ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

L Reunión Anual

Noviembre de 2015

ISSN 1852-0022
ISBN 978-987-28590-3-9

Assortative Mating and Household Income Inequality in Argentina.

Funes Leal, Víctor

Assortative Mating and Household Income Inequality in Argentina

Víctor Funes Leal*
Universidad Nacional de La Plata

Resumen

En este trabajo se busca estudiar cual es el impacto del emparejamiento selectivo sobre la distribución del ingreso total familiar en Argentina. Con éste objetivo se utiliza el método de microdescomposición de (Greenwood y col., 2014), el cual se basa en el coeficiente de Gini y las tablas de contingencia de la distribución de los niveles educativos de cada cónyuge para las parejas casadas para simular los resultados que surgirían del emparejamiento aleatorio.

El resultado obtenido es que el emparejamiento selectivo sólo cumple un rol secundario en la determinación de la desigualdad del ingreso total familiar, porque a pesar que el método afirma que explica casi el 10 % de la magnitud de la desigualdad, éste resultado carece de robustez a cambios en parámetros y, además, el cálculo de una serie de indicadores de selectividad contradice los resultados del método.

JEL: D31, C15, J12

Palabras clave: emparejamiento selectivo, distribución del ingreso, oferta laboral femenina, brecha de género.

Abstract

We use the decomposition method of (Greenwood et al., 2014) to assess the contribution of Assortative Mating to household income inequality for Argentina. We estimate the Gini coefficient and contingency tables for for spouses' educational attainment in order to simulate the outcomes of random mating.

We find that Assortative Mating plays a minor role in the determination of household income inequality, despite explaining about 10% of inequality, the results have very little robustness to changes in parameters and are also at odds with values of a number of sorting indicators.

JEL: D31, C15, J12

Keywords: assortative mating, income distribution, female labor supply, gender wage gap.

*victorefunes@gmail.com

Contents

1 Introduction

The determinants of family income distribution have attracted considerable attention in the last fifteen years, but most empirical analyses have concentrated in advanced economies, there are very few studies on this topic in emerging economies, such as Argentina. Most papers on the determinants of household income inequality have studied the following three factors:

1. **Wage gap and labor participation**: Starting in the mid twentieth century, female labor participation has increased significantly in western countries, and neither Argentina nor the remaining Latin American countries were the exception. This increase had a very strong impact on fertilty, intra-family income distribution, marriage and divorce rates (Killingsworth and Heckman, 1986).

2. **Fertility**: The increase in hours of work supplied in the market, required women to have fewer children since now they can no longer raise the same number of children as their mothers or grandmothers.

3. **Assortative mating**: According to this theory, long-term human mating decisions are influenced by a number of factors related to the degree of similarity between both spouses, among them are educational attainment, labor income, ethnic background, religion and many others.

Schwartz (2010) asserts that until the 1950s, marriage was an economic relationship between two persons with complementary income sources, typically males supplied their labor in the market, while females focused on providing public goods at home (cooking, cleaning, child rearing). This pattern of labor division changed since the 1960s, and it is evidenced by a change in the sign of the correlation of spouses' earnings from negative to positive. From this date on, high-earning males tended to marry with high-earning females, and consequently, spouses began to resemble each other in a number of observable characteristics beside earnings (such as educational attainment, age, etcetera.)

2 Literature review

In his pioneering work on the economics of the family, Becker (1991) argued that there is an optimal mating level, reached in the point when it is not possible to change the combination of spouses in order to improve one persons's welfare without hurting other person. In the same token, there is negative assortment when there are benefits arising from specialization, that is, when one of the spouses offers hours of work in the market while the other produces public goods at home. On the other hand, there is positive mating when there are no benefits from specialization, that is, productivity is maximized when both spouses supply hours of work in the market.

Mare, 1991 is another early study on educational Assortative Mating, he argues that both educational attainment and age at marriage affect a person's chance of finding another person with similar characteristics because the educational system forces homogeneity in age and social class. This homogeneity structures student's acquaintances increasing their likelihood to marry a person with similar characteristics later in their lives.

According to his theory, marital sorting may increase or decrease educational homogamy (marriages between two similar persons) depending on the relative strength of two opposing forces:

- Both the increase in women's participation and the closing of the wage gap have altered expectations of mean and women about marriage, if women can now attain higher income and education levels they may now seek partners more similar to them in these aspects, therefore increasing sorting measured using income or education levels.

- As a consequence of pursuing a career, mean and women marry at later ages than their parents or grandparents, this means they may lose a large part of their acquaintances from high school or university, this, a part of their "social capital", (the pool of individuals where they are supposed to meet their spouses). If this assertion is true, then sorting should decrease rather than increase in time, Mare shows evidence that the first effect is larger and therefore Assortative Mating is indeed increasing.

Later, Cancian and Reed (1999) used microeconometric decomposition methods to analyze the effects of assortative mating on household income inequality. They used a decomposition of the coefficient of variation (CV) of total household income, because it can be used as a measure of income inequality. This coefficient is preferred instead of other more popular indicators, such as the Gini or Theil indices because neither of them can be linearly decomposed in order to assess the contribution of each of its components.

Later works have continued this line of analysis, using other methods, they can be classified according to the following taxonomy:

Parametric decompositions : this is the method described in the previous paragraph, it uses the coefficient of variation as a measure of income inequality, despite an important shortcoming, namely, it has a tendency to assign greater weight to transfers between individuals with a high income differential. Nevertheless it is used because of it can be decomposed linearly without resorting to more advanced statistical techniques. A recent work by (Campos-Vázquez, Hicapié, and Rojas-Valdéz, 2012) uses this method to study household income inequality in Mexico, also Funes Leal (2015) does a similar analysis for Argentina. A second drawback of this decomposition is the fact that uses just one statistic and not a distribution as a counterfactual, that is just one point per period.

Non-parametric decompositions : a different decomposition method was devised by DiNardo, Fortin, and Lemieux (1996) which is based on simulating counterfactual distributions based on a set of covariates, this simulated distributions are obtained using a re-weighting approach, then an income inequality measure is measured both on the original distribution and the counterfactual distribution. Two important papers applied this method in the recent years. Daly and Valletta (2006) decomposed a number of inequality and poverty indicators for the United States[1] using data form the *Current Population Survey* in an attempt to measure the impact of the growing dispersion of male labor income on the household income distribution. Next, (Eika, Mogstad, and Zafar, 2014) used it to study the effects of assortative mating on income inequality for the United States and Norway using the correlation between educational attainment levels as a mating measure.

Semi-parametric decompositions : a most recent method is a combination of the previous, Greenwood et al. (2014) uses a semi-parametric method to decompose the Gini coefficient with IPUMS[2] data for the United States. Their decomposition method splits the data in a large number of groups and computes the Gini coefficient for each of them using their average income and size. Then a number of counterfactual scenarios are be estimated with the means of mating patterns and female labor participation.

[1] Median, standard deviation, coefficient of variation, quantile ratios, Gini and Theil coefficients, mean log deviation and poverty ratio $(FGT(0))$

[2] International Public Use Microdata Samples

3 Some stylized facts

3.1 Gini indices

In order to characterize the trend of income inequality in Argentina we can split our 21-year time span into many periods, we will use two of these classifications, the first one, on in this section, makes use of two clearly-defined periods: period one began in 1991-1992 and finished with the 2001-2002 crisis, it was characterized by a growing inequality in labor income due to larger returns to skilled labor, greater trade liberalization, growing labor force and increasing unemployment. The second period, started in late 2002 and was characterized by the opposite trend in income inequality (a phenomenon observed throughout Latin America), a much slow-growing female labor participation and lower unemployment. The second periodization, will be described in a later section, uses four different periods with the objective of increasing the number of observations and avoid business cycle effects.

The most widely used measure of income inequality by the literature is the Gini coefficient (Cowell and Flachaire, 2015), defined as:

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

Where y is a measure of income (total household income in this case), $F(y)$ is the cumulative distribution of y, and $\mu(F)$ is the mean of y. Income inequality, as measured by this index, fell from an average of 0.5 in the 1992-2002 decade to roughly 0.42 in the following decade. Figure 1 plots the estimated value of the Gini index per year along with its confidence intervals estimated using bootstrap[3] in order to accurately estimate standard errors. We used using $R = 500$ bootstrap replications, in the following way: let $\hat{G}_i$ for $i = 1, \ldots, R$ be the estimated value for the Gini index in each replication, $\bar{G} = \frac{1}{R}\sum_{i=1}^{R} \hat{G}_i$ the average value estimated over all replications, then the bootstraped standard errors will be:

$$SE(\hat{G}) = \sqrt{\frac{1}{R-1}\sum_{i=1}^{R}\left(\hat{G}_i - \bar{G}\right)^2} \tag{2}$$

Using this standard errors we can compute confidence intervals for the Gini Index, this are the shaded lines in Figure 1.

