

Anales del VII CONGRESO NACIONAL DE ESTUDIANTES DE POSTGRADO EN ECONOMÍA (CNEPE)

*DEPARTAMENTO DE ECONOMÍA
UNIVERSIDAD NACIONAL DEL SUR*

*INSTITUTO DE INVESTIGACIONES ECONÓMICAS Y SOCIALES DEL SUR (IIESS)
CONICET - UNIVERSIDAD NACIONAL DEL SUR*

Bahía Blanca

Mayo de 2015

ISBN: 978-987-1648-39-9



Departamento de Economía



I I E S S

Emparejamiento selectivo y distribución del ingreso en Argentina.

Funes Leal, Víctor Exequiel

Emparejamiento selectivo y distribución del ingreso en Argentina

Víctor Funes Leal

(Universidad Nacional de La Plata)

victorefunes@gmail.com

Códigos JEL: D31, C15, J12

Resumen

En este trabajo se busca evaluar cuál es el impacto del emparejamiento selectivo sobre dos variables vinculadas de forma muy cercana: la oferta laboral femenina y la desigualdad en la distribución del ingreso familiar. Para ello se utiliza la metodología de Greenwood y col. (2014) con el objeto de descomponer el impacto de una mayor selectividad en la formación de parejas, vía deciles de ingreso laboral en el primer caso y niveles educativos en el segundo sobre las horas de trabajo ofrecidas en el mercado y sobre la desigualdad medida a través del índice de Gini.

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.

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, tanto de ellas como de los hombres, 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, donde 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. Este panorama fue cambiando a medida que mayor cantidad de mujeres se integraron al mundo laboral, haciendo que, en primer lugar, las normas sociales se vuelvan más igualitarias y, en segundo lugar, se incremente la simetría entre las características de los cónyuges, particularmente los ingresos y el nivel educativo de cada uno de los cónyuges.

La creciente simetría mencionada es una consecuencia del emparejamiento selectivo y en este trabajo se mostrará en que grado este efecto positivo sirve como motor del incremento de las horas trabajadas por mujeres, puesto que, si las mujeres con mayor productividad se casan con hombres de similares características, entonces a mayor decil de ingreso de los hombres, mayor serán las horas ofrecidas por sus esposas en el mercado laboral puesto que el ingreso se maximiza en la medida que dos personas de similares productividades formen una pareja. En la siguiente sección se verá que, si bien el ingreso se maximiza, también se incrementa la desigualdad, por lo que existe un “trade-off” entre ingresos e igualdad.

2. 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

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

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.

El trabajo pionero de Cancian y Reed (1999) fue el primero en intentar una descomposición de los ingresos laborales por fuentes, descomponiendo dicha variable en sus componentes (ingreso del marido, de la mujer y otros); para la mencionada descomposición utilizan el coeficiente de variación del ingreso familiar:

$$CV_F = S_H^2 CV_H^2 + S_W^2 CV_W^2 + S_R^2 CV_R^2 + 2\rho_{HW} S_H S_W CV_H CV_W + 2\rho_{HR} S_H S_R CV_H CV_R + 2\rho_{RW} S_R S_W CV_R CV_W$$

Dónde:

S_K : Participación de la fuente K en el ingreso familiar total.

C_K : Coeficiente de variación de la fuente K ($CV_K = \sigma_K / \mu_K$)

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

μ_K : Ingreso medio de la fuente K , con K igual a F (ingreso familiar), W (ingreso de la mujer), H (ingreso del marido) o R (ingreso de otras fuentes).

Las autoras 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), el motivo reside en que dicho indicador puede descomponerse linealmente en sus componentes, tal como lo muestra la última ecuación.

Una vez efectuada la descomposición, se realizan cuatro experimentos contrafácticos para evaluar los efectos sobre la distribución del ingreso familiar: reducir el ingreso de las mujeres a cero, reducirlo marginalmente, mantener constante en el tiempo la media y la dispersión de los ingresos de éstas (μ_W y σ_W) y mantener constantes las correlaciones parciales del ingreso de las mujeres con las demás fuentes (ρ_{HW} y ρ_{RW}).

