer para los años posteriores al base y sirven como una medida del emparejamiento selectivo. Los coeficientes θ_t se utilizan para controlar por el incremento exógeno en los niveles educativos de hombres y mujeres.

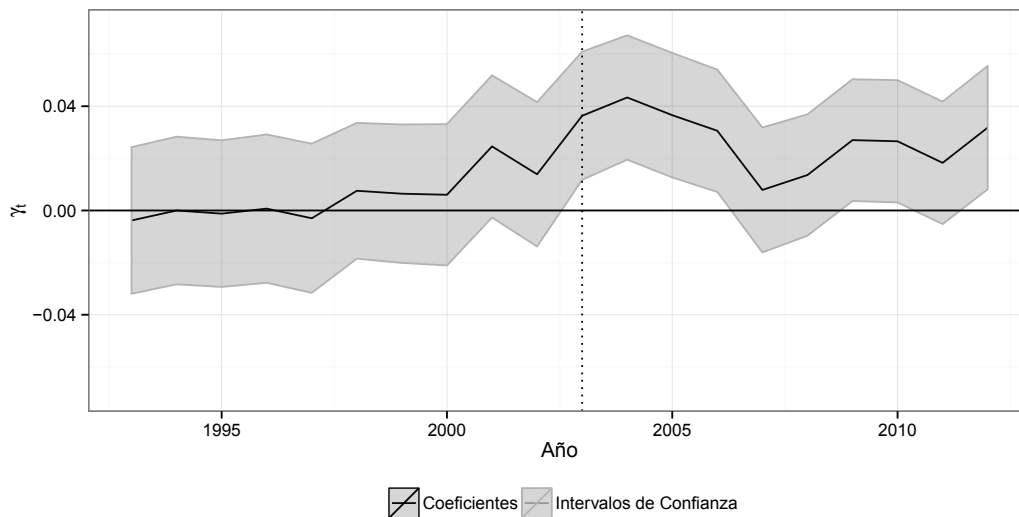


Figura 3.1: Incremento en el emparejamiento selectivo

La Figura 3.1 muestra que existe una leve tendencia al alza en el período estudiado, pero que gran parte de los coeficientes no son significativos al 5%, ya que su intervalo de confianza contiene al cero para la mayoría de los años del período, en general el incremento en la fuerza del emparejamiento es mayor en la segunda mitad del período que en la primera, aunque éste fenómeno no es tan notorio como sucede con los datos de Greenwood, en gran parte debido a que el período de tiempo es mucho más reducido al mismo tiempo que la frecuencia es mayor, lo que perjudica la significatividad de los resultados, dado que el fenómeno estudiado se caracteriza por ser estructural.

3.2.4. Tau de Kendall

Otra medida presentada por los autores es el coeficiente de correlación por rangos, denominado “Tau de Kendall”, el cual se utiliza para medir la fuerza de la asociación entre dos variables categóricas, en este caso se utiliza para medir la correlación entre los niveles educativos de los cónyuges.

El mencionado coeficiente (Kendall, 1970) se define de la siguiente forma:

$$\tau = \frac{S}{\frac{n(n-1)}{2}} \quad (3.6)$$

Dónde S es igual a la diferencia entre el número total de pares concordantes (P) y discordantes (Q), en este caso un par es concordante si tanto marido como mujer poseen el mismo nivel educativo y discordante en caso contrario, de modo tal que $P + Q = \frac{n(n-1)}{2}$, por lo que:

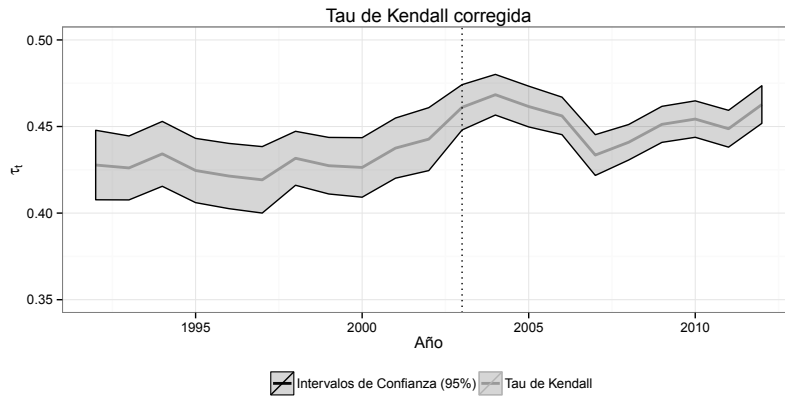
$$\begin{aligned} \tau &= \frac{P - Q}{\frac{n(n-1)}{2}} \\ &= 1 - \frac{2Q}{\frac{n(n-1)}{2}} \\ &= \frac{2P}{\frac{n(n-1)}{2}} - 1 \end{aligned}$$

El coeficiente de correlación por categorías de Kendall se distribuye como normal (Worner, 2006), pero en los datos existen numerosos “empates”, es decir, cuando un par no es ni concordante ni discordante, lo que puede hacer que el coeficiente se aleje de su rango para alguna pareja. Para resolver este problema se calculará el llamado “Z-score” de Fisher, con el objeto de poder estimar los intervalos de confianza de la tau de Kendall para cada año.

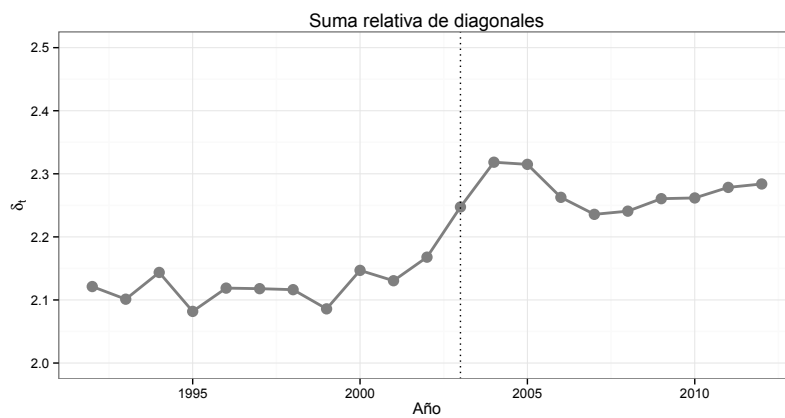
El Z-score es una transformación que permite estabilizar la varianza corrigiéndola por posible no normalidad y se define como la siguiente magnitud, dónde τ es la tau de Kendall y \log es el logaritmo neperiano:

$$Z = \frac{1}{2} [\log(1 + \tau) - \log(1 - \tau)]$$

El error estándar de éste estadístico es igual a: $SE(Z) = \sqrt{\frac{1}{n-3}}$, dónde n es el número de parejas en la muestra y con éstos errores es posible armar intervalos de confianza puesto que el estadístico también se distribuye como normal.



(a) Coeficiente de correlación por rangos para el nivel educativo de los cónyuges



(b) Ratio de la suma de diagonales principales

Figura 3.2: Indicadores de emparejamiento selectivo

Siguiendo a Worner (2006), puesto que el valor de la τ de Kendall toma valores entre -1 y 1 , su valor mide la fuerza de la correlación entre dos variables categóricas, en nuestro caso, los niveles educativos de los cónyuges. Un valor negativo del coeficiente indica que la posición de las variables en el ranking se incrementa de forma conjunta, mientras que un valor positivo indica que el de una se incrementa a medida que el de la otra se reduce. De ésta forma, si el valor del coeficiente es de, por ejemplo, $0,25$, si se selecciona a una familia aleatoriamente, existe una probabilidad de $0,25$ que los miembros de ésta familia posean el mismo nivel educativo.

No existe un consenso sobre que valores constituyen una relación “fuerte” o “débil”, pero si puede utilizarse la siguiente regla informal:

- $\pm 0,01 \leq \tau < \pm 0,3$ Relación débil
- $\pm 0,3 \leq \tau < \pm 0,7$ Relación moderada
- $\tau \geq \pm 0,7$ Relación fuerte

En la Figura 3.2 puede observarse que la evolución del coeficiente de correlación entre niveles educativos de los cónyuges evoluciona de manera muy similar al de los coeficientes de las interacciones en la regresión detallada anteriormente. El valor del coeficiente τ también se mantiene relativamente constante durante la década de 1990, registra una leve suba a principios de la década posterior⁸ para luego amesetarse en un nivel más alto a partir de 2005. Según la regla detallada anteriormente puede afirmarse que la relación entre los niveles educativos de los cónyuges es “moderada” puesto que el valor del estadístico se ubica por encima de 0,3, pero por debajo de 0,7.

3.2.5. Suma de diagonales

Otro indicador del emparejamiento selectivo surge a partir de la matriz de asociación entre niveles educativos de parejas casadas: sea a_{ij}^t la celda correspondiente al nivel educativo i del marido y j de su esposa para el año t en la respectiva matriz, entonces la suma de la diagonal principal (valores para los cuales $i = j$) se define como: $\sum_{i=1}^6 a_{ii}^t$.

Por otro lado, sea r_{ij}^t la celda equivalente en una matriz obtenida como el producto interno de los totales por filas de la matriz de asociación, con el objeto de simular la matriz que existiría si el emparejamiento fuese aleatorio y sea $\sum_{i=1}^6 r_{ii}^t$ su respectiva suma de las celdas diagonales, entonces se define a:

$$\delta_t = \frac{\sum_{i=1}^6 a_{ii}^t}{\sum_{i=1}^6 r_{ii}^t} = \frac{tr(\mathbf{A})_t}{tr(\mathbf{R})_t} \quad (3.7)$$

Dónde $tr(\mathbf{A})$ y $tr(\mathbf{R})$ son las trazas de las matrices de asociación empíricas y simulada respectivamente.

El valor de los coeficientes δ_t es mayor que uno en todos los años, lo que implica que el emparejamiento selectivo es positivo, sin embargo, concurda con los resultados obtenidos con los otros indicadores en el hecho que su comportamiento, si bien es creciente carece de la fuerza que se observa para los datos censales de Estados Unidos.

3.2.6. Tablas de contingencia

Otra medida del emparejamiento selectivo puede encontrarse observando las tablas de contingencia, las cuales miden las frecuencias de las combinaciones de los seis niveles educativos de los cónyuges para las parejas casadas en un

⁸En esta figura y en las posteriores, la línea vertical punteada marca el año 2003, cuando el INDEC modificó la metodología de muestreo de la Encuesta Permanente de Hogares, se señala este período para advertir por posibles cambios exógenos surgidos por ésta modificación.

período específico, para ello se mostrarán las correspondientes al primero (1992-1996) y al último (2008-2012) para mostrar su evolución:

Cuadro 3.2: Tabla de contingencia para 1992-1996

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,165	0,040	0,041	0,055	0,008	0,005
P	0,037	0,044	0,005	0,012	0,000	0,001
S-	0,027	0,004	0,082	0,034	0,032	0,014
S	0,059	0,011	0,047	0,067	0,015	0,011
U-	0,002	0,000	0,024	0,006	0,062	0,015
U	0,004	0,000	0,019	0,010	0,021	0,018
Dist. marginal	0,296	0,099	0,217	0,185	0,138	0,065

Fuente: Elaboración propia en base a INDEC.

Cuadro 3.3: Tabla de contingencia para 2008-2012

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,104	0,017	0,040	0,041	0,011	0,007
P	0,019	0,017	0,004	0,009	0,001	0,001
S-	0,025	0,003	0,116	0,033	0,060	0,028
S	0,033	0,006	0,046	0,062	0,020	0,018
U-	0,003	0,000	0,024	0,005	0,110	0,024
U	0,003	0,000	0,021	0,007	0,041	0,035
Dist. marginal	0,188	0,045	0,252	0,158	0,243	0,114

Fuente: Elaboración propia en base a INDEC.

Los cuadros 3.2 y 3.3 muestran la distribución conjunta de cada combinación de niveles educativos en los períodos inicial y final, allí puede verse cómo en la mayoría de los casos el valor más alto de cada fila/columna coincide con el de la diagonal principal. También se observa cómo las frecuencias de las combinaciones de niveles más altos (Superior incompleto y completo) se incrementan en el período 2008-2012 en relación a los valores del período 1992-1996, lo que refleja el incremento de los niveles educativos, particularmente el de las mujeres, tal como se observa en la fila de distribución marginal, donde los niveles de educación más altos son más frecuentes en el último período que en el primero, por ejemplo, las frecuencia de las mujeres con nivel superior completo se duplicó prácticamente en el período.

3.3. Indicadores de desigualdad

3.3.1. Definición de los indicadores

El segundo objetivo del trabajo consiste en evaluar el impacto del incremento del emparejamiento selectivo en la desigualdad, para ello se deben utilizar indicadores de ésta, en este trabajo se utilizarán dos, además del mencionado índice de Gini se utilizarán las curvas de Lorenz.

Sean:

f_{ij} : fracción de hogares del tipo i en el percentil de ingreso j : $f_{ij} = \frac{N_{ij}}{N}$, dónde N es el tamaño de la muestra y N_{ij} el del grupo (i, j) .

r_{ij} : ingreso de los hogares i relativo al ingreso familiar medio de la población:
 $r_{ij} = \frac{y_{ij}/N_{ij}}{Y/N}$, donde y_{ij} es el ingreso familiar total del grupo (i, j) y Y el ingreso familiar total de la muestra.

s_j : participación del ingreso agregado del percentil j : $s_j = \sum_i f_{ij} r_{ij}$.