In Figures 2(a) and 2(b) we replicate previous Figure for subsamples of males and females in order to show the variation of individual income inequality. Total income is defined in the same way as in **SEDLAC** database[4], this is, as the sum of both labor and non-labor income for all persons. Labor income is defined as all payments received by individuals from both their main and secondary sources of income. Non-labor income, by contrast, is equal to the sum of payments originated in pensions, capital benefits and transfers (unemployment insurance, scholarships, etcetera).

The behavior of male income inequality closely resembles the aggregate because it is by far its main component. But female income inequality has a very different trend compared to that of males, despite showing larger levels in the 1990s and much lower values in the 2000s, we can see that the value of the Gini index is significantly higher than that of either their counterpart or the aggregate. This is related to a much lower female labor participation, especially in the 1992-2001 decade.

In the same vein, it can be seen that the reduction in female income inequality in 2002-2012 decade is higher than that of males, this is also related to female labor participation since it had a large increment in this period. Finally it must be mentioned that the **Encuesta Permanente de Hogares (EPH)**, Argentina's oldest continuous household survey, had a methodology modification

[3]See James et al. (2013) for a comprehensive introduction to the bootstrap.

[4]Socio-Economic Database for Latin America and the Caribbean (`http://sedlac.econo.unlp.edu.ar`)

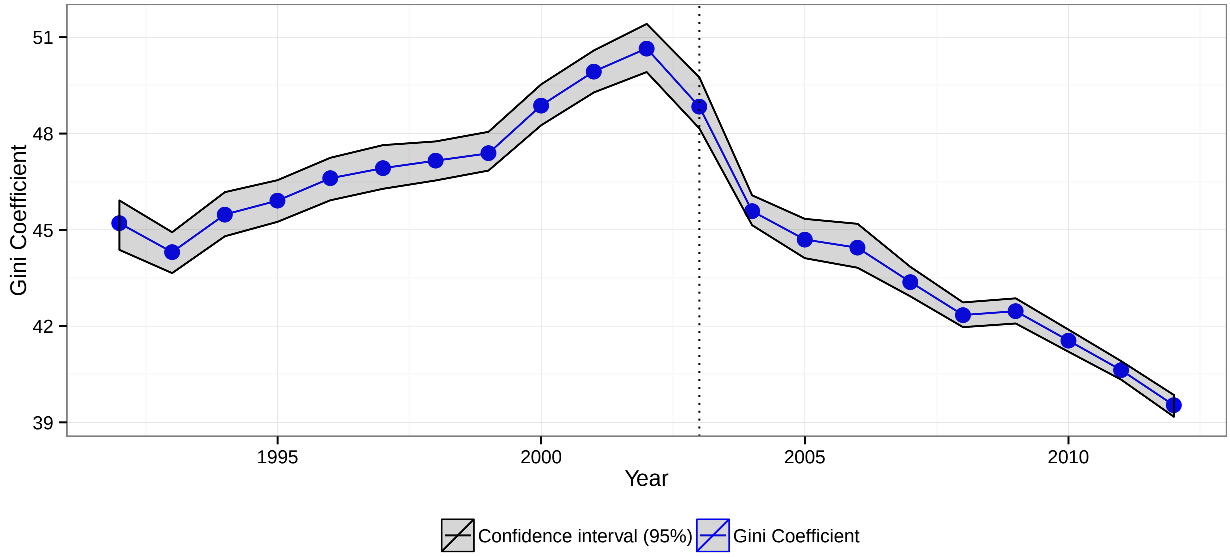


Figure 1: Gini Coefficient for real per capita household income and bootstrap confidence intervals

in the third quarter of 2003[5]. This modification consisted on a series of methodological improvements that had a sizeable impact on our inequality measures. It remains an open question the degree in which this modification changed our estimates, but we can be sure that the this is only a minor change in levels which left the rates of variation almost unchanged. A dotted vertical line was included in all time plots to mark the year 2003, when this change took place.

3.2 Income distributions

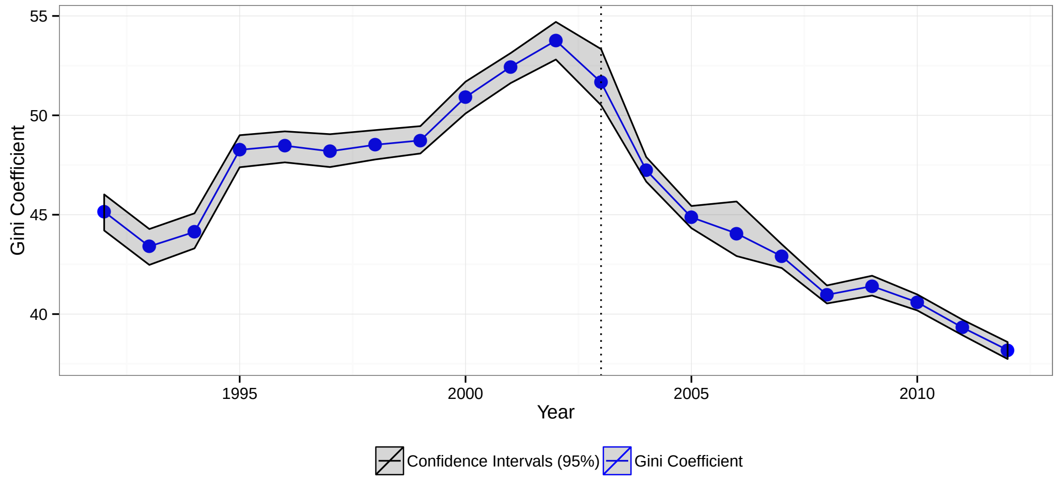
Another way to look at the evolution of income inequality is using a histogram of the variable, unfortunately, they make the distribution constant on each interval, causing discrete jumps at every end. One way to overcome this difficulty is employing a smoothing function, called a "kernel function" to estimate a distribution $\hat{f}(y_0)$ at every point in its support. In addition, given that the distribution of any income variable is highly asymmetric, because there usually are many households with low incomes and a comparatively small number of households with high incomes, we need to use a logarithmic transformation in order to make the distribution symmetric, but preserving the orderings given that the log function is monotonic.

The kernel function method developed by Rosenblatt and Parzen (Pagan and Ullah, 1999) approximates the true distribution of income $(f(y))$ using a function defined as:

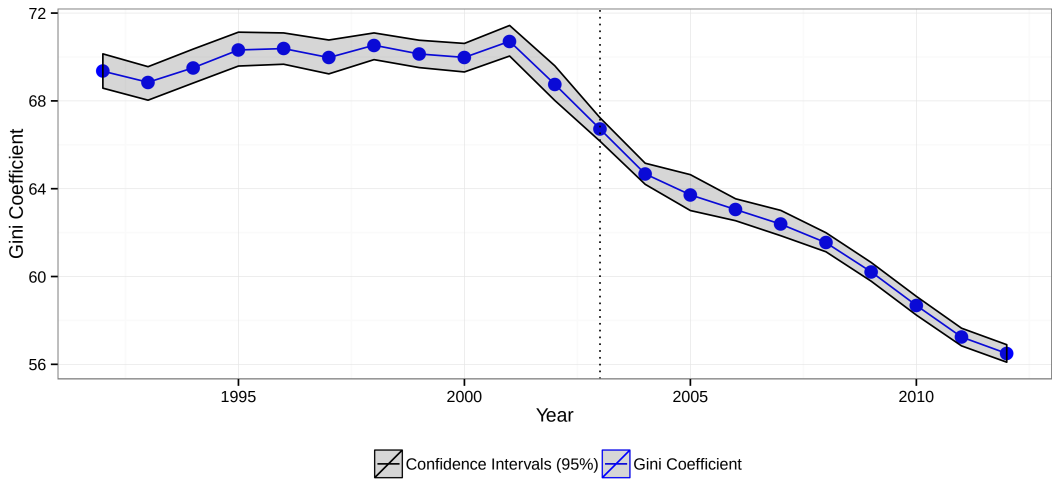
$$\bar{f}(y_0) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{y_i - y_0}{h}\right) \tag{3}$$

$K(.)$ is the kernel function, it measures how close an arbitrary observation y_i is from y_0, there is a large number of possible kernel functions, but in income distribution analysis two functions are widely used: normal and Epanechnikov(Cowell and Flachaire, 2015), but we will use the later:

[5]For the year 2003, we use the "old" EPH ("EPH Puntual") for the first and second quarters and the "new" EPH ("EPH Contínua") for the third and fourth quarters.