Ésta metodología es aplicada posteriormente por Campos-Vázquez, Hicapié y Rojas-Valdéz (2012) para estudiar los cambios en la distribución del ingreso familiar en México entre 1988 y 2012, en particular la magnitud del efecto del incremento en el ingreso laboral de las mujeres casadas. Con la mencionada metodología muestran que el principal determinante de la desigualdad del ingreso familiar es la desigualdad de ingresos de los maridos, en segundo lugar, la oferta laboral y la estructura de salarios de las

mujeres disminuye la desigualdad de salarios y por último, la correlación entre fuentes de ingresos carece de efecto sobre el ingreso familiar.

Otro camino seguido por la literatura consiste en descomponer algún índice de desigualdad por medio de distribuciones contrafácticas, siguiendo la metodología de DiNardo, Fortin y Lemieux (1996) o la de Machado y Mata (2005), como por ejemplo Daly y Valletta (2006) Daly y Valletta, 2006 en cuyo trabajo se 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.

El método consiste en simular distribuciones contrafácticas en base a parametrizaciones de 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. Dentro de ésta literatura, también es posible aplicar la mencionada metodología a la distribución conjunta de los niveles educativos entre las parejas, el reciente trabajo de Eika, Mogstad y Zafar (2014) en cuyo trabajo se comparan los resultados de Estados Unidos con los de una base de datos longitudinal de Noruega sobre niveles de escolaridad de individuos, la cual puede usarse para estudiar el emparejamiento selectivo por niveles educativos ya que es posible identificar en ella a las parejas en cada año, así como también obtener datos detallados no sólo del nivel de escolaridad más alto alcanzado sino también del título obtenido, permitiendo estudiar el emparejamiento a niveles de títulos universitarios inclusive.

Los últimos dos trabajos mencionados comparten con éste un aspecto metodológico importante, todos se basan en estimar una distribución de frecuencias conjunta (o, más precisamente una tabla de contingencia) de los deciles de ingresos y de los niveles de educación de las parejas, la cual es utilizada como “insumo” para realizar las simulaciones detalladas anteriormente y, con ellas simular distintos escenarios contrafácticos por medio de variaciones en diferentes parámetros, la metodología utilizada en este trabajo se detalla en la sección “Simulación”.

3. Tendencias de emparejamiento selectivo en Argentina

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 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

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

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.

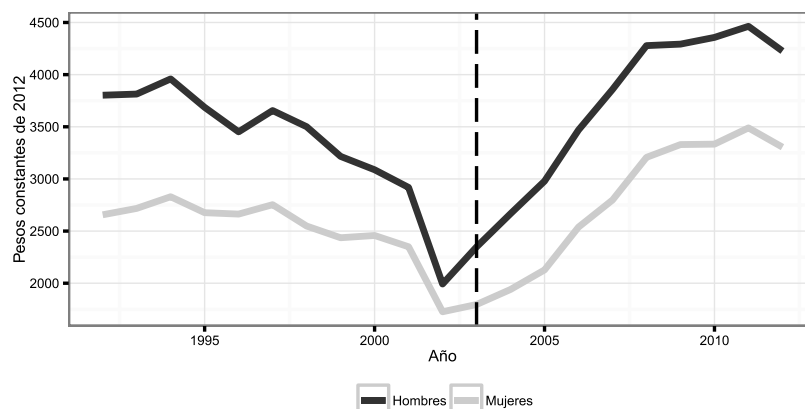


Figura 1: Ingresos laborales promedio por género y por año

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.

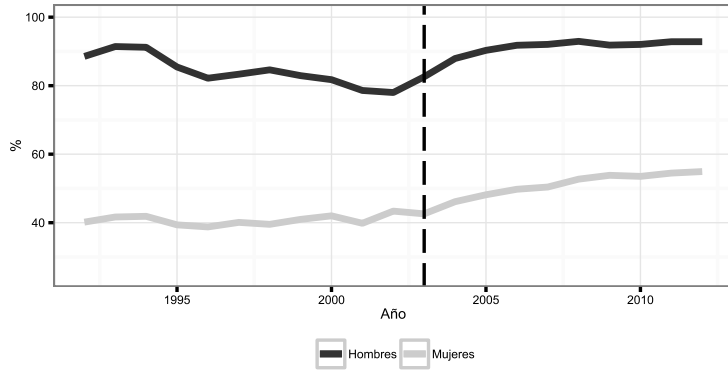
Las Figuras 2(b) and 2(c), replican al descomposición del Coeficiente de variación del ingreso total familiar desarrollada por Cancian y Reed (1999), definido como la suma de los ingresos laborales de cada uno de los cónyuges más un ingreso “residual” no atribuible a ninguna de las fuentes anteriores³.