I_p : participación acumulada del percentil p : $I_p = \sum_j^p s_j = \sum_j^p \sum_i f_{ij}$

Debido a que se trabaja con datos de encuestas, el tamaño de la muestra utilizada en este trabajo es mucho más reducida que la que utilizan Greenwood et Al, por tal motivo, ambos indicadores se calcularán por quintiles y no por deciles, con el objeto de reducir el número total de grupos, de modo tal que $j \in \{0,1; 0,2; 0,3; 0,4; 0,5\}$ y por tal motivo el número total de grupos será igual a $768 \times 5 = 3840$.

En base a ésto, la curva de Lorenz se define cómo el gráfico de la dupla (p, I_p) , dónde $p = \sum_j^p \sum_i f_{ij}$. Del mismo modo el coeficiente de Gini, igual a dos veces el área por debajo de la curva de Lorenz y una recta de 45° , se define cómo:

$$g = 2 \int_0^1 |I_p - p| dp, \quad \text{con: } 0 \leq g \leq 1 \quad (3.8)$$

Por lo que ambos índices son funciones de los f_{ij} y los r_{ij} :

$$\begin{aligned} l_p &= \text{Lorenz}_p(\{f_{ij}\}, \{r_{ij}\}) \\ g &= \text{Gini}(\{f_{ij}\}, \{r_{ij}\}) \end{aligned}$$

Para poder computar el coeficiente de Gini se utilizará la fórmula desarrollada por Rao (1969), la que permite descomponer dicho índice en un número arbitrario de n grupos de la siguiente forma:

$$g = \sum_{p=1/n}^{1-1/n} \left[pI_{p+1} - \left(p + \frac{1}{n} \right) I_p \right] \quad (3.9)$$

Para graficar la curva de Lorenz se divide el intervalo $[0, 1]$ en n segmentos equiespaciados: $j \in \mathcal{J} = \{\frac{1}{n}, \dots, 1 - \frac{1}{n}\}$, donde el valor de n es la cantidad de cuantiles que se utilizarán ($n = 4$: cuartiles, $n = 5$: quintiles, $n = 10$: deciles, $n = 100$: percentiles)

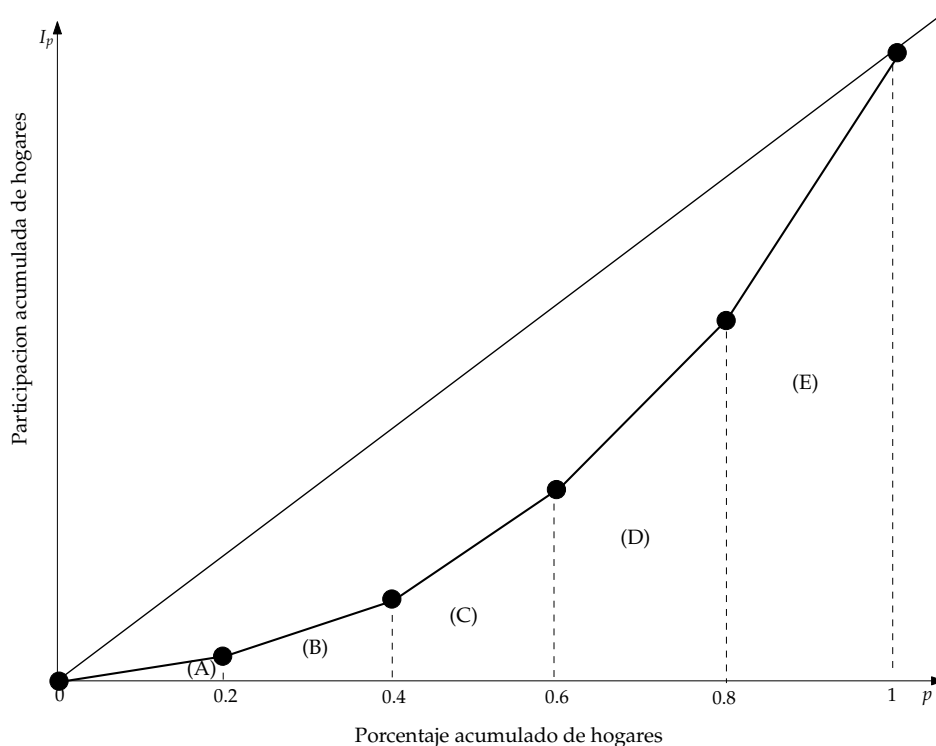


Figura 3.3: Curva de Lorenz

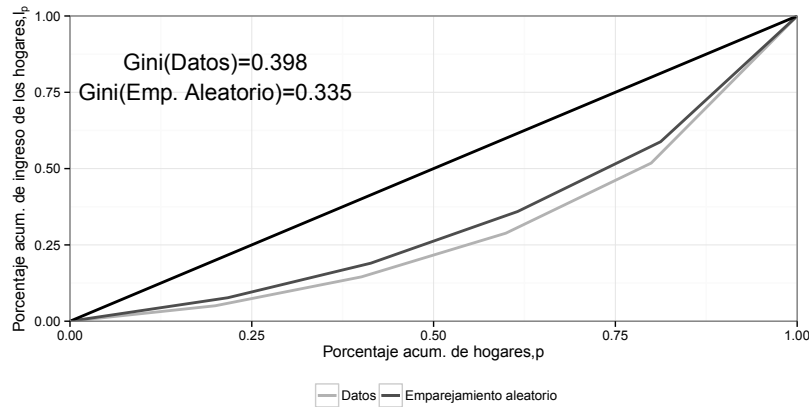
El coeficiente de Gini se calcula en base a las áreas de la Figura 3.3 de la siguiente manera:

$$g = 1 - 2 \left[\underbrace{\frac{p_1 I_1}{2}}_{(A)} + \underbrace{\frac{(I_1 + I_2)(p_2 - p_1)}{2}}_{(B)} + \underbrace{\frac{(I_2 + I_3)(p_3 - p_2)}{2}}_{(C)} + \underbrace{\frac{(I_3 + I_4)(p_4 - p_3)}{2}}_{(D)} + \underbrace{\frac{(I_4 + 1)(1 - p_4)}{2}}_{(E)} \right] \quad (3.10)$$

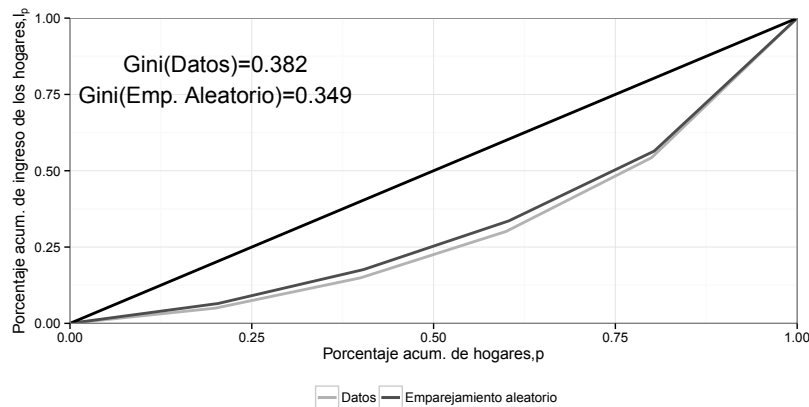
Esto es, el coeficiente de Gini es igual a: $1 - 2 \times (\text{Área bajo la curva de Lorenz})$,

dónde el área bajo la curva de Lorenz se aproxima como la suma de las áreas de los trapezoides (A), (B), (C), (D) y (E), puesto que $p_1 = 0,1$, $p_2 = 0,2$, $p_3 = 0,3$, $p_4 = 0,4$ y $p_5 = 1$, la fórmula del coeficiente de Gini puede reexpresarse cómo:

$$g = \frac{1}{5} [(I_2 - 2I_1) + (2I_3 - 3I_2) + (3I_4 - 4I_3) + (4 - 5I_4)] \quad (3.11)$$



(a) Período 1992-1996



(b) Período 2008-2012

Figura 3.4: Curvas de Lorenz: Datos versus emparejamiento aleatorio

En la figura 3.4 se muestran las curvas de Lorenz y los coeficientes de Gini en los períodos inicial y final, en ellos puede apreciarse que la variación en la desigualdad resulta ser relativamente baja, debido a que se calculan en base a promedios de años en los cuales la desigualdad creció (período 1) y luego se redujo (período 4), en el período 1992-1996 el 20 % más pobre ganaba apenas el 5 % del ingreso total, mientras que el 20 % más rico se quedaba con poco menos del 50 % del total del ingreso. En el período 2008-2012 los porcentajes se modificaron, pero en una magnitud reducida, ya que el quintil inferior ahora se queda con poco más del 5 % del total, mientras que el quintil más alto se queda con aproximadamente el 47 % del total del ingreso total familiar⁹.

⁹Por “Ingreso Total Familiar” se hará referencia siempre al ingreso familiar corregido por adulto

Por otro lado, cuando se compara la desigualdad que surge de los datos con la que surge de imponer el emparejamiento aleatorio, se observa que ésta cae 6,3 puntos (en el período 1992-1996) y 3,3 puntos (período 2008-2012).

3.4. Experimentos contrafácticos

Los experimentos contrafácticos se llevarán a cabo en base a cambios en dos características de la población: la distribución conjunta de niveles de educación entre los cónyuges y la participación laboral femenina; en base a cambios en la distribuciones de éstos se estimarán los coeficientes de Gini respectivos.

El coeficiente de Gini puede escribirse de la siguiente forma (Cowell y Flachaire, 2015):

$$G = 1 - 2 \int_0^1 L(F; q) dq \quad (3.12)$$

Dónde $L(F, q)$ es la q -ésima ordenada de la curva de Lorenz, que depende de la función de distribución del ingreso F , de la siguiente forma:

$$L(F, q) = \frac{C(F, q)}{\mu(F)} \quad (3.13)$$

$\mu(f)$ denota a la media de la distribución del ingreso y $C(F, q)$ es el funcional de ingreso acumulado definido como:

$$C(F, q) = \int_{\underline{y}}^{\bar{y}} y dF(y) \quad (3.14)$$

Juntando todos los términos, se obtiene la siguiente expresión para el coeficiente de Gini:

$$G = 1 - \frac{2}{\mu(F)} \int_0^1 \int_{\underline{y}}^{\bar{y}} y * f(y, E_H, E_W, LFP_W) dy dq \quad (3.15)$$

Reemplazando por las distribuciones condicionales y marginales de las características, se obtiene la siguiente expresión para $t = 1, 4$:

$$G_t = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_t} y * f^t(y/E_H, E_W, LFP_W) * g^t(E_H, E_W/LFP_W) * h^t(LFP_W) dy dq \quad (3.16)$$

equivalente de la forma detallada en la sección anterior.

Con:

$$\begin{aligned} f(y, E_H, E_W, LFP_W) &= f(y/E_H, E_W, LFP_W) \times g(E_H, E_W, LFP_W) \\ g(E_H, E_W, LFP_W) &= g(E_H, E_W/LFP_W) \times h(LFP_W) \end{aligned}$$

Los experimentos contrafácticos se basan en modificaciones en las distribuciones de los niveles educativos de los cónyuges $f(E_H, E_W)$ (emparejamiento aleatorio y estandarización de tablas de contingencia) y en cambios en la distribución de la participación laboral femenina según quintiles del ingreso per cápita familiar $f(LFP_W)$.

Emparejamiento aleatorio

El primer experimento contrafáctico consiste en simular los resultados que surgirían si el emparejamiento fuese totalmente aleatorio y calcular los indicadores de desigualdad que surgirían en ambos períodos. Sea $\mathcal{M} = 1, \dots, 576$ el índice para identificar a las parejas casadas, $\mathcal{S} = 577, \dots, 768$ el índice para las personas solteras y divorciadas, entonces el experimento contrafáctico parte de reemplazar las f_{ij} para $(i, j) \in \mathcal{M}$ observadas por las que existirían si el emparejamiento fuese aleatorio: \tilde{f}_{ij} para $(i, j) \in \mathcal{M}$, por lo que las curvas de Lorenz y coeficientes de Gini contrafácticos serán:

$$\begin{aligned} l_p &= \text{Lorenz}_p(\{f'_{ij}\}, \{r_{ij}\}) \\ g &= \text{Gini}(\{f'_{ij}\}, \{r_{ij}\}) \\ \{f'_{ij}\} &\equiv \{\tilde{f}_{ij}\}_{\mathcal{M}} \cup \{f_{ij}\}_{\mathcal{S}} \end{aligned}$$

En términos de las distribuciones conjuntas, el coeficiente de Gini para los períodos $i = 1, 4$ es ahora:

$$G_i^{RM} = 1 - \frac{2}{\mu(F)_i} \int_0^1 \int_{\underline{y}_i}^{\bar{y}_i} y f^i(y/E_H, E_W, LFP_W) g_R^i(E_H, E_W/LFP_W) h^i(LFP_W) dy dq \quad (3.17)$$

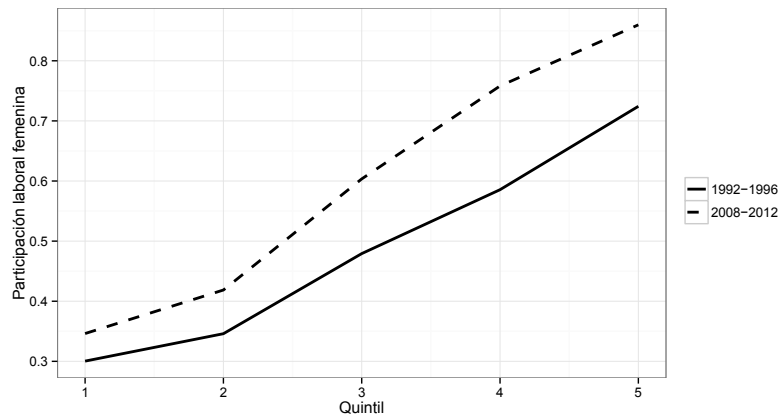
Dónde $g_R^i(E_H, E_W/LFP_W)$ es la distribución conjunta de niveles educativos contrafáctica suponiendo emparejamiento aleatorio, de modo tal que las diferencias entre los valores del índice de Gini para cada período se atribuyen a cambios en la distribución de niveles educativos únicamente.