(a) Gini coefficient for male total income



(b) Gini coefficient for female total income

Figure 2: Decompositions of total household income

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

It is also necessary to determine what value of the parameter h will be used, it is called the "bandwidth" and, according to Gasparini, Cicowiez, and Sosa Escudero (2013) the chosen value is even more important than the kernel function itself. To avoid further complications we will use Scott's method, which selects the optimal bandwidth (h^*) as:

$$h^* = 1.06\sigma n^{-1/5}$$

Where n is the sample size and σ is an estimator of the standard deviation of $\log(y)$. In Figure 3 the distribution of real total household income is plotted (in 2012 prices) using an adult-equivalent correction. The values are for the two extreme years (1992 and 2012) and for one intermediate year (2002), chosen since this was the year with the lowest average per capita income for the entire period (once again, due to the 2001-2002 economic crisis).

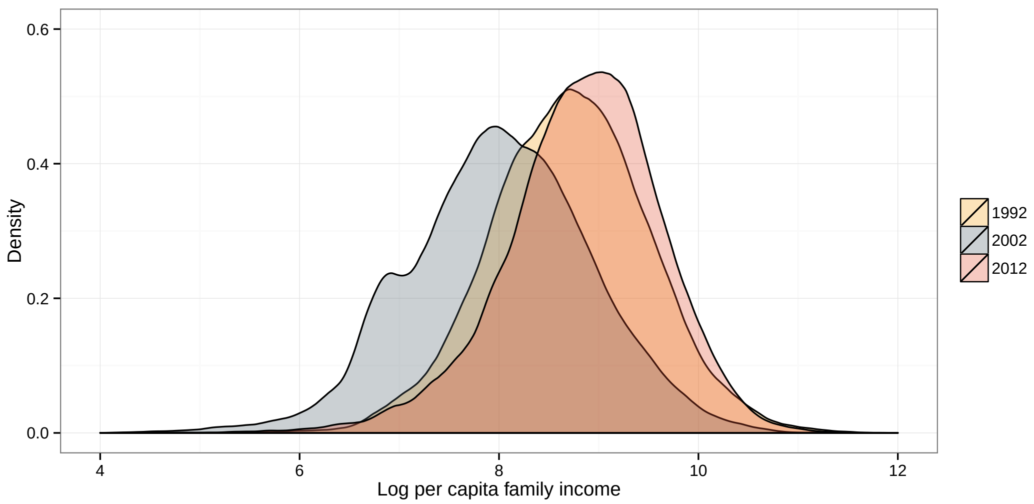


Figure 3: Distribution of log real total household income (2012 prices)

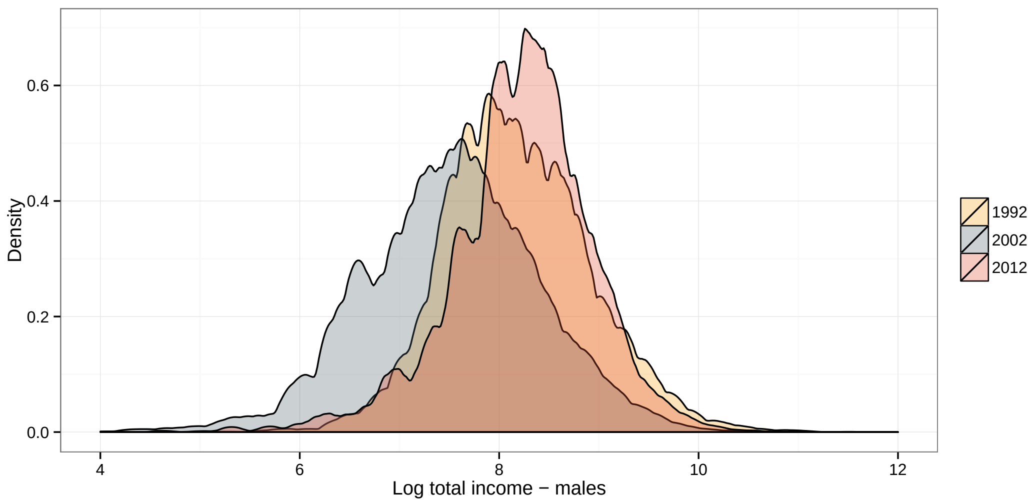
Firstly, this overlapping distributions show the profound effect of the 2001-2002 crisis: there was a significant decrease of both mean and median income relative to the initial year, nevertheless, in the following period they recovered steadily. Secondly, using a log scale distorts the variation in levels, for example, total household income fell from ARS 7806.78 in 1992 to ARS 4208.77 in 2002, then it increased again to ARS 9062.95 in 2012[6]. In other words, income almost halved as an effect of the crisis, but it reversed its trend to reach a 16% higher value relative to that of 1992.

In Figure 4 estimated distributions of total male and female income are plotted using the exact same Epanechnikov kernel function and optimal bandwidth parameter.

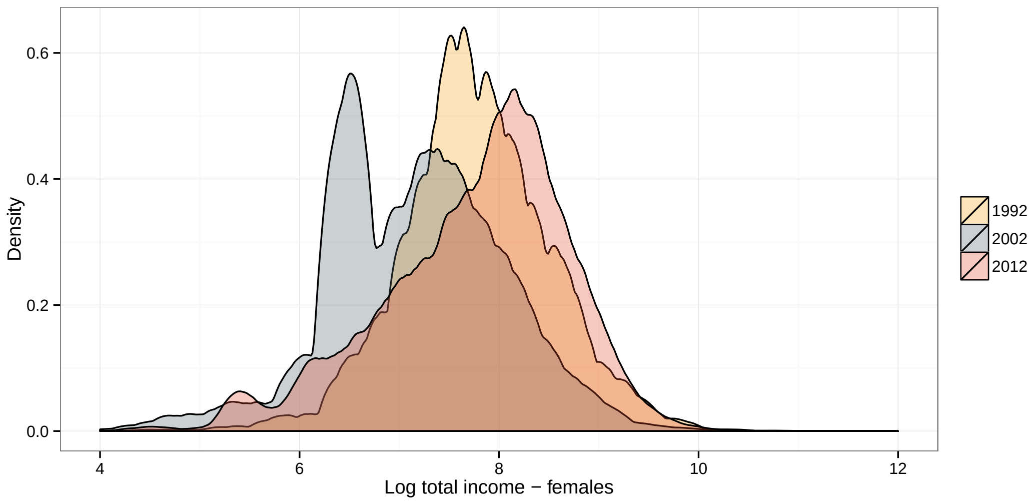
Figuras 4(a) and 4(b) plot the distribution of total male and female income for the same years as the previous plot. The following facts can be highlighted:

- Both distributions (for males and females) are remarkably similar in shape because the log transformation drops all observation with zero income, thus reducing the male-female wage gap.

[6]All magnitudes are measured in 2012 ARS



(a) Distribution of log total income - males



(b) Distribution of log total income - females

Figure 4: Distributions of log total income by gender (2012 prices)

- The 2002 financial crisis triggered a series of social protection policies which had a greater impact on females, due to their lower average wages, this is evidenced in 2002, when the distribution of the log income became bi-modal, and remained this way, but with a progressively smaller "peak" at lower wages.

- The male-female wage gap has decreased throughout the period, but the distribution of female total income is still asymmetric compared to that of males, it clearly has a larger density in the left tail.

4 Assortative mating and educational attainment

4.1 Introduction

The purpose of this section is to replicate the method used by (Greenwood et al., 2014), a microeconometric decomposition exercise to appraise the impact of assortative mating on income inequality as measured by the Gini index. The theory behind this exercise is described in Greenwood et al. (2015) where mating and labor supply decisions by individuals are modeled in a utility maximization setting, authors also provide a number of equilibrium and stable matching conditions.