En la Figura 2(b) se muestra las evolución de los coeficientes de variación de cada uno de los componentes del ingreso familiar, aquí se observa que la desigualdad, medida por medio del mencionado 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.

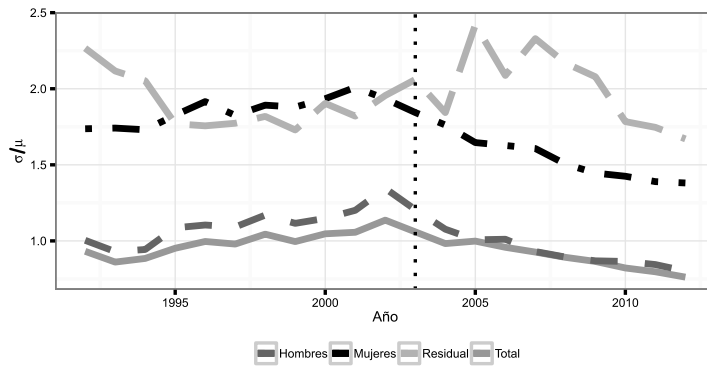
El coeficiente de variación del ingreso laboral femenino se comporta de forma similar, pero mostrando una mayor desigualdad (igual a poco menos del doble que el de sus maridos), pero mostrando también una tendencia decreciente, no ocurre así con la desigualdad del ingreso residual, cuyo comportamiento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

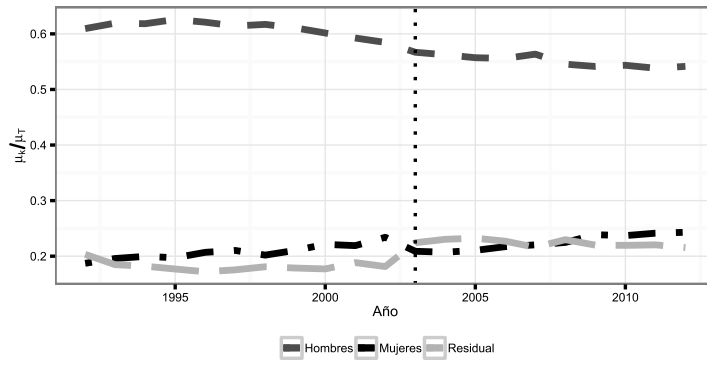
³Todas las magnitudes fueron calculadas a precios constantes de 2012



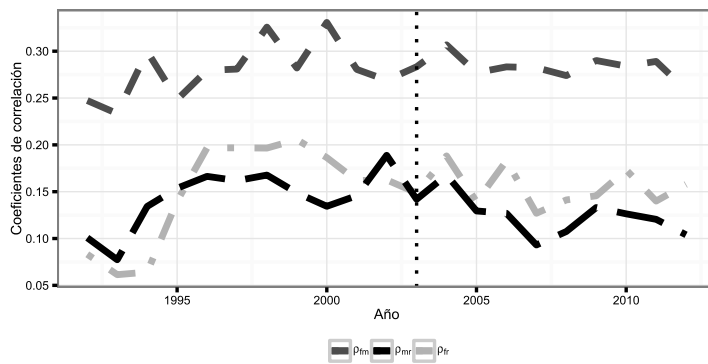
(a) Participación laboral por género y por año



(b) Coeficiente de variación



(c) Participación en el ingreso total familiar



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

la Figura 2(c) 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(d) 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 positiva⁴.

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%).

4. Emparejamiento selectivo y desigualdad

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, Cicowicz y Sosa Escudero (2013), pág 374).

El espíritu del método de simulación es similar al empleado por Bredemeier y Juesen (2013), ya que se trata de una microdescomposición no paramétrica 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. (2012), 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

⁴Más 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.

modesta en la desigualdad de 0,295 a 0,290 en 2004 utilizando datos de Seguridad Social⁵ y de 0,273 a 0,264 con datos del Panel Study of Income Dynamics (PSID).

5. 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 no paramétrica 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:

$$(f(y_m|x_m) * f(y_f|x_f)) f(x_m, x_f)$$

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.

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

$$f(y_m) * f(y_f)$$

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.