La modificación de los patrones de emparejamiento en cualquiera de los períodos arroja los resultados esperados, puesto que la desigualdad cae al aleatorizar la distribución de los niveles educativos de las parejas resultados, éste

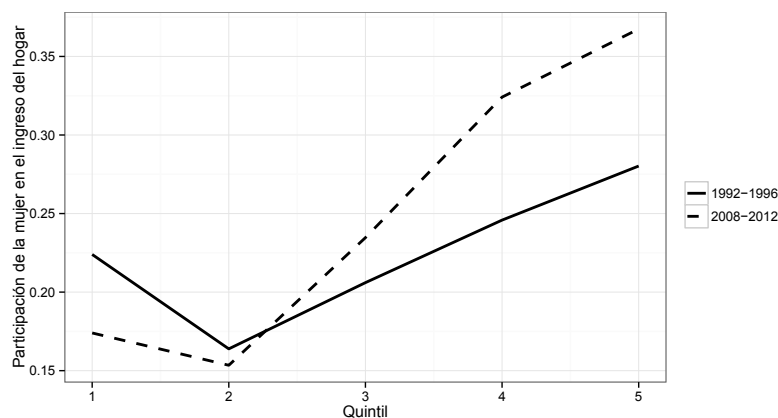
resultado sigue la línea de los encontrados por Greenwood y col. (2014), sin embargo resulta llamativo que el efecto del emparejamiento aleatorio sea mayor en el primer período que en el segundo (6,3 puntos contra 3,3) habida cuenta que según los indicadores mostrados en su respectiva sección muestran un leve incremento. Las tablas de contingencia que se obtiene utilizando la simulación de emparejamiento aleatorio se detallan en el anexo (Tablas 4.6 y 4.7).

Participación laboral femenina

Otro motor de la desigualdad en el ingreso familiar es la participación laboral de las mujeres, particularmente de las casadas, éste posee el obvio efecto de reducir la desigualdad intrahogar, aunque el efecto que ésta tiene no debe ser necesariamente lineal a lo largo de los quintiles. Del mismo modo, la mayor participación de la mujer en el ingreso total familiar (consecuencia de la anterior) tampoco debe serlo.



(a) Participación laboral femenina



(b) Participación de la mujer en el ingreso total familiar

Figura 3.5: Participación laboral de las mujeres casadas: 1992-1996 vs. 2008-2012

El efecto del emparejamiento selectivo depende de que las mujeres se en-

cuentren empleadas, en la Figura 3.5(a) se observa que ésto ocurre con mayor frecuencia en 2008-2012 que en 1992-1996, también, la variación en la participación es mayor en los quintiles más altos de la distribución del ingreso, en la Figura 3.5(b) puede verse que lo mismo ocurre con la participación del ingreso laboral femenino en el total del hogar, ésta también crece con los quintiles del ingreso.

En primer lugar se realizará un experimento contrafáctico que consiste en volver a hacer al emparejamiento aleatorio en ambos períodos, pero ahora haciendo que las mujeres en el período 1(4) participen como las del período 4(1), ésto equivale a modificar tanto la distribución condicional de niveles educativos como la distribución marginal de la participación laboral femenina:

$$G_t^{RM+\Delta LFP} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_t} y * f^t(y/E_H, E_W, LFP_W) * g_R^t(E_H, E_W/LFP_W) h^s(LFP_W) dydq \quad (3.18)$$

Dónde $s = 4$ si $t = 1$ y $s = 1$ si $t = 4$ de modo tal que el efecto agregado se debe a las variaciones de dos distribuciones, la distribución conjunta de niveles educativos y la distribución de la participación laboral femenina.

En la Figura 4.6 del anexo se observa que el hacer variar la participación laboral femenina reduce la desigualdad en el primer período de 0,398 a 0,334 y de la misma manera, reduce la desigualdad en el segundo período de 0,382 a 0,348, según los autores ésta reducción de la desigualdad se origina en la diversificación del ingreso familiar entre maridos y mujeres.

Nótese que casi toda la reducción en la desigualdad de ambos períodos se debe al emparejamiento aleatorio, ya que el efecto incremental de la participación laboral femenina es prácticamente despreciable, estos resultados contrastan con los obtenidos por Greenwood et Al, para los cuales el efecto de la participación laboral femenina refuerza al del emparejamiento aleatorio de forma notable. Las tablas de contingencia que surgen de imponer la participación laboral femenina del período contrapuesto corresponden con las Tablas 4.8 y 4.9.

Tablas de contingencia estandarizadas

Otro experimento consiste en reemplazar las tablas de contingencia de cada período por la del otro y analizar los resultados que se obtienen, éste ejercicio posee un inconveniente, ya que el nivel educativo promedio del segundo período es superior al del primero y por tras de ello las distribuciones marginales de los niveles educativos de maridos y mujeres han cambiado, ésto puede verse en las distribuciones marginales de los Cuadros 3.1 y 3.2.

Para evitar el problema que surge de utilizar tablas con distribuciones marginales diferentes que distorsionen los patrones de asociación entre dos variables cualitativas se utilizan tablas estandarizadas, las cuales pueden construirse a partir de las distribuciones marginales de ambos períodos, una para 1992-1996 con las distribuciones de 2008-2012 y otra con el orden inverso.

Las tablas se construyen utilizando el procedimiento iterativo detallado en Mosteller (1968):

1. Ingresar una iteración en la tabla de contingencia.
2. Esta tabla de contingencia posee una distribución marginal asociada con las filas (hombres) que se obtiene sumando cada fila a lo largo de las columnas para obtener un total por fila. Luego se divide cada fila en 6 hasta obtener una distribución marginal de las filas igual a: (1/6, 1/6, 1/6, 1/6, 1/6, 1/6).
3. Computar la distribución marginal asociada a las columnas (mujeres) sumando cada columna a lo largo de sus filas para obtener un total por columna, luego se divide cada columna por 6.
4. Se calcula nuevamente la distribución marginal de las filas, luego de los pasos anteriores debe haber cambiado. Se calcula al diferencia con relación a la distribución marginal deseada: (1/6, 1/6, 1/6, 1/6, 1/6, 1/6). Si se alcanzó al cercanía deseada, parar, caso contrario volver al paso 1.

Nuevamente, en términos de la fórmula del índice de Gini, aquí se repiten dos experimentos, en el primero se modifica sólo la distribución de niveles educativos, reemplazándola por la obtenida con el procedimiento descrito anteriormente y la segunda, es igual a ésta, agregándole variaciones en la distribución de la participación laboral femenina:

$$G_t^{ST} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_i} y * f^t(y/E_H, E_W, LFP_W) * g_E^t(E_H, E_W/LFP_W) * h^t(LFP_W) dydq$$

$$G_t^{ST+\Delta LFP} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_i} y * f^t(y/E_H, E_W, LFP_W) * g_E^t(E_H, E_W/LFP_W) * h^s(LFP_W) dydq$$

En la Figura 4.7(a) se muestran los cálculos detallados en base a la tabla de contingencia estandarizada, aquí se observa que casi no hay cambios con respecto al segundo experimento contrafáctico, se observa una reducción de menor magnitud en los coeficientes de Gini vis-à-vis el caso con emparejamiento aleatorio, en especial para el primer período, puesto que la desigualdad cae 5 y 3 puntos en los períodos 1 y 4 respectivamente.

Al igual que en la sección anterior, si se permite variar a la participación laboral femenina, calculando la tabla estandarizada de 1992-1996 (2008-2012) en base a las distribuciones marginales y participaciones laborales femeninas de 2008-2012 (1992-1996), el coeficiente de Gini presenta variaciones similares a su equivalente con emparejamiento aleatorio ya que se observan caídas en la desigualdad de 5 puntos para el período 1 y 4 para el período 4.

En conjunto todos los experimentos muestran que el rol del emparejamiento selectivo es muy importante, puesto que explica entre 4 y 5 puntos de la desigualdad de la distribución del ingreso, éstos resultados se encuentran en sintonía con los encontrados por Greenwood y col. (2014), para quienes el emparejamiento selectivo es una de las principales fuerzas determinantes de la mencionada desigualdad.

Éstos resultados son un aporte más a la discusión acerca del papel que cumple el emparejamiento selectivo en la desigualdad de la distribución del ingreso de los hogares, dónde en un lado se sitúan Greenwood y sus coautores, para los cuales la importancia de ésta es muy grande. Mientras que en el otro se sitúan Hryshko, Juhn y McCue (2014), para quienes el rol del emparejamiento selectivo es secundario relativo al cierre de la brecha salarial entre hombres y mujeres, de la misma forma, así como también Eika, Mogstad y Zafar (2014), quienes encuentran el mismo resultado para Estados Unidos y Noruega utilizando una metodología diferente. Las tablas de contingencia estandarizadas para cada período son las Tablas 4.10 y 4.11 respectivamente

Cuadro 3.4: Valores de los coeficientes de Gini para cada experimento contrafáctico

Experimento	1992-1996	2008-2012
Datos	0,398	0,382
Emp. aleatorio en 1	0,335	
Emp. aleatorio en 4		0,349
Emp. aleatorio en 1 y Δ en PLF	0,334	
Emp. aleatorio en 4 y Δ en PLF		0,341
Tabla est. en 1 con marg. de 4	0,346	
Tabla est. en 4 con marg. de 1		0,351
Tabla est. en 1 con marg. de 4 y Δ en PLF	0,345	
Tabla est. en 4 con marg. de 1 y Δ en PLF		0,342

Fuente: Elaboración propia en base a INDEC.

3.5. Cheques de robustez

El ejercicio de simulación depende del número de grupos en los que se subdivide a la población, tal como se mencionó con anterioridad, en este caso la población se separó en 5 quintiles en vez de 10 deciles, como lo hace el trabajo

de Greenwood y col. (2014) para evitar tener un gran número de grupos con cero individuos que no aporten nada al análisis.

Sin embargo, es válido preguntarse que sucedería si se modifica éste supuesto y, en segundo lugar, que sucedería si se mantiene el número de cuantiles, pero se reduce la cantidad de grupos. Para el primer experimento se adaptó el código de MATLAB (2013) para volver a tener 10 deciles, esto implica que el número de grupos será de $768 \times 10 = 7680$, el objetivo consiste en replicar la metodología original al pie de la letra para corregir por posibles errores en la modificación del código.

En segundo lugar se modifica el número de grupos, eliminando la variable número de hijos, de modo tal que el número de grupos se reduce a $128 \times 5 = 640$, con éstos dos chequeos puede verse cuan robusta es la metodología a variaciones en el número de grupos, debe destacarse que los autores del trabajo original no la realizan, en parte quizá por el hecho que disponen de varios millones de observaciones, lo cual no sucede en éste caso.

Para todos los chequeos se muestra el valor del índice de Gini para la totalidad de los experimentos en cada uno de los escenarios alternativos, dónde por período 1 se hace referencia 1992-1996 y por período 4 a 2008-2012:

Cuadro 3.5: Chequeos de robustez ante cambio en el número de grupos: coeficientes de Gini

Simulación	Cambio en deciles		Cambio en grupos	
	1992-1996	2008-2012	1992-1996	2008-2012
Experimento				
Datos	0,426	0,409	0,403	0,383
Emp. aleatorio en 1	0,352		0,389	
Emp. aleatorio en 4		0,368		0,400
Emp. aleatorio en 1 y Δ en PLF	0,350		0,386	
Emp. aleatorio en 4 y Δ en PLF		0,358		0,376
Tabla est. en 1 con marg. de 4	0,363		0,401	
Tabla est. en 4 con marg. de 1		0,370		0,408
Tabla est. en 1 con marg. de 4 y Δ en PLF	0,362		0,399	
Tabla est. en 4 con marg. de 1 y Δ en PLF		0,360		0,384

Fuente: Elaboración propia en base a INDEC.

Los chequeos muestran que hay algunas discrepancias al usar deciles en vez de quintiles, al igual que cuando se reduce el número de grupos, en el primer caso, la reducción en la desigualdad es de mayor magnitud, fruto de que el valor del índice de Gini se incrementa de 0,398 a 0,426 al utilizar deciles en vez de quintiles, éste resultado siembra algunas dudas sobre la exactitud del algoritmo de cómputo del coeficiente de Gini, puesto que los valores no deberían verse modificados de forma significativa ante ésta modificación. Lo mismo sucede al reducir el número de grupos, ya que esta modificación eleva el valor del coeficiente de Gini calculado a partir de los datos de 0,398 a 0,403.

Otro efecto llamativo es que la desigualdad aumenta en el segundo período al imponer tanto emparejamiento aleatorio como la matriz de asociación estandari-

zada en el caso con un menor número de grupos, mientras que éstos efectos se cancelan por completo al incluir las variaciones en la oferta laboral femenina.

Las variaciones en el coeficiente de Gini que surgen de utilizar los deciles en vez de los quintiles de la distribución del ingreso familiar pueden deberse a las siguientes causas, en primer lugar, puede ocurrir un problema en el algoritmo original o, en segundo lugar puede suceder que la modificación realizada para éste trabajo sea errónea; en cualquiera de los dos casos se requerirá un estudio más profundo de los códigos utilizados para generarlos, vale decir, reescribirlos para tener en cuenta el número de deciles como una variable independiente (“input”) de éstos.