According to Greenwood et Al, if couples were formed randomly using US data, the Gini index for 2005 would have fallen from 0.43 to 0.34 using American Community Survey data. On the other hand, Eika, Mogstad, and Zafar (2014), using data from the Current Population Survey and a different decomposition method (that of DiNardo, Fortin, and Lemieux (1996)) found a more modest decrease in inequality due to assortative mating, from 0.403 to 0.384. Another method was used by Hryshko, Juhn, and McCue (2014), who also found a much lower decrease in inequality with social security data from the United States[7], according to this authors, values of the Gini index fall from 0.295 to 0.290 in 2004, they also replicate this experiment with data from the Panel Study of Income Dynamics (PSID), finding a decrease from 0.273 to 0.264.

4.2 Simulation method

Greenwood et al. (2014) introduced a semi-parametric decomposition method based on contingency tables, that is joint distribution of educational classes for both spouses, with the objective of simulating outcomes of random mating. It is useful at this point to use the taxonomy proposed by Harmenberg (2014), who claims there are two competing approaches for studying assortative mating with microeconometric decompositions. The first method, called *addition approach* computes the income from each "pseudo-household" adding up spouses' incomes, the second one, called *imputation approach*, imputes the pseudo-income assuming that it is distributed in a similar way as actual income for households with similar observable characteristics.

Let:

x_m, x_f : individual characteristics for males and females (educational attainment, labor participation, number of children, etc.)

y_m, y_f : individual income of both spouses, such that total income for household i is defined as: $y_i = y_{mi} + y_{fi}$.

$f(z_i)$: true distribution of variable z_i

Under the addition approach, attributes x_m, x_f are kept constant for all couples, this means that the following distribution is estimated:

[7] Data from the Survey of Income and Program Participation Panels (SIPP-SSA)

$$(f(y_m|x_m) * f(y_f|x_f))\, f(x_m, x_f) \tag{5}$$

This method assumes that $m(f)$ are randomly distributed, such that $y_m|x_m$ and $y_f|x_f$ are independent.

Under non-conditional randomization, attributes are kept constant and the expression is reduced to:

$$f(y_m) * f(y_f) \tag{6}$$

The imputation approach computes the following expression: $f(y|x_m, x_f) * f(y_m) * f(y_f)$ and household income is calculated with its observable characteristics.

Both approaches have advantages and disadvantages, the addition approach avoids wasting information by randomizing the existing population, but fixes labor supply and income since both are exogenous to the decision of starting a household. Imputation approach takes labor supply and income into account when creating a household, but wastes information because observable attributes do not allow perfect imputation.

(Greenwood et al., 2014) use a form of the the imputation method, because they approximate the distribution of income using empirical quintiles and all combinations of educational levels, and finally they compute the counterfactual distribution for each combination of educational levels. With these data, they use an estimated distribution of educational levels as weights and aggregate the income distribution by quintiles along all education levels in order to obtain the new household income distribution.

The decomposition method requires a re-estimation of the Gini index using a large number of groups on the basis of: marital status, educational attainment, labor participation and number of children.

$\mathcal{M}_{E_H}$: husband educational level index, $E_H \in \{P-, P, S-, S, U-, U\}$ where each category represents, primary incomplete, primary complete, secondary incomplete, secondary complete, higher education (university and/or technical degrees) incomplete and higher education complete.

$\mathcal{M}_{E_W}$: wife educational level index, $E_W \in \{P-, P, S-, S, U-, U\}$.

$\mathcal{M}_{LFP_H}$: husband labor participation index, $LFP_H \in \{WORK_H, \sim WORK_H\}$, where the first category refers to those who participate in the labor market, while the second refers to those who don't.

$\mathcal{M}_{LFP_W}$: wife labor participation index, $LFP_W \in \{WORK_W, \sim WORK_W\}$.

$\mathcal{M}_{KIDS}$: index for number of children $KIDS \in \{0, 1, 2, 2+\}$, here the last category is for couples with more than two kids.

The full set of indices is, then:

$$\mathcal{M} = \bigcup_{E_H, E_W, LFP_H, LFP_W, KIDS} (\mathcal{M}_{E_H} \cap \mathcal{M}_{E_W} \cap \mathcal{M}_{LFP_H} \cap \mathcal{M}_{LFP_W} \cap \mathcal{M}_{KIDS}) \tag{7}$$

Therefore, for married couples we have a total of: $6 \times 6 \times 2 \times 2 \times 4 = 576$ groups.

For non-married households we have the following group indices:

$\mathcal{M}_{MARSTAT}$: index for individual's marital status,

$MARSTAT \in \{NMAR_M, NMAR_F, DIV_M, DIV_F\}$, and its categories are single men, single women, divorced men and divorced women. .

$\mathcal{M}_E$: index for educational level, $E \in \{P-, P, S-, S, U-, U\}$

$\mathcal{M}_{LFP}$: index for labor force participation, $LFP \in \{WORK, \sim WORK\}$.

$\mathcal{M}_{KIDS}$: index for number of children: $KIDS \in \{0, 1, 2, 2+\}$.

The full set of indexes for non-married couples is:

$$\mathcal{M} = \bigcup_{MARSTAT,E,LFP,KIDS} (\mathcal{M}_{MARSTAT} \cap \mathcal{M}_E \cap \mathcal{M}_{LFP} \cap \mathcal{M}_{KIDS}) \tag{8}$$

For single and divorced households, the number of groups is: $4 \times 6 \times 2 \times 4 = 192$, so that adding up we end up with $576 + 192 = 768$ groups.

4.3 Data: sources and description

We use data from the Encuesta Permanente de Hogares (EPH) for year 1992 to 2012, instead of using IPUMS data as Greenwood et al. (2014) for two reasons, first, Argentine population censuses do not gather any income data, and consequently they cannot be used for distributional analyses and, second, even if this was possible at the time of writing this document the full dataset from the latest population census (2010) had not been uploaded yet.

The Encuesta Permanente de Hogares underwent a very significant methodological improvement in 2002, it mainly consisted in the following modifications[8]:

- Survey frequency changed from bianually to quarterly.
- Sampling methodology was modified from a simple stratified sample to a rotating panel, for this reason the survey is called "continuous-EPH" from that time on.
- The questionnaire was heavily modified, as a side effect, some compatibility between the old and the new EPH was lost in important questions such as length of unemployment spell and others related to income sources.
- The number of urban areas in the survey increased in order to make the survey representative of all urban population (roughly 90% of total population of Argentina.)

These changes in methodology, however, are not likely to cause a large impact on our results because all estimation are based on aggregates on a group level and unchanged family characteristics (employment status, age, education and total income.) Nevertheless, this greater frequency from 2003 onwards increases the number of observations per year by a factor of two.

For our analysis we group years in four blocks, the rationale for doing this is twofold: avoiding business cycle effects on labor participation and incomes and also increasing the size of our sample[9], much smaller than those of IPUMS. The blocks are listed below, but our analysis will be centered on periods 1 and 4:

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

[8] Beccaria and Groisman, 2008.

[9] This is similar to Bredemeier and Juessen (2013)

4.4 Data cleaning

In order to attain the required data we must first perform a series of manipulations. First, we drop all members of the family other than the head, his or her spouse, and all children under 18 years old. Next we keep only families which have both a head and a spouse with ages between 24 and 55 years old in order to capture the smallest possible number of students and pensioners.

The next step is to classify households according to the categories previously described according to marital status and gender of its members. Here is were we depart from Greenwood et al. (2014) because we use six education levels instead of five, we do this because the former classification would leave the highest level ("postgraduate education") barely unpopulated, especially in period 1, due to a relatively lower average education level in Argentina compared to that of the United States.

The fourth step involves estimating an adult-equivalent total household income, once again we depart from the original study because we do not use the OECD equivalence scales[10]. We use instead the equivalence scale form Argentina's National Statistics Office (INDEC):

Table 1: Adult equivalence Table (INDEC)

Age	**Sex**	**Adult-equivalent Value**
Under a year old	Both	0.33
1 year		0.43
2 years		0.50
3 years		0.56
4 to 6 years		0.63
7 to 9 years		0.72
10 to 12 years	Males	0.83
13 to 15 years		0.96
16 to 17 years		1.05
10 to 12 years	Females	0.73
13 to 15 years		0.79
16 to 17 years		0.79
18 to 29 years	Males	1.06
30 to 59 years		1.00
60 years and over		0.82
18 to 29 years	Females	0.74
30 to 59 years		0.74
60 years and over		0.64

Source: Bergés (2011).