⁵Los datos provienen de la Survey of Income and Program Participation panels (SIPP-SSA)

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.

5.1. 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)⁶:

\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: $LFP_W \in \{WORK_W, \sim WORK_W\}$.

\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:

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

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 unipersonales, en cambio, las categorías son las siguientes:

⁶Se utiliza la misma nomenclatura que Greenwood y col. (2014) para facilitar la comparación.

$\mathcal{M}_{MARSTAT}$: 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:

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

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 IPUMS⁷ 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 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 grandes magnitudes 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.

⁷Integrated 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.

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 éstos⁸ 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:

Cuadro 1: Tabla de adulto equivalente-INDEC

Edad	Sexo	Valor Adulto equivalente
Menor a un año		0,33
1 año		0,43
2 años	Ambos	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

⁸ *menos que secundario completo, secundario completo, universitario incompleto, universitario completo y más que universitario completo*

la **Población Económicamente Activa** (PEA), para tal caso se genera una variable dummy `part` 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.

5.2. Simulación

Para realizar los cálculos que el método requiere se utilizó el código provisto por los autores⁹, 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 " 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 tres programas en R¹⁰, 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 eficiente¹¹.

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:

⁹Disponible en <http://www.aeaweb.org/articles.php?doi=10.1257/aer.104.5.348>

¹⁰Un 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 más 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.

¹¹Utilizando 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 propuesto originalmente por Kendall e implementado en Stata.

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.

5.3. Coeficientes de regresión

La primer medida parte de la estimación de la ecuación:

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

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 $Año_{st} = 1$ si $s = t$ y $Año_{st} = 0$ 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.

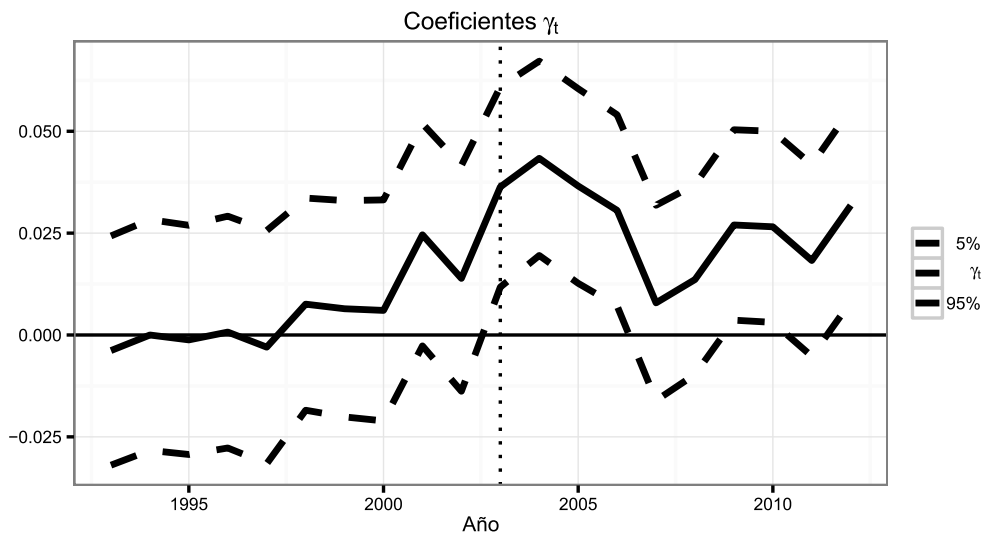


Figura 3: Incremento en el emparejamiento selectivo

La Figura 3 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 et Al, 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.

5.4. Tau de Kendall

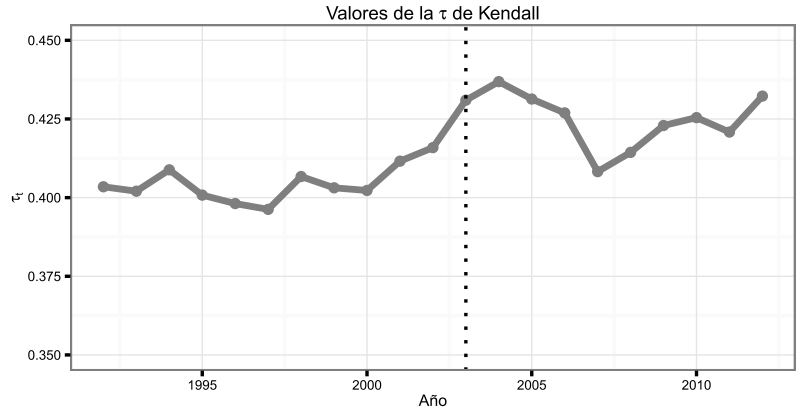
Otra medida presentada por los autores es el coeficiente de correlación por rangos, denominado “Tau de Kendall”, el cual se utiliza para medir la fuerza de la asociación entre dos variables categóricas, en este caso se utiliza para medir la correlación entre los niveles educativos de los cónyuges.