3.6. Conclusión

Los resultados de ésta sección resultan ser contradictorios, puesto que, si se utilizan sólo indicadores de emparejamiento, éste apenas parece haber experimentado variaciones en los 21 años que se estudian, mientras que si se utiliza la metodología de micro-descomposición detallada anteriormente, la desigualdad cae entre 3 y 6 puntos.

Los patrones observados son consistentes con un nivel de emparejamiento relativamente constante en el tiempo pero cuyo impacto sobre la distribución del ingreso no está del todo claro, puesto que la metodología utilizada para medirlo posee problemas serios. Los resultados obtenidos se alinean con los de Harmenberg (2014) quién encuentra que el método aquí utilizado tiende a exagerar el impacto del emparejamiento selectivo en el coeficiente de Gini en comparación con otros métodos, aquí se agrega a ésta evidencia contra el método el hecho que éste también carece de robustez ante cambios en parámetros de gran importancia como ser el número de grupos o el de cuantiles en los que se divide a la distribución del ingreso. Además de la causa mencionada en la sección previa (problemas en el código), surgen una serie de motivos adicionales para justificar las diferencias con respecto a los resultados encontrados por otros autores, más allá de las obvias que surgen de utilizar datos para un país diferente:

Período de tiempo : dada la escasez de datos característica de las economías emergentes, para éste trabajo se optó por utilizar el período 1992-2012, ya que los datos de la EPH de la década de 1980 sufren de una serie de inconvenientes (problemas de medición de ingresos fruto de la elevada inflación, menor cobertura de agregados urbanos, etc.), lo que implica un período de sólo 20 años, además, el no disponer de datos censales obligó a juntar 5 años en un sólo período, reduciendo al ventana de análisis. Como consecuencia de esto, se redujo el numero de cuantiles, a fines de evitar que un gran número de grupos posean tamaño cero.

Metodología : Hryshko, Juhn y McCue (2014) y Eika, Mogstad y Zafar (2014) utilizan bases de encuestas en vez de censos y, al igual que aquí encuen-

tran variaciones en la desigualdad mucho más acotadas, a pesar de que el período de tiempo es mayor que el usado aquí, lo que lleva a postular como causa a los datos usados por Greenwood y col. (2014).

Motivos estructurales : esta causa es más bien una consecuencia de la primera, dado que los cambios en el emparejamiento selectivo son estructurales y no coyunturales, es de esperarse que no varíe significativamente año a año, tal como se observa en los indicadores descritos en la respectiva sección. Una explicación para este fenómeno es que los patrones de emparejamiento en Argentina son muy diferentes relativo a los de Estados Unidos, vale decir, que ya eran altos en el pasado y sólo crecieron levemente en el período estudiado, ésto explicaría la poca contribución a la desigualdad del emparejamiento selectivo.

Sea cual fuere la explicación, aún queda lugar para realizar estudios posteriores sobre el tema, ya que las metodologías de microdescomposición utilizadas no permiten separar las contribuciones a la desigualdad de cada una de las causas mencionadas con anterioridad (emparejamiento selectivo, fecundidad y brecha salarial), en éste sentido el trabajo de Eika, Mogstad y Zafar (2014) utiliza un método que, en un principio, podría adaptarse para responder éste interrogante utilizando la metodología de microdescomposición desarrollada por Firpo, Fortin y Lemieux (2009), la cual permitiría, separar las contribuciones mencionadas.

A la luz de la evidencia de ésta sección y la anterior también debe reconocerse que el emparejamiento selectivo no parecería ser un fenómeno que posea una importancia clave a la hora de influenciar la desigualdad de la distribución del ingreso total familiar. Éste resultado resulta ser bastante positivo desde la óptica de las políticas públicas, puesto que si éste no fuera el caso, es muy difícil generar variaciones en sus niveles por los canales usuales (subsidios y/o impuestos, cuotas, etcétera).

Capítulo 4

Emparejamiento selectivo y oferta laboral femenina

4.1. Introducción

En este capítulo se estudian las horas de trabajo promedio de las mujeres en Argentina durante el período 1992-2012, para ello se utilizará el modelo detallado en Bredemeier y Juessen (2013) con el cual se realiza un ejercicio de simulación basado en un conjunto de parámetros calibrados.

En Argentina se observa un patrón de comportamiento de las horas trabajadas por mujeres similar al de Estados Unidos, donde las mujeres casadas con hombres de mayor nivel salarial son las que experimentan un mayor crecimiento en sus horas de trabajo totales, aunque las variaciones que éstas experimentan son considerablemente inferiores a las de Estados Unidos. Las variaciones en la cantidad de horas trabajadas se explican por medio de variaciones en el emparejamiento selectivo, puesto que éstas son consistentes con un patrón de emparejamiento creciente, pero de pequeña magnitud.

El marco de análisis cuantitativo parte de las distribuciones observadas de salarios en la economía, y de la distribución conjunta de deciles de ingreso laboral de los cónyuges, la cual se utiliza para medir el nivel de emparejamiento selectivo. La evidencia muestra que el efecto del emparejamiento selectivo es relativamente bajo, especialmente en relación con el observado en Estados Unidos debido a que la variación en el emparejamiento selectivo fue también reducida.

4.2. Modelo

4.2.1. Datos y fuentes

Los datos que se utilizan para estimar el modelo provienen de la **Encuesta Permanentes de Hogares (EPH)** elaborada por el Instituto Nacional de Estadísticas (INDEC) de Argentina, tal como se detalla en Gasparini, Cicowiez y Sosa Escudero, 2013, pág. 719. El período abarca los años 1992-2012, aquí hay una importante desviación con respecto al trabajo original, que utiliza datos de la Current Population Survey desde 1970 a 2010, el motivo radica en que, en primer lugar, en la década de 1970 sólo se realizó una encuesta (en el año 1974) y, en segundo lugar, las encuestas de la década de 1980 poseen una serie de problemas como ser, cobertura geográfica limitada a pocos aglomerados y errores de imputación, por último, se decidió comenzar en 1992 para evitar problemas con la unidad monetaria, puesto que en dicho año entró en vigencia la actual moneda de curso legal (el peso).

A fines de deflactar los valores de los salarios nominales se utilizó la siguiente estrategia: para el período 1992-2006 se utilizó el IPC-INDEC mientras que para el período 2007-2012 se utilizó el IPC-BPP, calculado por el “Billion Prices Project” (BPP) del MIT¹ en base a datos on-line, es decir en base a un relevamiento automático de datos publicados en internet (Cavallo (2013)).

Los motivos para utilizar esta alternativa radican en que, utiliza una canasta similar a la del INDEC y que, si bien los precios se relevan en un número reducido de locales, el agregado se comporta de manera similar al IPC, como consecuencia de la **Ley de los grandes números**. El ingreso laboral se deflacta, entonces utilizando este índice empalmado con base 2012=100, es decir, todas las magnitudes están expresadas en pesos constantes de 2012.

Posteriormente, siguiendo a Bredemeier y Juessen (2013), se eliminan los hogares con un único miembro, el motivo reside en que las estimaciones del modelo requieren la existencia de dos cónyuges porque si hubiese sólo un miembro no existiría una distribución conjunta de ingresos por decil de la pareja, así como también a los hogares secundarios. Como paso final se eliminan todas las parejas en las que alguno de los cónyuges sea menor a 25 años o mayor de 60 años con el objeto de evitar captar a la menor cantidad posible de estudiantes y jubilados.

Un detalle adicional que cabe mencionar es que en este caso, a diferencia de los autores citados, no se eliminan a los individuos que poseen ingresos por cuenta propia motivado en que en las economías emergentes, como es el caso de Argentina, existe un alto grado de informalidad laboral, y estos trabajadores suelen ser considerados independientes por las estadísticas oficiales, fruto del tipo de vínculo que poseen con sus empleadores.

¹www.bpp.mit.edu

La base de datos se completa agregando a éstos en cuatro bloques a fines de evitar el efecto de los ciclos económicos sobre los salarios y las horas trabajadas²:

- Período 1: 1992-1996
- Período 2: 1997-2002
- Período 3: 2003-2007
- Período 4: 2008-2012

4.2.2. Análisis cuantitativo

Bredemeier y Juessen modelan el problema decisorio de las parejas como una optimización conjunta de una función de utilidad CES que depende del consumo de bienes privados c_i y del consumo agregado de bienes públicos d ; si $i = M, F$:

$$u_i = \left[\left(\frac{1}{1 + \psi} \right)^{1/\phi} c_i^{\frac{\phi}{\phi-1}} + \left(\frac{\psi}{1 + \psi} \right)^{1/\phi} d^{\frac{\phi}{\phi-1}} \right]^{\frac{\phi-1}{\phi}} \quad (4.1)$$

Dónde ψ es la ponderación relativa del consumo de bienes producidos por el hogar (d) y ϕ es la elasticidad de sustitución entre ambos bienes.

Los bienes producidos en el mercado se consiguen intercambiándolos por horas de trabajo (n_i) de cada uno de los cónyuges i , de modo que la restricción presupuestaria de la familia es:

$$c_F + c_M = w_F n_F + w_M n_M \quad (4.2)$$

Los bienes del hogar se producen por medio de una función Cobb-Douglas que utiliza como insumo las horas de trabajo que los cónyuges no ofrecen en el mercado (h_i):

$$d = h_M^\theta h_F^{1-\theta} \quad (4.3)$$

Finalmente, dado que la totalidad del tiempo puede asignarse al consumo en el hogar o fuera de él, la totalidad de horas (normalizadas) debe ser igual a 1:

$$h_i + n_i = 1 \quad i = M, F \quad (4.4)$$

El problema de optimización de las parejas consiste en maximizar (4.1) sujeto a (4.2), (4.3) y (4.4). La solución del problema dependerá del salario relativo de la pareja, para algunos valores sólo un cónyuge participa del mercado laboral y

²Aunque debe recalcar que el grueso del análisis se realizará en los períodos 1 y 4

para otros ambos participan. Si $w_F < w_M$ el hombre es el principal asalariado de la pareja, en el caso contrario, lo es la mujer, Bredemeier y Juessen muestran que la función de oferta de horas de trabajo en el mercado de las mujeres tienen la siguiente forma, para $\omega = w_F/w_M$:

$$n_F(\omega) = \begin{cases} 0 & \omega < \frac{\psi}{2 + \psi} \\ \frac{2}{2 + 2\psi} - \frac{\psi}{2 + 2\psi} \omega^{-1} & \frac{\psi}{2 + \psi} \leq \omega < \frac{2 + \psi}{\psi} \\ \frac{2}{2 + 2\psi} & \omega > \frac{2 + \psi}{\psi} \end{cases} \quad (4.5)$$

De esta ecuación se infiere que las horas de trabajo que las mujeres ofrecen en el mercado laboral son una función decreciente del salario de sus maridos, para un dado salario propio, y la pendiente de dicha relación es una función de la valuación que cada mujer le otorga al consumo en el hogar.

La simulación descansa en dos supuestos:

1. **Distribuciones marginales de salarios:** el soporte de las distribuciones marginales de los salarios se normaliza a 1, lo que implica que la brecha salarial es una diferencia absoluta constante entre los cuantiles de las distribuciones de salarios, la que se denota como α de manera tal que los salarios de las mujeres están distribuidos uniformemente en el intervalo $(w_{min}, w_{min} + 1)$ y los de los hombres en $(w_{min} + \alpha, w_{min} + \alpha + 1)$, además, cuando ambos cónyuges se ubican en el mismo decil de la distribución, ambos optan por trabajar, lo que implica que: $\frac{w_{min}}{w_{min} + \alpha} \geq \frac{2}{2 + \psi}$.
2. **Distribución conjunta de salarios:** el emparejamiento selectivo se modela suponiendo que una fracción ξ de los individuos se casa con una mujer de su mismo cuantil, mientras que el resto lo hace aleatoriamente, de modo que el emparejamiento perfecto y el aleatorio son casos especiales en los que $\xi = 1$ y $\xi = 0$.

El patrón de horas trabajadas por las mujeres es igual al total de horas de la población, dados los supuestos puede calcularse integrando a lo largo de toda la distribución condicional de salarios condicionales, la cual depende del emparejamiento selectivo:

$$\bar{n}_F(w_M) = \int n_F\left(\frac{w_F}{w_M}\right) f(w_F|w_M) dw_F = \Psi + (1 - \xi)a(w_M) + \xi b(w_M) \quad (4.6)$$

Donde $\bar{n}_F(w_M)$ son las horas trabajadas promedio por mujeres casadas con hombres cuyo salario es w_M y las funciones que resultan de evaluar la integral son.

$$\begin{aligned}\Psi &= \frac{2}{2+2\psi} + (1-\xi)\frac{2}{2+2\psi}w_{min} \\ a(w_M) &= -\frac{2}{2+2\psi} \left(\frac{2-\psi}{\psi} + \psi \log \frac{2+\psi}{\psi} \right) w_M \\ b(w_M) &= -\frac{\psi}{2+2\psi} \left(\frac{w_M}{w_M+\alpha} \right)\end{aligned}$$

Estas dos funciones son claves para entender como la distribución de horas trabajadas por mujeres según el decil de sus cónyuges depende del grado de emparejamiento selectivo, en primer lugar, $a'(w_M) < 0$ y $a''(w_M) = 0$ y luego, $b'(w_M) > 0$ y $b''(w_M) < 0$, por lo que a medida que se incrementa el salario, hay dos fuerzas contrapuestas que determinan la cantidad de horas trabajadas, una hace que las mujeres trabajen más a medida que se incrementa el salario de sus esposos y la otra tiene el efecto opuesto, el peso que tendrá cada función dependerá de la magnitud de ξ , por lo que el patrón de comportamiento de las horas depende del grado de emparejamiento selectivo. En la figura 4.1(a) puede verse que las horas trabajadas crecen en la medida que aumenta el salario de sus parejas, esto es consistente con un valor alto de ξ o sea, de un alto grado de emparejamiento selectivo.