We also need to identify persons who offer their hours of work in the labor market, for this end, we create a dummy variable equal to 1 if the person is employed and 0 otherwise. Finally, we generate another categorical variable for the number of children (with values equal to: no children, one child, two children and more than two children).

5 Assortative Mating Indicators

The model of Greenwood et al. (2014) seeks to answer the next two questions:

1. Is there any evidence of an increase of Assortative Mating in households during the period 1992-2012?

[10]On this scale, an adult male has a weight equal to 1, an adult female has a weight equal to 0.5, and all children under 14 years old have a weight equal to 0.3.

2. What is the contribution of Assortative Mating to household income inequality?

Question number 1 can be answered looking at a series of indicators, but question number 2 requires a more sophisticated analysis, for this reason we will employ a simulation method.

5.1 Indicator 1: Regression Coefficients

This indicator is based on estimating this equation:

$$EDU^f_{it} = \alpha + \beta EDU^m_{it} + \sum_{t\in\mathcal{T}} \gamma_t * EDU^m_{it} * \text{Year}_{st} + \sum_{t\in\mathcal{T}} \theta_t * \text{Year}_{st} + \epsilon_{it} \quad \epsilon_{it} \sim N(0,\sigma) \tag{9}$$

Where:

EDU$^f_{it}$, EDU$^m_{it}$: are education levels of males (m) and females (f) who constitute household i in year t.

Year$_{st}$: are year dummies $t = 1992, \ldots, 2012$ such that Year$_{st} = 1$ if $s = t$ and Year$_{st} = 0$ otherwise.

Coefficient β measures the strength of the association between spouses' education levels in the base year (1992), while, coefficients γ_t measure the relative contribution of husband's education relative to that of their wives for all years after the base year, this means that they measure the degree of Assortative Mating. Coefficients θ_t are used to control for the exogenous increase in education levels of both males and females.

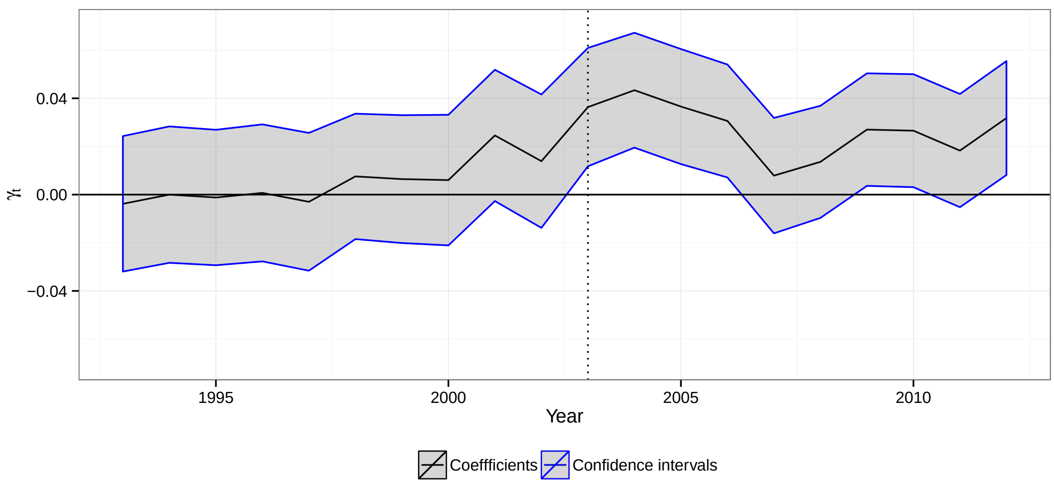


Figure 5: Increase in Assortative Mating

Figure 5 shows a slight upward trend in the range, however coefficients are not significant at 5% because their confidence intervals contain the value of zero for most years. The strength of Assortative Mating is higher in the second half of the period, but much smaller than that for the United States, partly due to the smaller period and higher frequency since this phenomenon is a structural one with very little year to year variation.

5.2 Indicator 2: Kendall's Tau

Kendall's tau is a rank correlation coefficient used to measure the strength of a correlation between two categorical variables, in this case, education level of spouses. This coefficient (Kendall, 1970) is defined as:

$$\tau = \frac{S}{\frac{n(n-1)}{2}} \tag{10}$$

S is the difference between the number of concordant pairs (P) and the number of discordant pairs (Q), a pair of observations is called *concordant* if both husband and wife attained the same education level and *discordant* otherwise, then: $P + Q = \frac{n(n-1)}{2}$, and:

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

Kendall's tau is normally distributed (Worner, 2006), in the range $[-1, 1]$, but this may not be the case if there is a large number of *ties* (pairs that are neither concordant nor discordant.) In the presence of ties, a continuity correction is used to account for them as described in Kendall, 1970, p. 58.

Confidence intervals for this statistic are computed using Fisher's **Z-score**:

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

Its standard error is equal to: $SE(Z) = \sqrt{\frac{1}{n-3}}$, where n is the number of couples in the sample. We can use these standard errors to estimate confidence intervals given that the Z-sore has a normal distribution.

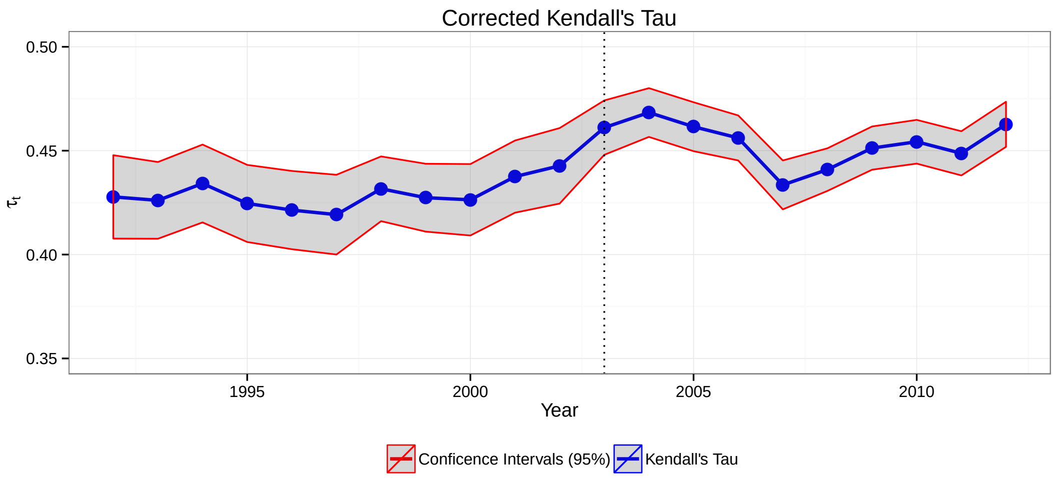


Figure 6: Rank correlation coefficient for couples' education attainment

Following Worner, 2006, and given that the values of Kendall's tau are bounded between -1 and 1, a positive value for the coefficient indicates that both categorical variables increase together, but a negative value means that while one increases, its counterpart decreases. Absolute values of this coefficient can be interpreted as probabilities, for example, a value of -0.25, means that, if we select a random couple, there is a probability of 0.25 that both spouses do not have the same education levels.

There is no consensus of what is a "high" or a "low" value for Kendall's tau, but it is possible to use this informal rule:

$\pm 0.01 \leq \tau < \pm 0.3$	weak correlation
$\pm 0.3 \leq \tau < \pm 0.7$	moderate correlation
$\tau \geq \pm 0.7$	strong correlation

In Figure 6 values of Kendall's tau and confidence intervals for spouses' educational attainment are plotted. Its trend is similar to that of the regression coefficients, but their significance is somewhat higher, the value of τ was almost constant throughout the 1990s. In the 2000s it had a slight increase again, but it does not deviate further than 0.46, having started in 0.42, according to the categories above, it can be stated that correlation is moderate, but showing no clear signs of increasing, unlike the U.S. data from Greenwood et al., 2014.

5.3 Indicator 3: Relative sum of diagonals

Our third indicator of Assortative Mating is originated on the association matrix of couples' educational attainment, if a^t_{ij} is the cell for husband's educational level i and wife's j in year t. Then, the sum of all diagonal entries in matrix $\mathbf{A}$ (values with $i = j$) can be defined as: $\sum_{i=1}^{6} a^t_{ii}$.