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

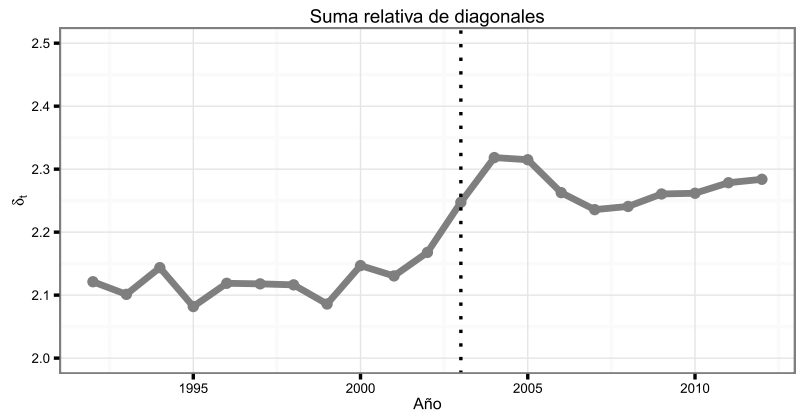
$$\tau = \frac{S}{\frac{n(n-1)}{2}}$$

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:

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



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



(b) Ratio de la suma de diagonales principales

Figura 4: Indicadores de emparejamiento selectivo

En la Figura 4 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 posterior¹² para luego amesetarse en un nivel más alto a partir de 2005.

5.5. Suma de diagonales

Por último, 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: $\sum_{i=1}^6 a_{ii}^t$.

¹²En 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.

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:

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

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.

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

Cuadro 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: 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 2 y 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.

6. Indicadores de desigualdad

6.1. Definición de 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 : $f_{ij} = \frac{N_{ij}}{N}$, donde 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:

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

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

$$l_p = \text{Lorenz}_p(\{f_{ij}\}, \{r_{ij}\})$$

$$g = \text{Gini}(\{f_{ij}\}, \{r_{ij}\})$$

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:

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

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)