4.2.3. Estimación de salario de mujeres

Para estimar el modelo es necesario calcular de forma apropiada los deciles de la distribución de los ingresos laborales de ambos cónyuges, para el caso de los maridos, ésto no es mayor problema, ya que su tasa de participación laboral es cercana al 90 % en la mayoría de los años, pero no ocurre lo mismo para las mujeres, quienes participan del mercado laboral en menos del 50 %, esto genera problemas para el cálculo de los deciles, debido a que una parte importante de ellas poseen ingresos laborales iguales a cero. Una forma de solucionarlo consiste en estimar un salario “contrafáctico” igual al que recibirían en el mercado laboral, si estas participaran en función a sus características observables.

Siguiendo de cerca nuevamente a Bredemeier y Juessen se replicará el análisis, el cual consiste en estimar una regresión de Heckman, también llamado “heckit” (Amemiya (1985)) la cual consta de dos subecuaciones, la primera, denominada “ecuación de participación” determina si la mujer participará o no en el mercado laboral, es esencialmente un modelo de clasificación logit que determina la participación (1) o no (0) en función a un conjunto de características: la cantidad de niños menores a 6 años de edad, la edad de la mujer y su cónyuge (la que se incluye como términos lineales y cuadráticos) y el nivel educativo, definida como una variable cualitativa que toma 7 valores posibles (desde no escolarizado hasta superior completo), de los cuales 6 se incluyen en la ecuación para evitar multicolinealidad perfecta.

Luego se estima la llamada “ecuación salarial” que no es otra cosa que una ecuación de Mincer (Mincer (1958)) la cual relaciona al logaritmo del salario con el logaritmo de los años de experiencia (medidos como los años pasados después de los 17), el cuadrado de dicha variable, los años de educación y un conjunto de dummies que representan los efectos fijos por cada año del período.

La estimación del modelo se hizo por medio del paquete “sampleSelection” (Toomet y Henningsen (2008)) el cual permite estimar estos modelos denominados “Tobit-tipo 2” (Amemiya (1985), pág. 385) ya sea por el método de dos pasos (estimar la ecuación de selección por logit, calcular la razón de Mills inversa y luego realizar una regresión por MCO) o bien estimar directamente por máxima verosimilitud, en este caso se optó por la primera forma, dada su simplicidad, por ser la forma escogida por los autores y a que los resultados no difieren significativamente entre un método y otro.

En el Cuadro 4.1 pueden observarse las estimaciones del modelo para los dos períodos considerados, en primer lugar, las ecuaciones de selección muestran que tener hijos en edad preescolar reduce la probabilidad de participar del mercado laboral y de una forma estadísticamente significativa, del mismo modo, el efecto de la experiencia de las mujeres es también significativo, tanto en su coeficiente lineal como en el cuadrático, por lo que el efecto tendrá un valor máximo en un punto fácilmente determinable, mientras que el efecto de la experiencia de su pareja es negativo y lineal.

Por último, el efecto de los niveles de educación propios sobre la probabilidad de participar del mercado laboral son todos positivos y significativos, no así con los de su pareja, cuyos efectos carecen de significancia estadística, salvo para las observaciones del primer período.

En el caso de la ecuación de resultado, la experiencia posee un efecto similar que en la ecuación anterior, ya que los coeficientes son significativos, siendo positivo el del término lineal y negativo el del cuadrático, siendo ambos significativos. A medida que aumenta el nivel educativo, se incrementa el valor del logaritmo del salario, tal como habría de esperarse en función de lo propuesto por Mincer.

4.2.4. Imputación de salarios

El ingreso familiar total puede desagregarse según sus fuentes en los retornos provenientes del factor trabajo y en ingresos provenientes del factor capital, donde “capital” debe entenderse en sentido amplio ya que incluye desde activos líquidos como dinero en efectivo hasta inversiones fijas como inmuebles. No obstante, es sabido que las encuestas de hogares (principal fuente de información de éste y la mayoría de los trabajos sobre distribución del ingreso con microdatos) no miden apropiadamente éstos retornos, de aquí que la gran mayoría de los estudios se han centrado en estudiar la desigualdad del ingreso laboral y, en particular la descomposición del ingreso laboral familiar en sus elementos.

Cuadro 4.1: Regresiones de Heckman

	<i>log (salario_M)</i>			
	1992-1996	1997-2002	2003-2007	2008-2012
Ecuaciones de Selección				
<i>log(edad_F/17)</i>	3,41*** (0,28)	3,54*** (0,22)	2,93*** (0,16)	3,20*** (0,14)
<i>log(edad_F/17)²</i>	-2,02*** (0,17)	-2,09*** (0,13)	-1,65*** (0,10)	-1,78*** (0,08)
<i>log(edad_M/17)</i>	-0,62 (0,30)	-0,95*** (0,23)	-0,13 (0,17)	-0,52*** (0,15)
<i>log(edad_M/17)²</i>	0,12 (0,18)	0,32** (0,14)	0,05 (0,10)	0,14 (0,09)
Pri. Inc. (F)	-0,02 (0,06)	0,10* (0,06)	0,07 (0,06)	0,10** (0,05)
Pri. Comp. (F)	-0,01 (0,06)	0,18*** (0,06)	0,14*** (0,05)	0,18*** (0,05)
Sec. Inc. (F)	0,09 (0,06)	0,22*** (0,06)	0,23*** (0,05)	0,27*** (0,05)
Sec. Comp. (F)	0,39*** (0,06)	0,46*** (0,06)	0,45*** (0,05)	0,46*** (0,05)
Sup. Inc. (F)	0,50*** (0,07)	0,64*** (0,06)	0,62*** (0,05)	0,60*** (0,05)
Sup. Comp. (F)	1,35*** (0,06)	1,35*** (0,06)	1,32*** (0,05)	1,37*** (0,05)
Pri. Inc. (M)	-0,15*** (0,06)	0,05 (0,06)	0,11** (0,05)	-0,004 (0,05)
Pri. Comp. (M)	-0,26*** (0,06)	-0,04 (0,06)	0,04 (0,05)	-0,01 (0,05)
Sec. Inc. (M)	-0,30*** (0,06)	-0,04 (0,06)	0,06 (0,05)	0,01 (0,05)
Sec. Comp. (M)	-0,40*** (0,07)	-0,13** (0,06)	0,03 (0,05)	-0,04 (0,05)
Sup. Inc. (M)	-0,28*** (0,07)	-0,06 (0,06)	0,16*** (0,05)	0,07 (0,05)
Sup. Comp. (M)	-0,40*** (0,07)	-0,19*** (0,06)	0,01 (0,05)	-0,01 (0,05)
Hijos2	-0,32*** (0,03)	-0,32*** (0,02)	-0,32*** (0,02)	-0,36*** (0,01)
Hijos3	-0,49*** (0,05)	-0,50*** (0,04)	-0,54*** (0,04)	-0,53*** (0,04)
Hijos4	-0,88*** (0,17)	-0,55*** (0,11)	-0,72*** (0,12)	-0,62*** (0,13)
Hijos5	-0,77 (0,55)	-0,98** (0,37)	0,11 (0,39)	-0,77* (0,41)
Constante	-1,15*** (0,12)	-1,32*** (0,11)	-1,57*** (0,09)	-1,40*** (0,08)
Ecuaciones de Resultado				
<i>log(edad_F/17)</i>	1,28*** (0,23)	1,51*** (0,20)	0,54*** (0,10)	0,63*** (0,12)
<i>log(edad_F/17)²</i>	-0,67*** (0,14)	-0,65*** (0,13)	-0,10* (0,09)	-0,18* (0,07)
Pri. Inc. (F)	-0,44*** (0,04)	0,17*** (0,06)	0,14** (0,06)	0,08 (0,05)
Pri. Comp. (F)	-0,18*** (0,04)	0,44*** (0,06)	0,41*** (0,06)	0,31*** (0,05)
Sec. Inc. (F)	-0,02 (0,04)	0,64*** (0,06)	0,57*** (0,06)	0,50*** (0,05)
Sec. Comp. (F)	0,30*** (0,04)	1,11*** (0,06)	1,03*** (0,06)	0,91*** (0,05)
Sup. Inc. (F)	0,44*** (0,04)	1,33*** (0,07)	1,19*** (0,06)	1,04*** (0,06)
Sup. Comp. (F)	0,57*** (0,05)	1,64*** (0,08)	1,44*** (0,07)	1,37*** (0,06)
Constante	6,83*** (0,14)	5,11*** (0,16)	6,36*** (0,13)	6,58*** (0,11)
Observaciones	36.790	60.435	112.278	140.403
ρ	0,14	0,38	-0,22	-0,17
Razón inversa de Mills (λ)	0,09 (0,06)	0,29*** (0,06)	-0,18*** (0,06)	-0,13** (0,05)

Nota: *p<0,1; **p<0,05; ***p<0,01

Una vez estimados los salarios de las mujeres, se realiza la misma estimación para el salario de los hombres, sólo que utilizando una regresión de Mincer idéntica a la ecuación de resultados de la sección anterior. Con estos dos resultados se procede a simular los ingresos de hombres y mujeres que reportan ingreso igual a cero, la metodología de trabajo es la misma que se describe en Gasparini, Cicowiez y Sosa Escudero, 2013, pág 547 y el objetivo es estimar los deciles de la distribución conjunta de los ingresos laborales de las parejas.

A partir de los coeficientes generados y el error cuadrático medio (RMSE) de la regresión se genera un vector de residuos independientes e idénticamente distribuidos como normal con media cero y varianza igual al cuadrado del RMSE, luego se predicen los resultados para los valores faltantes (esto es, para los que reportan salario igual a cero) y se les adiciona el residuo estimado. Por último, dado que la regresión es en logaritmos de salarios, el salario estimado se obtiene calculando el exponente de los valores estimados anteriormente.

Luego, con estos dos vectores de salarios estimados (uno para varones y otro para mujeres) se calculan los deciles con igual número de individuos, para ello se replicó en **R** (Team, 2013) la rutina de **Stata** "gcuan.do" descrita en Gasparini, Cicowiez y Sosa Escudero (2013) (pág 154), para cada período. Con estos deciles se calculan las distribuciones de horas trabajadas de mujeres por deciles del salario de sus parejas.

4.2.5. Obtención de datos para el modelo

La simulación del modelo de Bredemeier y Juessen requiere obtener para los cuatro períodos en estudio los siguientes insumos:

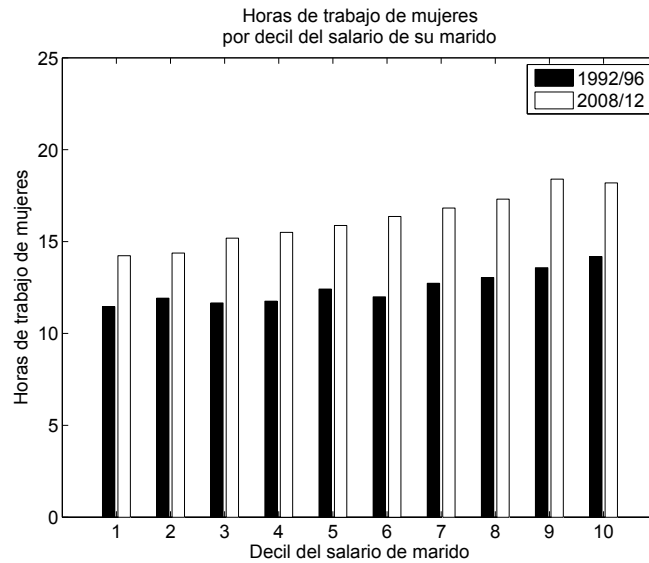
1. Horas trabajadas de hombres y mujeres según los deciles del ingreso de los hombres.
2. Salarios de hombres y mujeres promedio por sus respectivos deciles
3. Matrices de asociación entre los posiciones de los deciles de cada uno de los cónyuges (esto es, la distribución conjunto de cada uno de los deciles de cada miembro de la pareja)
4. Matrices de distribución del número de hijos menores de 18 años por deciles de cada uno de los padres
5. Coeficientes de correlación entre deciles de cónyuges

Los salarios por decil se obtienen calculando el salario promedio de hombres y mujeres por cada decil de su propia distribución. Por su parte las matrices de asociación se obtuvieron por medio de un simple programa que genera una matriz de dimensión 10×10 calculando las medias de cada uno de los 100 grupos de deciles de varones y mujeres.

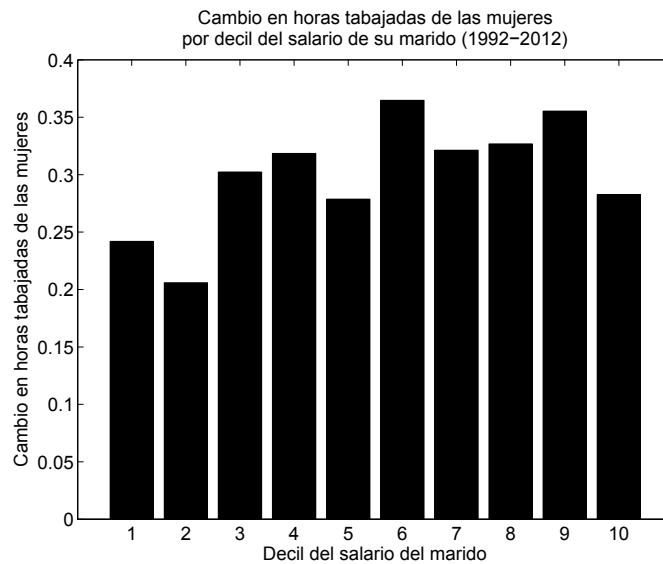
En la Figura 4.1 se muestran las horas trabajadas de las mujeres por decil de sus maridos en los dos períodos de tiempo extremos (1992-1996 y 2008-2012) y la variación de éstos³.