On the other hand, let r^t_{ij} be the equivalent cell in a matrix obtained as the internal product of row and column totals of association matrix $\mathbf{A}$, attempting to simulate the association matrix that would have arisen if mating was random. Let $\sum_{i=1}^{6} r^t_{ii}$ be its sum of diagonals, then we can define:

$$\delta_t = \frac{\sum_{i=1}^{6} a^t_{ii}}{\sum_{i=1}^{6} r^t_{ii}} = \frac{tr(\mathbf{A})_t}{tr(\mathbf{R})_t} \tag{11}$$

$tr(\mathbf{A})$ y $tr(\mathbf{R})$ are the traces of both observed and simulated association matrices.

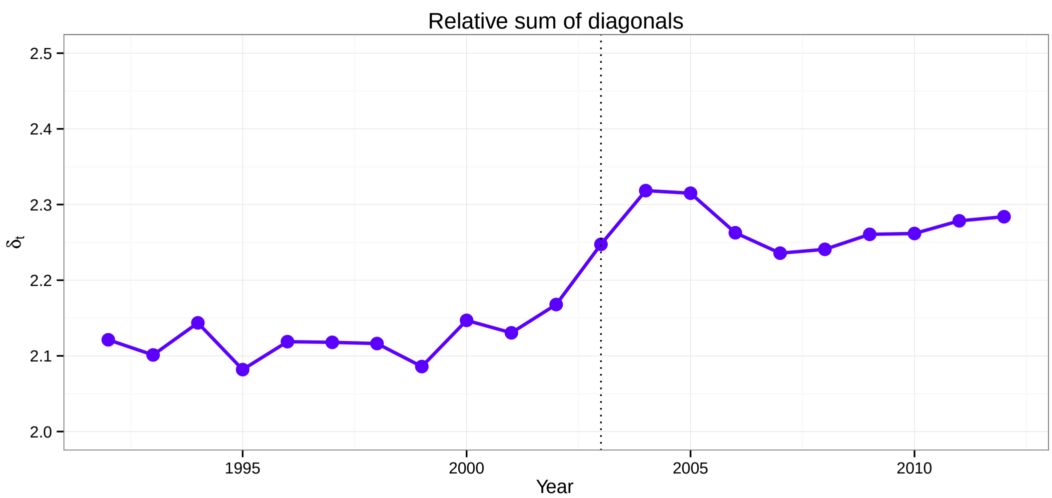


Figure 7: Ratio of sums of diagonal entries of association matrices per year

In Figure 7, values for δ_t per year are plotted, its value is greater then 1 in the entire span of the series, implying positive Assortative Mating. Nevertheless, its behaviour resembles that of previous indicators in both level and rate of variation.

5.4 Indicator 4: Contingency tables

The final Assortative Mating indicator used are association or contingency tables, that is, the joint distribution of all six couples' educational levels by period. In this case the levels for periods 1992-1996 and 2008-2012 are shown below:

Table 2: Contingency table for period 1992-1996

Husband	Wife					
	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**
Marginal dist.	0.296	0.099	0.217	0.185	0.138	0.065

Source: Household surveys (EPH).

Table 3: Contingency table for period 2008-2012

Husband	Wife					
	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**
Marginal dist.	0.188	0.045	0.252	0.158	0.243	0.114

Source: Household surveys (EPH).

Tables 2 and 3 show the joint distribution of each combination of educational levels in period 1 (1992-1996) and 4 (2008-2012). In most cases the highest value of the association matrix coincides with that in the main diagonal, it can also be seen that frequencies of the combinations from the highest education levels increase en period 4 relative to period 1. This shows that average educational level has increased, especially for women, in particular, the frequency of wives with the highest level of education almost doubled between both periods.

6 Inequality indices

Our second stated objective in this paper is to assess the impact of Assortative Mating on income inequality, to achieve this, two inequality indicators are used: the Gini index and Lorenz curves.

Let:

f_{ij} : fraction of type i households in income percentile j: $f_{ij} = \frac{N_{ij}}{N}$, where N is the sample size and N_{ij} is the size of group (i, j).

r_{ij} : income of households i relative to mean household income of the entire population: $r_{ij} = \frac{y_{ij}/N_{ij}}{Y/N}$, y_{ij} is total household income of group (i,j) and Y is the household income for the whole sample.

s_j : share of aggregate income for percentile j: $s_j = \sum_i f_{ij} r_{ij}$.

I_p : cumulative share of percentile p: $I_p = \sum_j^p s_j = \sum_j^p \sum_i f_{ij}$

We depart again from Greenwood et al., 2014, since we are using survey instead of census data, this implies that our sample size will be much smaller, one way to mitigate its impact is reducing the number of percentiles from 10 to 5 (that is, from deciles to quintiles), this allows us to reduce the number of groups without affecting results significantly. However, the analysis will also be carried out with deciles in the "robustness checks" section. Let these quintiles be: $j \in \{0.1; 0.2; 0.3; 0.4; 0.5\}$, then we will have just $768 \times 5 = 3840$ groups, instead of $768 \times 10 = 7680$ groups.

Lorenz curves are defined as the plot of (p, I_p), where $p = \sum_j^p \sum_i f_{ij}$, and $I_p = \sum_j^p \sum_i f_{ij}$. The Gini coefficient equals twice the area under the Lorenz curve and a $45°$ line:

$$g = 2\int_0^1 |I_p - p| dp, \quad \text{with: } 0 \leq g \leq 1 \tag{12}$$

Consequently, both indices are functions of f_{ij} and r_{ij}:

$$l_p = Lorenz_p(\{f_{ij}\}, \{r_{ij}\}) \tag{13}$$

$$g = Gini(\{f_{ij}\}, \{r_{ij}\}) \tag{14}$$

Since we have a very large number of groups, we will use the decomposition formula devised by Rao (1969), which decomposes the coefficient in any arbitrary number of groups, denoted by n:

$$g = \sum_{p=1/n}^{1-1/n} \left[pI_{p+1} - \left(p + \frac{1}{n}\right) I_p \right] \tag{15}$$

Thus, in order to plot the Lorenz Curve we need to split the interval $[0,1]$ into n equally spaced line segments: $j \in \mathcal{J} = \{\frac{1}{n}, \ldots, 1 - \frac{1}{n}\}$, the value of n is the number of quantiles used ($n = 4$: quartiles, $n = 5$: quintiles, $n = 10$: deciles, $n = 100$: percentiles.)

In Figure 8 Lorenz curves and Gini coefficients for periods 1992-1996 and 2008-2012 are plotted. Inequality had a very small variation because these numbers were calculated as an average of years when inequality grew (period 1) and later decreased (Period 4.) In the first period, the poorest 20% earned less than 5% of total income, while the richest 20% earned almost 50% of total income. In the latter period, this shares decreased but not significantly, since now the lowest quintile gets a little over 5% of total income, while the highest quintile gets about 47% of total household income[11].

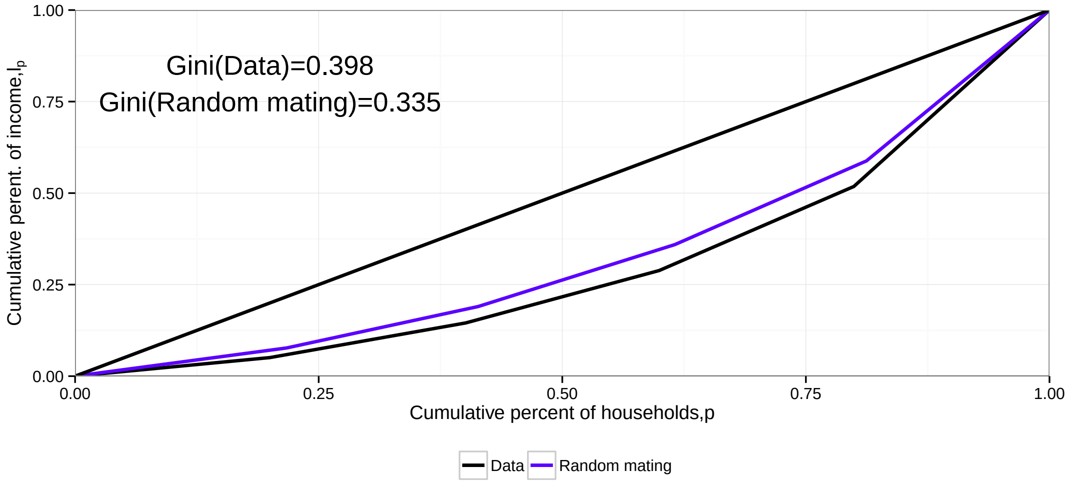
Finally, if we compare the observed inequality with the level that arises after imposing random matching, it can be seen that it falls nearly 6.3 points in period 1 and 3.3 points in period 4.