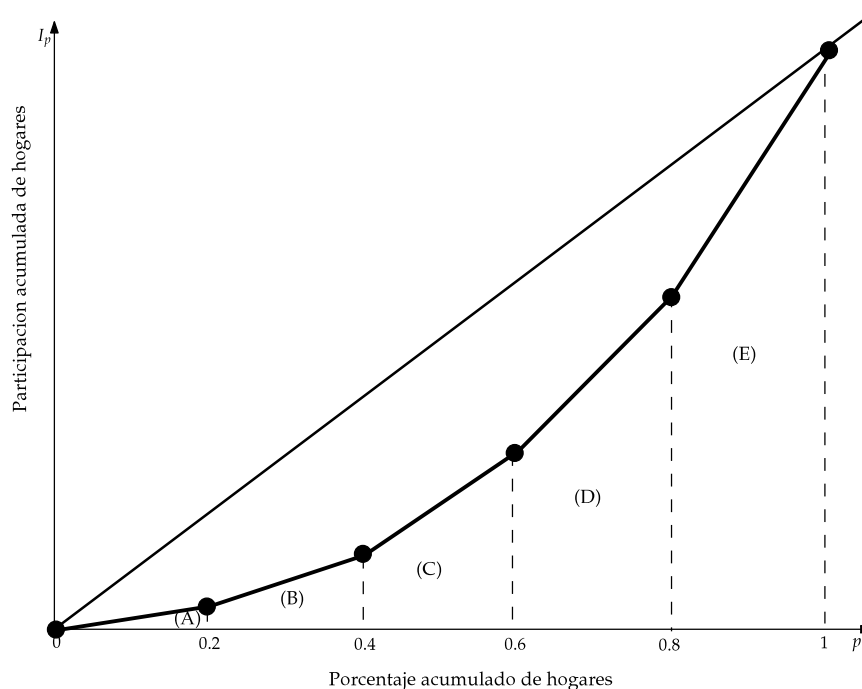


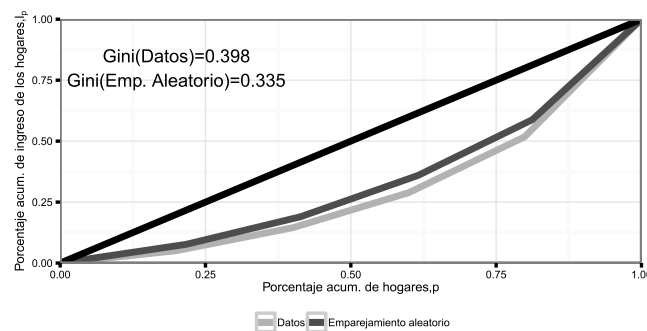
Figura 5: Curva de Lorenz

El coeficiente de Gini se calcula en base a las áreas de la Figura 5 de la siguiente manera:

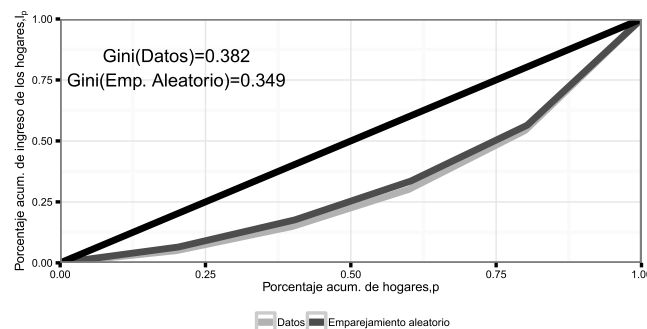
$$g = 1 - 2 \left[\underbrace{\frac{p_1 I_1}{2}}_{(A)} + \underbrace{\frac{(I_1 + I_2)(p_2 - p_1)}{2}}_{(B)} + \underbrace{\frac{(I_2 + I_3)(p_3 - p_2)}{2}}_{(C)} + \underbrace{\frac{(I_3 + I_4)(p_4 - p_3)}{2}}_{(D)} + \underbrace{\frac{(I_4 + 1)(1 - p_4)}{2}}_{(E)} \right]$$

Esto es, el coeficiente de Gini es igual a: $1 - 2 \times (\text{Área bajo la curva de Lorenz})$, donde el área bajo la curva de Lorenz se aproxima como la suma de las áreas de los trapecoides (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 como:

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



(a) Período 1992-1996



(b) Período 2008-2012

Figura 6: Curvas de Lorenz: Datos versus emparejamiento aleatorio

En la figura 6 se muestran las curvas de Lorenz y los coeficientes de Gini para los períodos inicial y final, la variación de ambos es muy poca, 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 familiar¹³.

7. 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.

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

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

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:

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

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

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

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

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

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

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

Con:

¹³Por "Ingreso Total Familiar" se hará referencia siempre al ingreso familiar corregido por adulto equivalente de la forma detallada en la sección anterior.

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

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 $f(LFP_W)$.

7.1. 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, el experimento contrafáctico consiste en reemplazar las f_{ij} para $(i, j) \in \mathcal{M}$ observadas por las que existirían si el emparejamiento fuera 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:

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

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 puntos contra 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 son las Tablas 4 y 5.

Cuadro 4: 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.

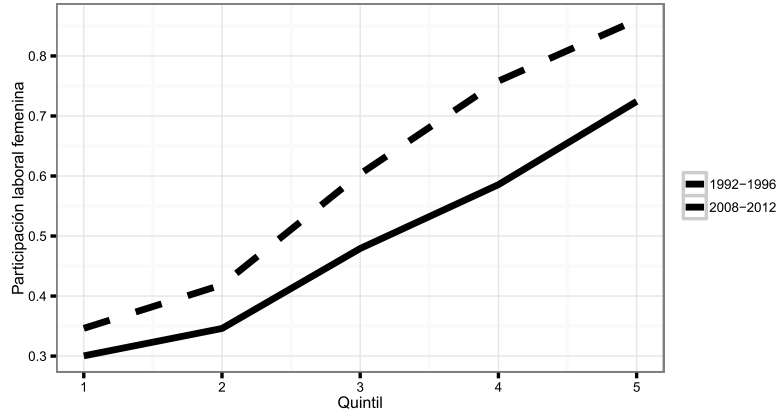
Cuadro 5: Tabla de contingencia para 2008-2012 con emparejamiento aleatorio