Se observa que, a diferencia de Estados Unidos, el incremento en horas trabajadas promedio por decil es uniforme a medida que se avanza en los deciles de los maridos y que, éstas variaciones son de una magnitud mucho menor que las reportadas para Estados Unidos por Bredemeier y Juessen o por Juhn y Murphy (1997). Este fenómeno también se ve reflejado en las modestas tasas de variación entre los dos períodos que, además, no poseen un patrón de comportamiento demasiado claro.

³Debe mencionarse que Bredemeier y Juessen no aclaran si para el cálculo del promedio de horas trabajadas por decil se tuvo en cuenta sólo a las mujeres que trabajan un número positivo de horas o al total como en el gráfico, en cualquier caso el patrón es el mismo, mientras que el promedio de horas trabajadas se duplica en el segundo caso.



(a) Horas trabajadas promedio de mujeres por decil de su cónyuge



(b) Variación en las horas trabajadas de mujeres por decil de su cónyuge

Figura 4.1: Horas semanales trabajadas por mujeres

Este patrón de comportamiento, es consistente con un alto grado de emparejamiento selectivo positivo, ya que implica que las mujeres que más horas trabajan son las que se casan con hombres de mayores ingresos y, tal como se vio en la sección anterior, el efecto que ésto tendría en la distribución del ingreso sería negativo, ya que incrementa la desigualdad medida por medio del coeficiente de Gini. Dicho resultado contrasta notablemente con el de Estados Unidos, donde los datos muestran que el emparejamiento selectivo creció en el tiempo, mientras que en Argentina puede verse que ya era alto en la década de 1990 y que se creció durante toda la década y en la primera mitad de la siguiente, para experimentar una leve caída posteriormente.

4.3. Simulación

4.3.1. Calibración

En esta sección se analizará en qué grado el modelo puede replicar horas trabajadas consistentes con las observadas en la realidad si se modifican los parámetros que éste utiliza como insumos⁴.

En primer lugar, el parámetro ψ , que mide el peso del consumo en el hogar en la función de utilidad, depende del número de hijos de la pareja (k):

$$\psi = \psi_0 + \psi_1 k^{\frac{\psi_2 - 1}{\psi_2}} \quad (4.7)$$

El primer parámetro a recalibrar son las horas totales semanales disponibles para el trabajo. En los datos se observa que la elevada prevalencia de mujeres que no trabajan reduce a casi la mitad el promedio de horas trabajadas, por lo que el stock de horas se reduce de 40 horas semanales a 30, lo que mejora el ajuste del modelo calibrado.

Cuadro 4.2: Valores de los parámetros

Parámetro	Interpretación	Valor
θ	Elasticidad de la producción hogareña (mujeres)	0,75
μ	Elasticidad de las participaciones en el consumo	1
ϕ	Elasticidad de sustitución	2
ψ_0	Valuación no condicional del consumo en el hogar	0,30
ψ_1	Efecto del primer hijo en la Valuación del cons. en el hogar	0,18
ψ_2	Curvatura del efecto de los hijos en el cons. del hogar	-0,90

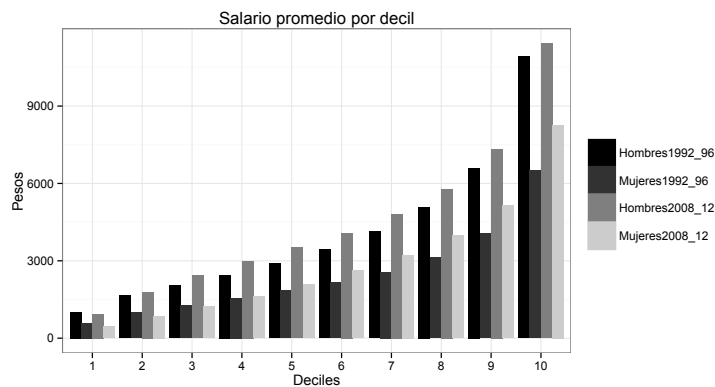
El valor del parámetro de la elasticidad de la producción hogareña para las mujeres es el mismo que en el trabajo original, debido a que no se poseen estimaciones de las horas trabajadas en el hogar en la base de datos, lo mismo ocurre con la elasticidad de las participaciones en el consumo, la que determina la fuerza con la que las participaciones en el consumo reacciona a los cambios en los salarios y se define como:

$$\frac{c_i}{c_M + c_F} = \frac{w_i^\mu}{w_M^\mu + w_F^\mu} \quad (4.8)$$

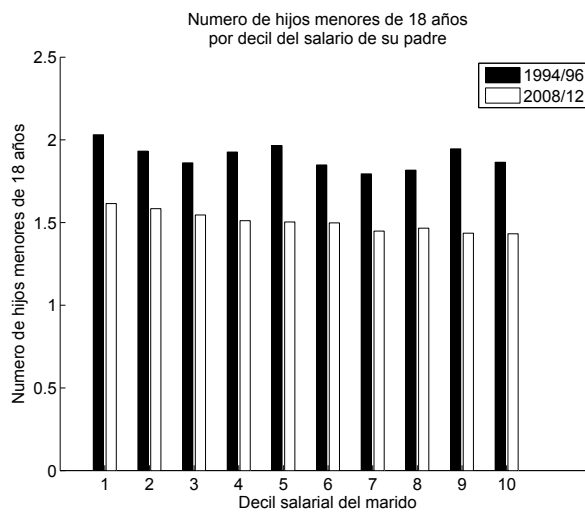
El valor de ϕ se fija en 2, ya que los datos son consistentes con una mayor elasticidad de sustitución entre consumo en el hogar y en el mercado, éste parámetro se obtuvo por modificaciones sucesivas en el rango aceptable [1, 57; 4].

⁴Para realizar la simulación se utilizó el código en **MATLAB** (MATLAB (2013)) provisto por Falko Juessen, donde se tuvo que recalibrar algunos de los parámetros para adaptarlos a los datos de Argentina.

La valuación del consumo en el hogar depende del número de hijos en el hogar, por ello es que requiere como insumo la matriz de asociación del número de hijos por decil de cada uno de sus padres, se utilizará nuevamente la parametrización original. El primer parámetro, ψ_0 , indica el valor que tiene el consumo en el hogar para las parejas sin hijos; el segundo, ψ_1 indica el efecto que tiene el primer hijo sobre el mencionado consumo mientras que el último, ψ_2 muestra cómo cada hijo adicional modifica la valuación del trabajo en el hogar.



(a) Distribuciones marginales de salarios



(b) Distribución de hijos menores de 18 años

Figura 4.2: Horas semanales trabajadas por mujeres

4.3.2. Datos de entrada

En función a lo mencionado en la sección anterior, el modelo requiere como insumos a las distribuciones salariales marginales (W), la matriz de asociación entre las posiciones de deciles de los cónyuges (S) y la matriz de asociación del número de hijos por cada decil de los cónyuges (K), el patrón de horas trabajadas por mujeres según el decil de sus parejas es una función que depende de dichas variables: $\bar{n}_F(q_M | \{W, S, K\})$.

La matriz de asociación es una matriz de dimensión 10×10 que mide la asociación en términos del número de parejas que existen entre los diferentes deciles de la distribución específica de cada pareja, cada una de las 100 celdas mide la frecuencia relativa de cada una de estas posibles combinaciones. En símbolos, cada elemento s_{ij} es igual a la frecuencia conjunta de parejas en las que el hombre pertenece al decil i y la mujer al decil j . Las Tablas 4.3 y 4.4 son éstas matrices de asociación para los períodos inicial y final respectivamente.

En el Cuadro 4.5, en el anexo, se muestran las diferencias porcentuales entre cada elemento de las matrices de asociación para cada período, siendo estas diferencias porcentuales iguales a $[(s_{ij}^{2008-12} - s_{ij}^{1992-96})/s_{ij}^{1992-96}] \times 100$.

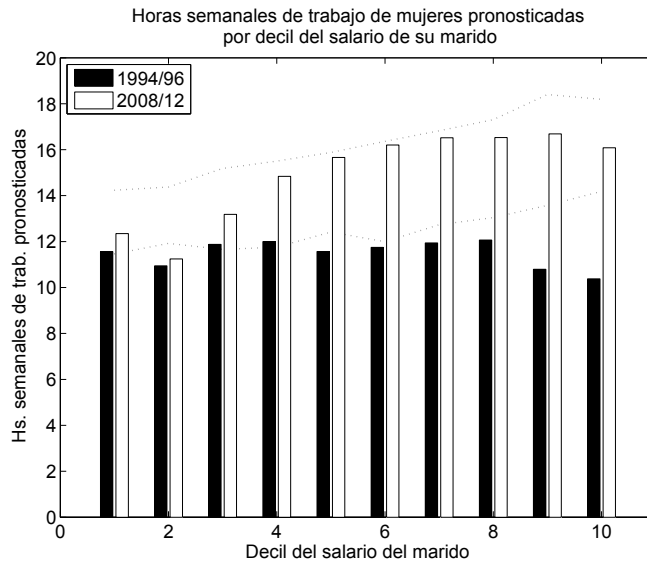
En el cuadro se observa una caída en las celdas donde hay una diferencia importante en las posiciones de deciles de ambos cónyuges, mientras que los mayores incrementos se ubican en la diagonal principal y en las celdas aledañas. Esto es otra evidencia a favor de un incremento en el emparejamiento selectivo entre los períodos 1992-1996 y 2008-2012.

En el gráfico 4.2(a) se muestran los niveles de salarios de mujeres y hombres por cada decil para los dos períodos extremos de la muestra. Éstos se denotan como $w_M(q)$ y $w_F(q)$, los salarios medios, medidos a moneda constante, experimentaron un notorio incremento, en ambos casos, el incremento porcentual es mayor cuanto más alto es el decil; en el caso de las mujeres, el cambio no es uniforme, porque los cuantiles más bajos o bien no han experimentado cambios significativos o bien tuvieron una leve caída, mientras que a medida que subimos en los deciles, el incremento toma fuerza.

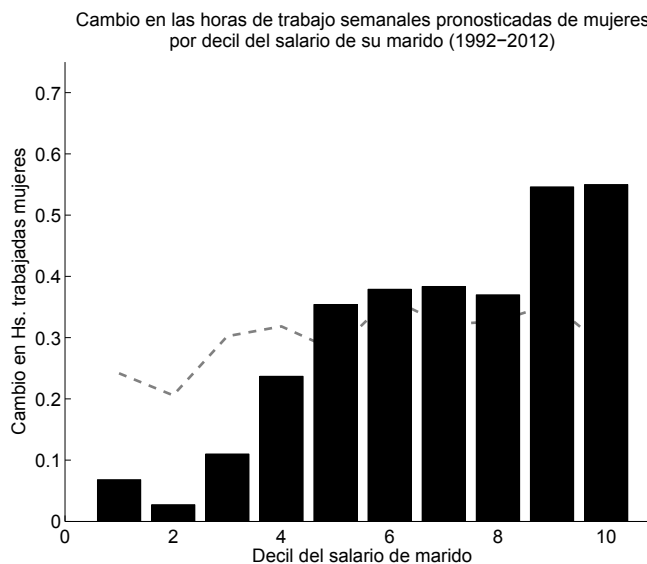
En el gráfico 4.2(b) se muestra la distribución del número de hijos por decil del padre, puede verse una caída en la cantidad promedio de hijos, aunque tampoco uniforme, porque todos los deciles llegan a igualar casi el promedio, mientras que en el período base existía una notoria diferencia. El número de hijos según la combinación de deciles de sus padres se denota como K , una matriz de 10×10 .

4.3.3. Resultados

Una vez calculadas las variables que utilizará el modelo: $\{W, S, K\}$ se calcula la oferta estimada de trabajo femenino para cada una de las 100 combinaciones posibles de deciles de la matriz H , definida como la matriz de horas trabajadas por las mujeres. Las horas promedio trabajadas por las mujeres se obtienen por medio de la fórmula: $(h_i \times s'_i) / \sum_{j=1}^{10} s_{ij}$ donde h_i es la i -ésima fila de la matriz H y s_i es la correspondiente fila de la matriz de asociación S .