7 Counterfactual experiments

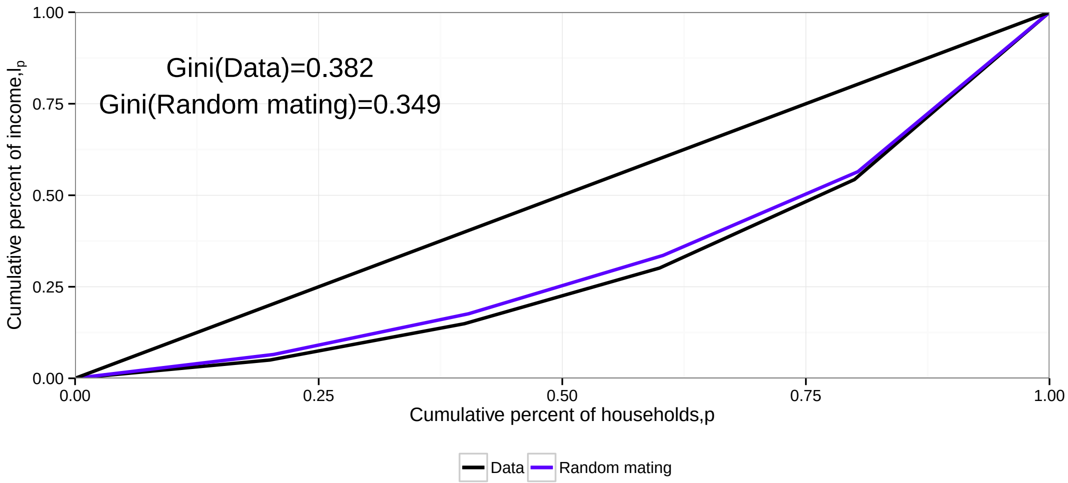
The counterfactual experiments carried out are based on two population characteristics: the joint distribution of education levels of spouses and female labor participation, Gini coefficients are calcualted after changing this distributions.

Another way to write the Gini coefficient is (Cowell and Flachaire, 2015):

[11] "Total household income" always refers to household income with an adult-equivalent correction.



(a) 1992-1996



(b) 2008-2012

Figure 8: Lorenz curves: data versus random mating

$$G = 1 - 2\int_0^1 L(F;q)dq \tag{16}$$

$L(F,q)$ is the q-th ordinate of the Lorenz curve, which depends on the distribution of function of income F, in this way:

$$L(F,q) = \frac{C(F,q)}{\mu(F)} \tag{17}$$

$\mu(F)$ is the average income, (a function of its distribution) and $C(F,q)$ is the cumulated income functional defined as:

$$C(F,q) = \int_{\underline{y}}^{\bar{y}} ydF(y) \tag{18}$$

Joining both terms we obtain an expression for the Gini coefficient:

$$G = 1 - \frac{2}{\mu(F)}\int_0^1 \int_{\underline{y}}^{\bar{y}} y * f(y, E_H, E_W, LFP_W)dydq \tag{19}$$

Finally, replacing both conditional and marginal distributions, we get, for $t = 1,4$:

$$G_t = 1 - \frac{2}{\mu(F)_t}\int_0^1 \int_{\underline{y}_t}^{\bar{y}_t} y * f^t(y/E_H, E_W, LFP_W) * g^t(E_H, E_W/LFP_W) * h^t(LFP_W)dydq \tag{20}$$

With:

$$f(y, E_H, E_W, LFP_W) = f(y/E_H, E_W, LFP_W) \times g(E_H, E_W, LFP_W) \tag{21}$$
$$g(E_H, E_W, LFP_W) = g(E_H, E_W/LFP_W) \times h(LFP_W) \tag{22}$$

All counterfactual experiments depend upon changes of spouse's educational attainment distributions $f(E_H, E_W)$ and changes in female labor participation as a function of quintiles of household income $f(LFP_W)$.

7.1 Random mating

The first counterfactual experiment consists on simulating the result that would have arisen if mating was completely random and then estimating inequality indicators from this scenario. Let $\mathcal{M} = 1,\ldots,576$ be the index used to identify married couples, $\mathcal{S} = 577,\ldots,768$ the index for single and divorced persons, then our counterfactual experiment departs from replacing the f_{ij} for $(i,j) \in \mathcal{M}$ observed for those that would be observed if mating was random: $\tilde{f}_{ij}$ for $(i,j) \in \mathcal{M}$, finally, counterfactual Lorenz curves and Gini coefficients will be:

$$l_p = Lorenz_p\left(\{f'_{ij}\}, \{r_{ij}\}\right) \tag{23}$$
$$g = Gini\left(\{f'_{ij}\}, \{r_{ij}\}\right) \tag{24}$$
$$\{f'_{ij}\} \equiv \{\tilde{f}_{ij}\}_{\mathcal{M}} \cup \{f_{ij}\}_{\mathcal{S}} \tag{25}$$

In terms of joint distributions, the Gini coefficient for periods $i = 1,4$ is now:

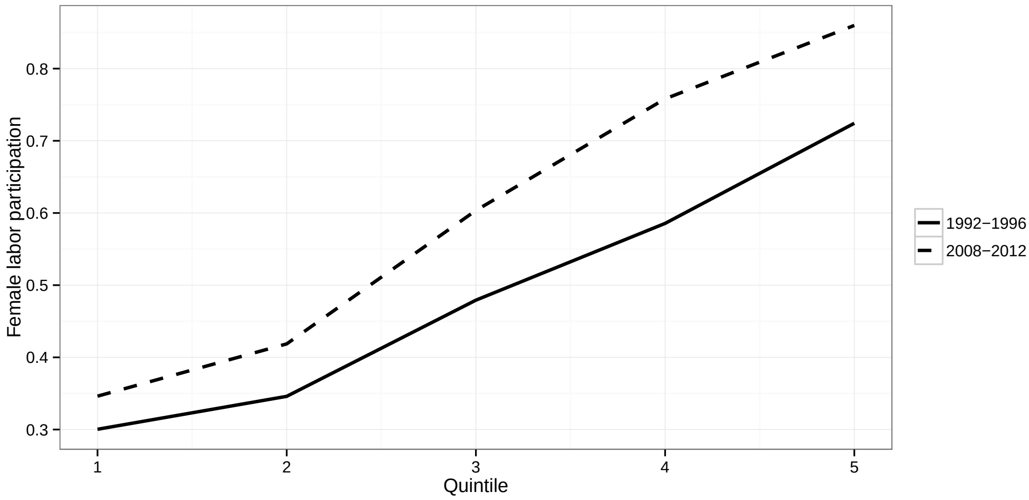
$$G_i^{RM} = 1 - \frac{2}{\mu(F)_i}\int_0^1 \int_{\underline{y}_i}^{\bar{y}_i} yf^i(y/E_H, E_W, LFP_W)g_R^i(E_H, E_W/LFP_W)h^i(LFP_W)dydq \tag{26}$$

$g_R^i(E_H, E_W/LFP_W)$ is the counterfactual joint distribution of education levels assuming random mating, such that differences in Gini coefficients can only be attributed to changes in the distribution of this education level distribution.

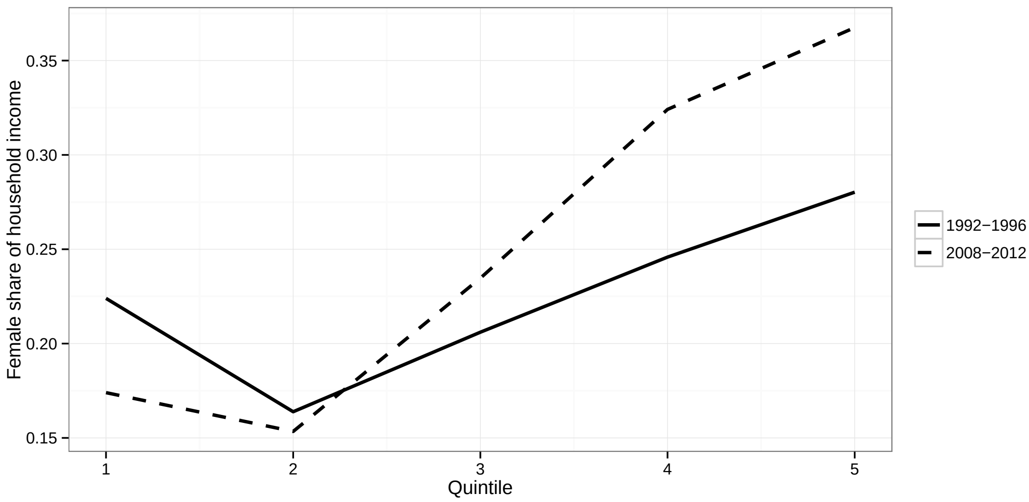
After modifying mating patterns we see that income inequality falls, this result is in line with those of Greenwood et al. (2014), but the effect of random mating is higher in period 1992-1996 than in 2008-2012 (6.3 points versus 3.3.) This contrasts with the slight increase in Assortative Mating shown in the previous section. Contingency tables of this simulation exercise are shown in the appendix (Tables 6 and 7).