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

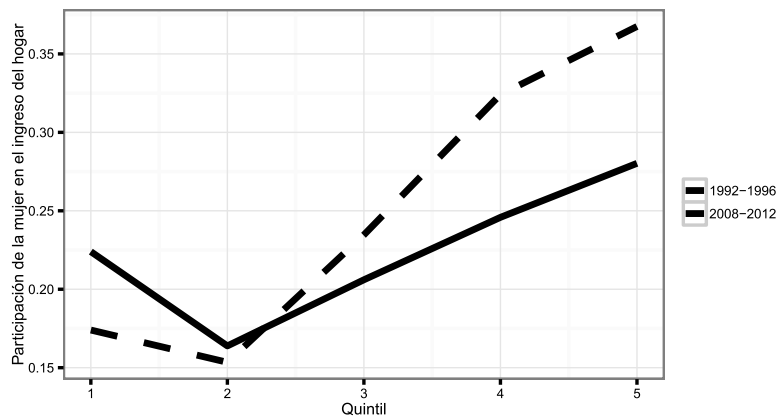
Fuente: Elaboración propia en base a INDEC.

7.2. 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 7: Participación laboral de las mujeres casadas: 1992-1996 vs. 2008-2012

El efecto del emparejamiento selectivo depende de que las mujeres se encuentren empleadas, en la Figura 7(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 7(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:

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

Dónde $j = 4$ si $i = 1$ y $j = 1$ si $i = 4$ 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.

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 6 y 7.

Cuadro 6: 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 7: 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.

7.3. Tablas de contingencia estandarizadas

OtroOtro 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 1 y 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 la diferencia con relación a la distribución marginal deseada: $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$. Si se alcanzó la 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:

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

$$G_i^{ST+\Delta LFP} = 1 - \frac{2}{\mu(F)_i} \int_0^1 \int_{\underline{y}_i}^{\bar{y}_i} y f^i(y/E_H, E_W, LFP_W) g_S^i(E_H, E_W/LFP_W) h^j(LFP_W) dy dq$$

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.

Cuadro 8: 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.

8. Chequeos 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 se puede ver que tan 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 9: 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 estandarizada 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.

Cabe destacar que el artículo de Greenwood y col. (2014) no realiza éstos chequeos de robustez, lo que hace dudar de los resultados, en particular debido a la forma en la que se calcula el coeficiente de Gini, puesto que no debería mostrar variaciones en función del número de cuantiles en los que se dividen los datos, esto puede deberse a las siguientes causas, en primer lugar, puede suceder que sea 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.

9. Conclusión

Los resultados de ésta sección muestran que el emparejamiento selectivo explica poco menos del 10% de la desigualdad total, lo que refuerza a las conclusiones alcanzadas en la sección previa, es decir, que el emparejamiento selectivo cumple un rol secundario en la determinación de la desigualdad en la distribución del ingreso. Sin embargo se muestra también que éstos resultados no son robustos a modificaciones en un grupo de parámetros muy importantes, 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 número 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 encuentran 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, esto 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.

Referencias

- Abrevaya, Jason (1999). «Computation of the maximum rank correlation estimator». En: *Economics Letters* 62, págs. 279-285.
- 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, Massachusetts: 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.
- 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/B978044459428000072>.
- 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>.
- 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/>.
- Greenwood, Jeremy y col. (2012). *Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment and Married Female Labor-Force Participation*. Working Paper 17735. National Bureau of Economic Research. DOI: 10.3386/w17735. URL: <http://www.nber.org/papers/w17735>.
- 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>.
- 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).
- 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.
- Machado, José A. y José Mata (2005). «Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regressions». En: *Journal of Applied Econometrics* 20, págs. 445-465.
- MATLAB (2013). *version 8.10.0.604 (R2013a)*. Natick, Massachusetts: The MathWorks Inc.
- Mosteller, Frederick (1968). «Association and Estimation in Contingency Tables». En: *Journal of the American Statistical Association*, 63.321, págs. 1-28.
- 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.
- 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>.