(a) Horas trabajadas pronosticadas y efectivas



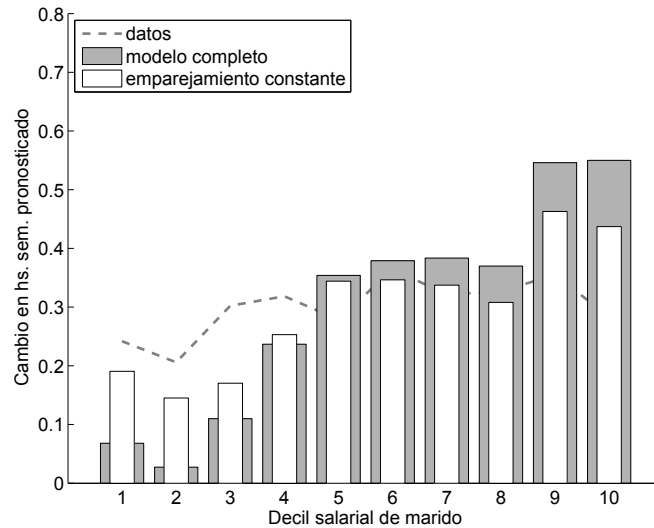
(b) Cambio en horas pronosticadas por decil

Figura 4.3: Horas semanales trabajadas por mujeres

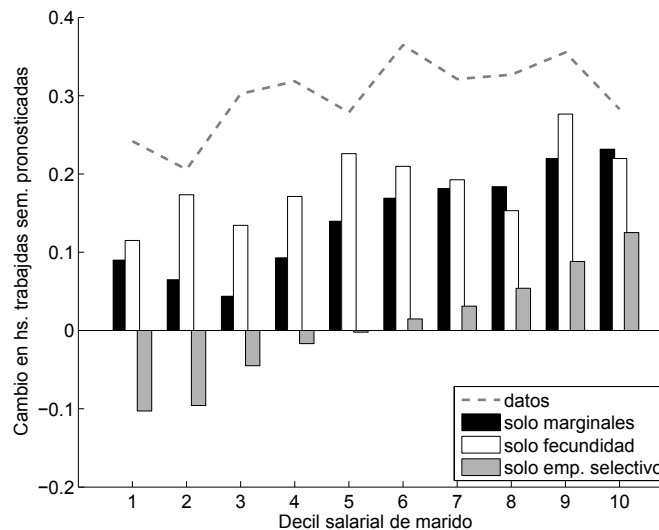
En la Figura 4.3 se muestra la comparación entre las distribuciones de horas trabajadas empírica (líneas punteadas) y estimada (barras), las que corresponden a $\bar{n}_F(q_M|W_{1992-96}, S_{1992-96}, K_{1992-96})$ y a $\bar{n}_F(q_M|W_{2008-12}, S_{2008-12}, K_{2008-12})$ puede verse que el ajuste es relativamente bueno, a pesar que la diferencia entre los valores pronosticados y teóricos tiende a crecer a medida que se incrementan los deciles.

El modelo logra reproducir el patrón creciente de horas trabajadas por decil del marido en el segundo período, pero no en el primero, puesto que las horas para los deciles más altos tienden a disminuir en vez de aumentar, sin embargo la diferencia no es cuantitativamente significativa, salvo en los deciles 9 y 10 donde la diferencia llega al 20 % del total.

La siguiente etapa consiste en replicar los resultados de los experimentos contrafácticos de los autores, el primero consiste en mantener constante la matriz de asociación en ambos períodos e igual a la de 1992-1996 mientras que las distribuciones marginales de salarios varían entre períodos con el objeto de cancelar cualquier tipo de tendencia en el emparejamiento selectivo.



(a) Asociación constante vs. variante



(b) Separación de efectos

Figura 4.4: Experimentos contrafácticos

En el gráfico 4.4(a) se muestra la distribución que resulta de mantener constante la matriz de asociación (barras blancas) contra la resultante de utilizar el modelo completo (barras grises) y la distribución empírica (línea punteada). Cuando se mantiene constante la asociación, se pronostica una variación en las horas trabajadas mucho más fuerte en los deciles bajos y mucho más débil en los deciles altos, esto quiere decir que los cambios en horas trabajadas más importantes los experimentaron las mujeres casadas con hombres cuyos ingresos

son más altos, cabe notar que este fenómeno también se observa en los datos originales de Estados Unidos.

El gráfico 4.4(b) separa los cambios en las horas trabajadas en sus tres componentes principales, las barras de color azul muestran los cambios que surgen sólo de las distribuciones marginales de salarios, la forma de éstas muestra que la caída en la brecha salarial de género contribuye al incremento en las horas trabajadas de las mujeres, pero dicho incremento no es lineal, su tasa de variación es decreciente en los deciles 1 a 3 y creciente en los restantes.

Las columnas verdes muestran la contribución de la caída en la fecundidad sobre las horas trabajadas, las cuales surgen de mantener la distribución de hijos del primer período y cambiando la matriz de asociación y las distribuciones marginales de salarios, claramente, a menor cantidad de hijos, más probable que una mujer ofrezca sus horas de trabajo en el mercado, éstas también muestran un patrón de comportamiento no lineal con picos en los deciles 2, 6 y 9.

Finalmente, las columnas de color blanco muestran el efecto del emparejamiento selectivo, éstas surgen de cambiar la matriz de asociación manteniendo constantes las distribuciones marginales de salarios. Dichas barras poseen un patrón de comportamiento lineal, negativo en los deciles más bajos y positivo en los más altos, pero su efecto sobre las horas trabajadas de las mujeres no es tan importante como las dos anteriores.

El gráfico permite llegar a una conclusión similar a la del trabajo original, el incremento en horas trabajadas por las mujeres se debe en su mayor parte a la reducción en la fecundidad y a la caída en la brecha salarial en mucha mayor medida que al emparejamiento selectivo, sin embargo es el emparejamiento selectivo el principal motivo por el cual se incrementaron las horas semanales trabajadas por las mujeres casadas con hombres de mayores ingresos.

4.4. Conclusión

En este trabajo se adaptó el modelo de Bredemeier y Juessen (2013) a los datos de Argentina, salvando las dificultades que estos presentan, como ser disponer de sólo 20 años de datos y la variabilidad de la disponibilidad de éstos. El objetivo consistió en determinar cuál es el grado de importancia del emparejamiento selectivo en la desigualdad de la distribución del ingreso laboral, para ello se estudió el comportamiento de las horas trabajadas por mujeres según el decil salarial de sus parejas, ya que es una de las formas de medir el emparejamiento selectivo (la otra es por medio de los niveles educativos de los cónyuges).

Los resultados muestran un patrón de comportamiento de las horas trabajadas compatible con un alto grado de emparejamiento selectivo por parte de las mujeres, al igual que los datos de Estados Unidos, pero con la importante diferencia que en vez de crecer en el tiempo, se mantiene en un nivel alto durante todo el período estudiado.

Como parte del trabajo se realizaron un conjunto de simulaciones con el objetivo de plantear distintos escenarios contrafácticos en el que se mantuvieron constantes las distribuciones marginales de salarios, la matriz de asociación y el número de hijos entre los períodos inicial y final. El resultado obtenido muestra que la caída en la fecundidad y el incremento salarial son los principales causantes del aumento en las horas trabajadas, pero no así el emparejamiento selectivo, debido a que éste no se incrementó significativamente, mientras que los otros si lo hicieron.

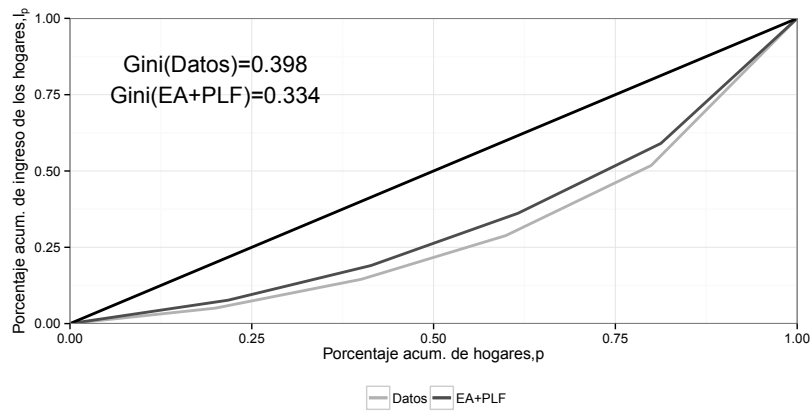
Bibliografía

- Abrevaya, Jason (1999). «Computation of the maximum rank correlation estimator». En: *Economics Letters* 62, págs. 279-285.
- Amemiya, Takeshi (1985). *Advanced Econometrics*. Inglés. 1.^a ed. Cambridge, Massachussets: Harvard University Press. ISBN: 0-674-00560-0.
- Beccaria, Luis y Fernando Groisman (2008). «Informalidad y Pobreza en Argentina». En: *Investigación Económica* 67.266.
- Becker, Gary (1991). *A Treatise on the Family*. Inglés. 1.^a ed. Cambridge, Massachussets: Harvard University Press. ISBN: 0-674-90699-3.
- Bergés, Miriam (2011). «Escala de equivalencias en el consumo para Argentina». Tesis doct. Universidad Nacional de La Plata.
- Bredemeier, Christian y Falko Juessen (2013). «Assortative mating and female labor supply». En: *Journal of Labor Economics* 31.3. URL: <http://www.jstor.org/stable/10.1086/669820>.
- Campos-Vázquez, Raymundo, Andrés Hicapié y Rubén Rojas-Valdéz (2012). «Family Income Inequality and the Role of Married Females' Earnings in Mexico: 1988-2010». En: *Latin American Journal of Economics* 49.1, págs. 67-98.
- Cancian, Maria y Deborah Reed (1999). «The Impact of Wives' Earnings on Income Inequality: Issues and Estimates». En: *Demography* 36.2, págs. 173-184.
- Cavallo, Alberto (2013). «Online and official price indexes: Measuring Argentina's inflation». En: *Journal of Monetary Economics* 60.2, págs. 152-165. ISSN: 03043932. DOI: 10.1016/j.jmoneco.2012.10.002. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0304393212000967>.
- Cowell, Frank A. y Emmanuel Flachaire (2015). «Chapter 6 - Statistical Methods for Distributional Analysis». En: *Handbook of Income Distribution*. Ed. por Anthony B. Atkinson y François Bourguignon. Vol. 2. Handbook of Income Distribution. Elsevier, págs. 359 -465. DOI: <http://dx.doi.org/10.1016/B978-0-444-59428-0.00007-2>. URL: <http://www.sciencedirect.com/science/article/pii/B9780444594280000072>.
- Daly, Mary C. y Robert G. Valletta (2006). «Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour». En: *Economica* 73.289, págs. 75-98.
- DiNardo, John, Nicole Fortin y Thomas Lemieux (1996). «Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach». En: *Econometrica* 64.5, págs. 1001-1044.
- Eika, Lasse, Magne Mogstad y Basit Zafar (2014). *Educational Assortative Mating and Household Income Inequality*. Working Paper 20271. National Bureau of

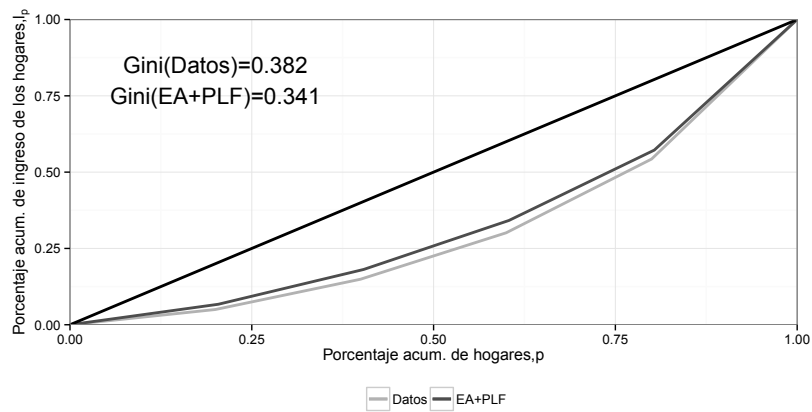
- Economic Research. DOI: 10.3386/w20271. URL: <http://www.nber.org/papers/w20271>.
- Firpo, Sergio, Nicole M. Fortin y Thomas Lemieux (2009). «Unconditional Quantile Regressions». En: *Econometrica* 77.3, págs. 953-973. ISSN: 1468-0262. DOI: 10.3982/ECTA6822. URL: <http://dx.doi.org/10.3982/ECTA6822>.
- Fortin, Nicole, Thomas Lemieux y Sergio Firpo (2011). «Chaper 1-Decomposition Methods in Economics». En: *Handbook of Labor Economics*. Ed. por David Card y Orley Ashenfelter. Vol. 4A. Elsevier, págs. 1-102.
- Gasparini, Leonardo, Martín Cicowiez y Walter Sosa Escudero (2013). *Pobreza y Desigualdad en América Latina. Conceptos, herramientas y aplicaciones*. Español. 1.^a ed. Buenos Aires: Temas. ISBN: 978-987-1826-45-2. URL: <http://www.depeco.econo.unlp.edu.ar/cedlas/libro-gcse-1/>.
- Gasparini, Leonardo y col. (2015). *Female Labor Participation in Latin America: Evidence of Deceleration*. Documento de Trabajo 181. CEDLAS, Universidad Nacional de La Plata. URL: http://cedlas.econo.unlp.edu.ar/download.php?file=archivos_upload/doc_cedlas181.pdf.
- Greenwood, Jeremy y col. (2014). «Marry Your Like: Assortative Mating and Income Inequality». En: *American Economic Review* 104.5, págs. 348-53. DOI: 10.1257/aer.104.5.348. URL: <http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.5.348>.
- (2015). «Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment and Married Female Labor-Force Participation». En: *American Economic Journal: Macroeconomics*.
- Harmenberg, Karl (2014). «A Note: The Effect of Assortative Mating on Income Inequality». Instituto de Estudios Económicos Internacionales-Universidad de Estocolmo.
- Hryshko, Dmytro, Chinhui Juhn y Kristin McCue (2014). *Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important is Assortative Matching?* Inf. téc. 8729. Institute for the Study of Labor (IZA).
- James, Gareth y col. (2013). *An Introduction to Statistical Learning with applications in R*. New York: Springer-Verlag.
- Juhn, Chinhui y Kevin Murphy (1997). «Wage Inequality and Family Labor Supply». En: *Journal of Labor Economics* 15.3. URL: <http://www.jstor.org/stable/2535315>.
- Kendall, Maurice G. (1970). *Rank Correlation Methods*. 4.^a ed. Londres: Charles Griffin y Co.
- Killingsworth, Mark y James Heckman (1986). «Female Labor Supply, A Survey». En: *Handbook of Labor Economics*. Ed. por Robert Layard y Orley Ashenfelter. Vol. 1. Elsevier Science Publishers, págs. 103-204.
- Lam, David (1988). «Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications». En: *Journal of Human Resources* 23.4, págs. 462-487.
- MATLAB (2013). *version 8.10.0.604 (R2013a)*. Natick, Massachusetts: The MathWorks Inc.

- Mincer, Jacob (1958). «Investment in Human Capital and Personal Income Distribution». En: *Journal of Political Economy* 66.4, págs. 281-302. URL: <http://www.jstor.org/stable/1827422>.
- Mosteller, Frederick (1968). «Association and Estimation in Contingency Tables». En: *Journal of the American Statistical Association*, 63.321, págs. 1-28.
- Pagan, Adrian y Amman Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press.
- Rao, V. M. (1969). «Two Decompositions of Concentration Ratio». En: *Journal of the Royal Statistical Society. Series A (General)* 132.3, págs. 418-425.
- Schwartz, Christine R. (2010). «Earnings Inequality and the Changing Association between Spouses' Earnings». En: *American Journal of Sociology* 115.5, págs. 1524-1557.
- Team, R Core (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org/>.
- Toomet, Ott y Arne Henningsen (2008). «Sample Selection Models in R : Package sampleSelection». En: *Journal of Statistical Software* 27.7.
- Wickham, Hadley (2011). «The Split-Apply-Combine Strategy for Data Analysis». En: *Journal of Statistical Software* 40.1, págs. 1-29. ISSN: 1548-7660. URL: <http://www.jstatsoft.org/v40/i01>.
- (2014). «Tidy Data». En: *Journal of Statistical Software* 59.10, págs. 1-23. ISSN: 1548-7660. URL: <http://www.jstatsoft.org/v59/i10>.
- Worner, Shane Mathew (2006). *The Effects of Assortative Mating on Income Inequality: A Decompositional Analysis*. CEPR Discussion Papers 538. Centre for Economic Policy Research, Research School of Economics, Australian National University. URL: <http://ideas.repec.org/p/auu/dpaper/538.html>.

4.5. Anexo: gráficos

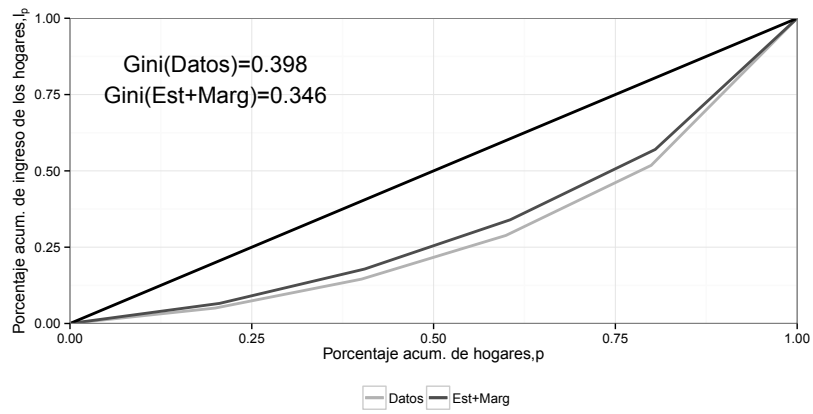


(a) Período 1992-1996

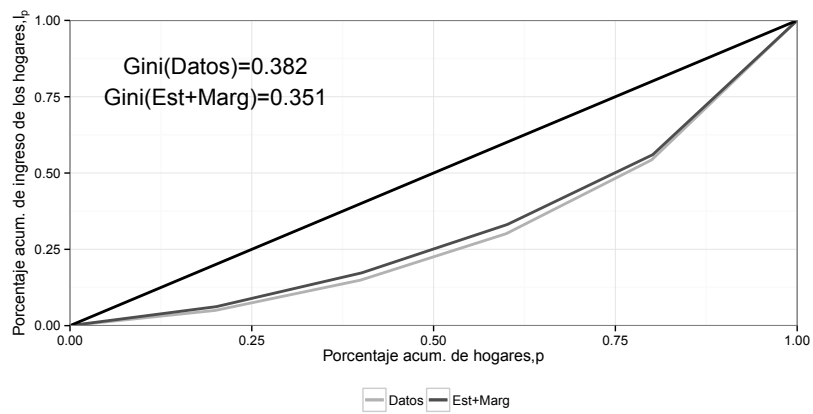


(b) Período 2008-2012

Figura 4.5: Curvas de Lorenz: Datos versus emparejamiento aleatorio en y cambios en la participación laboral femenina

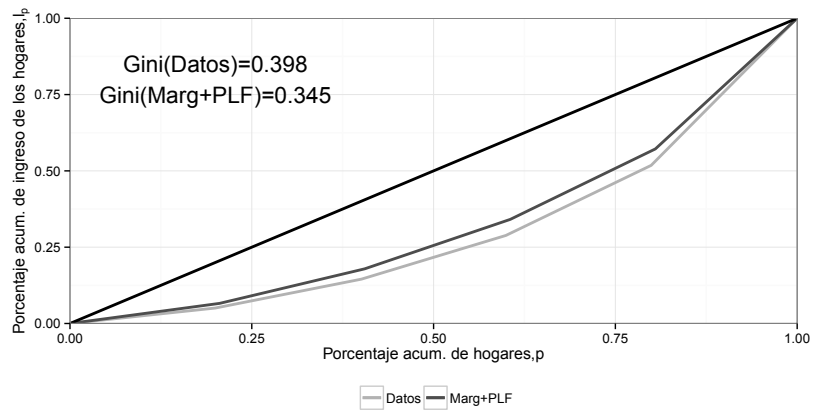


(a) Período 1992-1996

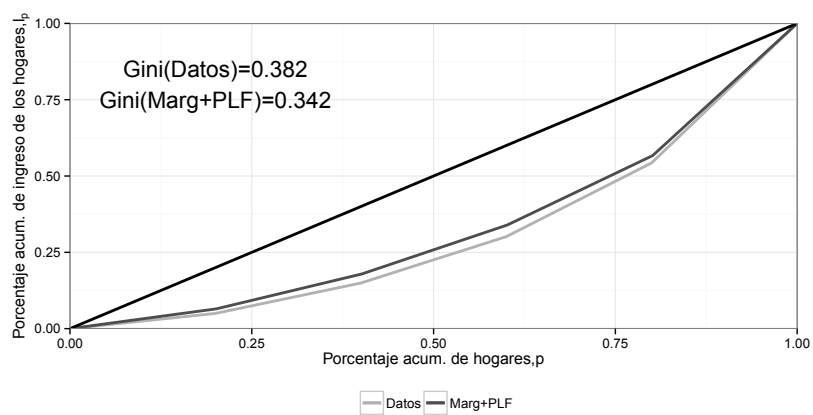


(b) Período 2008-2012

Figura 4.6: Curvas de Lorenz: Datos versus estandarización de matriz de asociación y distribuciones marginales del período opuesto



(a) Período 1992-1996



(b) Período 2008-2012

Figura 4.7: Curvas de Lorenz: Datos versus estandarización de matriz de asociación, distribuciones marginales del período opuesto y variación en la participación laboral femenina

4.6. Anexo: cuadros

Cuadro 4.3: Tabla de asociación por deciles para 1992-1996

Marido	Mujer									
	1	2	3	4	5	6	7	8	9	10
1	1,805	1,432	1,392	1,185	0,951	0,843	0,777	0,631	0,587	0,397
2	1,343	1,305	1,193	1,381	1,036	0,995	0,813	0,737	0,693	0,506
3	1,242	1,196	1,191	1,182	1,085	1,185	0,908	0,777	0,701	0,533
4	1,185	1,128	1,057	1,090	1,079	1,060	1,044	0,957	0,750	0,650
5	0,968	1,017	1,063	1,000	1,150	1,060	1,136	1,120	0,862	0,625
6	0,951	0,984	0,976	1,038	1,022	1,085	1,017	1,196	0,976	0,756
7	0,704	0,932	0,878	0,878	1,011	0,905	1,066	1,212	1,370	1,044
8	0,756	0,794	0,829	0,867	0,998	1,027	1,161	1,109	1,223	1,237
9	0,582	0,693	0,739	0,745	0,921	0,973	1,036	1,147	1,329	1,835
10	0,465	0,519	0,682	0,633	0,747	0,867	1,044	1,114	1,509	2,419

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.4: Tabla de asociación por deciles para 2008-2012

Marido	Mujer									
	1	2	3	4	5	6	7	8	9	10
1	1,956	1,632	1,403	1,070	0,991	0,801	0,667	0,556	0,494	0,429
2	1,500	1,399	1,293	1,326	1,244	0,876	0,771	0,605	0,546	0,441
3	1,254	1,214	1,194	1,099	1,211	1,243	0,880	0,769	0,600	0,536
4	1,112	1,169	1,113	1,117	1,083	1,150	1,204	0,875	0,654	0,523
5	1,003	1,024	1,014	1,089	1,041	1,090	1,234	1,028	0,814	0,662
6	0,822	0,887	0,960	1,006	1,011	1,036	1,155	1,375	0,944	0,803
7	0,729	0,835	0,864	0,955	0,935	1,043	1,115	1,202	1,377	0,945
8	0,648	0,702	0,787	0,860	0,890	0,984	1,085	1,223	1,491	1,330
9	0,532	0,622	0,737	0,771	0,853	0,934	0,949	1,207	1,498	1,895
10	0,444	0,516	0,634	0,706	0,739	0,843	0,939	1,159	1,583	2,437

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.5: Cambio porcentual en las matrices de asociación entre los períodos 1992-1996 y 2008-2012

Marido	Mujer									
	1	2	3	4	5	6	7	8	9	10
1	37,6	32,4	6,5	13,5	16,1	-4,1	-1,5	-24,3	-5,5	-12,6
2	20,3	14,8	2,0	-6,9	0,7	-5,5	-24,3	-20,8	-28,6	-20,3
3	7,5	7,7	8,5	10,0	15,3	15,0	2,8	-7,0	-17,4	-16,9
4	24,8	10,0	14,3	18,4	2,4	11,8	-0,9	-17,5	-23,8	-9,1
5	-2,6	-2,0	-0,2	-6,0	-7,9	10,2	23,8	-3,8	-18,1	-5,8
6	-3,2	17,9	-4,5	-11,0	1,5	-5,3	-0,7	24,6	-15,7	-15,4
7	-16,9	-4,5	-15,8	-11,1	-8,8	-5,8	-6,0	0,3	-3,5	-15,8
8	-9,4	-9,2	-13,3	-10,9	-9,2	-0,5	-12,8	-9,7	20,9	-3,1
9	-17,2	-17,5	-3,0	-7,9	-11,6	-10,6	-4,7	9,0	13,0	5,4
10	-6,2	2,4	11,5	7,8	1,6	-3,9	-12,5	7,5	5,1	8,4

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.6: Tabla de contingencia para 1992-1996 con emparejamiento aleatorio

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,094	0,031	0,068	0,059	0,044	0,020
P	0,030	0,010	0,021	0,018	0,014	0,006
S-	0,058	0,019	0,042	0,036	0,027	0,012
S	0,063	0,021	0,045	0,039	0,029	0,014
U-	0,032	0,011	0,023	0,020	0,015	0,007
U	0,022	0,007	0,016	0,014	0,010	0,005

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.7: Tabla de contingencia para 2008-2012 con emparejamiento aleatorio

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,041	0,010	0,056	0,035	0,054	0,025
P	0,010	0,002	0,013	0,008	0,013	0,006
S-	0,050	0,012	0,067	0,042	0,065	0,030
S	0,035	0,008	0,047	0,029	0,045	0,021
U-	0,031	0,008	0,042	0,026	0,040	0,019
U	0,020	0,005	0,027	0,017	0,026	0,012

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.8: Tabla de contingencia para 1992-1996 con emparejamiento aleatorio y PLF de 2008-2012

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,068	0,039	0,018	0,032	0,004	0,006
P	0,037	0,105	0,005	0,017	0,001	0,002
S-	0,018	0,006	0,058	0,032	0,028	0,025
S	0,035	0,015	0,030	0,056	0,012	0,018
U-	0,003	0,001	0,027	0,009	0,084	0,044
U	0,006	0,001	0,029	0,021	0,039	0,072

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.9: Tabla de contingencia para 2008-2012 con emparejamiento aleatorio y PLF de 1992-1996

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,071	0,036	0,019	0,031	0,004	0,006
P	0,037	0,102	0,006	0,020	0,001	0,002
S-	0,020	0,008	0,061	0,028	0,024	0,025
S	0,030	0,017	0,029	0,063	0,010	0,019
U-	0,004	0,001	0,025	0,009	0,087	0,041
U	0,005	0,002	0,027	0,016	0,041	0,075

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.10: Tabla de contingencia estandarizada para 1992-1996

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,101	0,019	0,039	0,043	0,011	0,008
P	0,020	0,018	0,004	0,008	0,001	0,001
S-	0,023	0,003	0,110	0,037	0,065	0,028
S	0,038	0,005	0,048	0,054	0,023	0,017
U-	0,002	0,000	0,027	0,006	0,105	0,026
U	0,003	0,000	0,023	0,010	0,038	0,034

Fuente: Elaboración propia en base a INDEC.

Cuadro 4.11: Tabla de contingencia estandarizada para 2008-2012

Marido	Mujer					
	P-	P	S-	S	U-	U
P-	0,172	0,037	0,041	0,053	0,008	0,005
P	0,037	0,043	0,005	0,013	0,001	0,001
S-	0,030	0,005	0,085	0,030	0,029	0,014
S	0,051	0,012	0,045	0,076	0,013	0,012
U-	0,004	0,000	0,021	0,005	0,064	0,014
U	0,004	0,001	0,017	0,008	0,023	0,020

Fuente: Elaboración propia en base a INDEC.