7.2 Female labor participation

Other driver of inequality is female labor participation, especially married women, at a first glance, it should reduce intra-household income inequality, but this effect may not be linear along quantiles. On the same token, a greater female share on household income (as a consequence of the former) must not be necessarily linear, either.



(a) Female labor participation



(b) Female share of household income

Figure 9: Married female labor participation: 1992-1996 versus 2008-2012

The effect of Assortative Mating depends on women being employed, in figure 9(a) is shown that this is more frequent in 2008-2012 than in 1992-1996. This increase in participation is larger the higher the quintile of income distribution. This effect is also present in the share of females in household income 9(b) since it is also greater the higher the quintile.

The next counterfactual experiment randomizes mating in both periods, but now women in period 1(4) participate in the labor market according to that of period 4(1). This is equivalent to changing both conditional and marginal distributions of education levels as well as the marginal distribution of female labor participation:

$$G_t^{RM+\Delta LFP} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_t} y * f^t(y/E_H, E_W, LFP_W) * g_R^t(E_H, E_W/LFP_W) h^s(LFP_W) dy dq \tag{27}$$

Where $s = 4$ if $t = 1$ and $s = 1$ if $t = 4$, this way, the aggregate effect is due to variations in two distributions: the joint distribution of education levels and that of female labor participation by quintile of income.

This variation in female labor participation decreases inequality in the first period from 0.398 to 0.334, and reduces inequality in the fourth period from 0.382 to 0.348, Greenwood et al. (2014) argue that the source of this change is the diversification of household income between husbands and wives.

Almost the entire reduction in inequality is due to random mating, because the effect of female labor participation is negligible. This results are at odds with those of Greenwood, who finds that female labor participation reinforces that of random mating. Contingency tables for this experiment are Tables 8 and 9 and their respective Lorenz curves are plotted as 10(a) and 10(b) in the appendix.

7.3 Standardized contingency tables

The last counterfactual experiment is similar to the first one, but we now take into account the fact that simulated contingency tables under random mating do not take into account the increase in husbands' and wives' education levels from 1992-1996 to 2008-2012. This increase affects marginal distributions estimated from these contingency tables, as a solution to this issue, we use **Standardized Contingency Tables** in which their marginal distributions are all equal to $1/6$ by means of an algorithm by Sinkhorn and Knopp using a standardization method proposed by Mosteller, 1968 and summarized in Greenwood et al., 2014, p. 6, but in our case the desired marginal distribution is $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$.

In terms of our notation for the Gini coefficient, this experiment can be expressed as:

$$G_t^{ST} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_i} y * f^t(y/E_H, E_W, LFP_W) * g_E^t(E_H, E_W/LFP_W) * h^t(LFP_W) dy dq \tag{28}$$

$$G_t^{ST+\Delta LFP} = 1 - \frac{2}{\mu(F)_t} \int_0^1 \int_{\underline{y}_t}^{\bar{y}_i} y * f^t(y/E_H, E_W, LFP_W) * g_E^t(E_H, E_W/LFP_W) * h^s(LFP_W) dy dq \tag{29}$$

We also repeat the outcome resulting from variations in female labor participation, as in experiment 2, we label this as experiment 3 (where we impose a standardized contingency table) and 4 (where we impose both a standardized contingency table and female participation of the opposite period) standardized contingency tables and Lorenz curves for this exercise are also shown in the appendix (Tables 10 and 11 and Figures 10(c), 10(d), 10(e) and 10(f)).

7.4 Results

Table 4: Gini coefficients for counterfactual experiments

Experiment	1992-1996 (1)	2008-2012 (4)
Data	39.8	38.2
Random mating in 1	33.5	
Random mating in 4		34.9
Random mating in 1 and ΔFLP	33.4	
Random mating in 4 and ΔFLP		34.1
Standardized table in 1 with marginals from 4	34.6	
Standardized table in 4 with marginals from 1		35.1
Standardized table in 1 with marginals from 4 and ΔFLP	34.5	
Standardized table in 4 with marginals from 1 and ΔFLP		34.2

Source: Own estimations based on Greenwood et al., 2014.

Results from simulations are summarized in Table 4, they are somewhat similar to those found for the United States because most simulations result in a decrease of the Gini coefficient of about 4 or 5 points. With random mating (experiment 1) inequality falls 6.3 and 3.3 points in periods 1 and 4 respectively, if we also allow variations in female labor participation (experiment 2), inequality falls in a similar amount (6.4 points for period 1 and 4.1 points for period 4.)

Experiments 3 and 4, as mentioned, are similar to 1 and 2, but using standardized contingency tables, with the standardized table only (experiment 4) inequality falls 5.2 points in period 1 and 3.1 points in period 4. Finally in experiment 4 we add variations in female participation to the standardized contingency table to find variations of 5.1 points and 4 points in periods 1 and 4.

The results summarized in the table show that assortative mating plays an important role on income inequality, because it explains about 3 to 6 points of it, this is along the same lines as those found by Greenwood et al. (2014) for the U.S.

There is an ongoing discussion regarding the importance of assortative mating on income inequality, the former paper is on one side (those who argue that there is a large impact on inequality), on the other side we find the results by Eika, Mogstad, and Zafar (2014) and Hryshko, Juhn, and McCue (2014) who argue the contrary. This differences in results can be accounted by the fact that these papers use different databases and different decomposition methods, as well as different time spans, nevertheless it can be argued that methodological differences explain these results, this will be explored in the following section.

7.5 Robustness checks

The results of our experiments depend on the way we parametrize the distribution of income and the number of groups we use, since we deviate form the analysis of Greenwood et al. (2014) in those items. These changes were made in order to reduce the number of groups, otherwise a large number of them would have had zero observations turning them useless for the analysis.

We can instead, carry out the analysis using either the same number of quantiles or an equivalent number of groups as Greenwood et al. The first robustness check consists on using deciles instead of quintiles of household income distribution, regardless of the large number of groups that will be left without observations. This check is also useful because it uses original Matlab codes instead of modified codes, thus reducing the likelihood of errors, now we have $768 \times 10 = 7680$ groups.

The second robustness check is based on the exact opposite modification, since we now decrease the number of groups dropping the variable for the number of children, while using income quintiles for grouping again. This modification reduces the number of groups to $128 \times 5 = 640$. After performing this checks we will be able to assess how robust this methodology is to changes in the number of groups, Greenwood et al., 2014 do not perform any of these checks, perhaps because they have millions of observations.

Table 5: Robustness checks: Gini coefficients

Simulation	**Change in quantiles**		**Change in groups**	
Experiment	**1992-1996**	**2008-2012**	**1992-1996**	**2008-2012**
Data	42.6	40.9	40.3	38.3
Random mating in 1	35.2		38.9	
Random mating in 4		36.8		40.0
Random mating in 1 + ΔFLP	35.0		38.6	
Random mating in 4 + ΔFLP		35.8		37.6
Standardized table in 1	36.3		40.1	
Standardized table in 4		37.0		40.8
Standardized table in 1 + ΔFLP	36.2		39.9	
Standardized table in 4 + ΔFLP		36.0		38.4

Source: Own estimations based on Greenwood et al., 2014.

These checks show a number of discrepancies with respect to baseline estimations (Table 4):

- Gini coefficients estimated with observed data increases in 2.8 points (period 1) and 2.7 points (period 4) if the number of quantiles is doubled, on the other hand it also increases in 0.5 points (period 1) and 0.1 (period 4) when we decrease the number of groups. This variations cast a shadow of doubt on the decomposition algorithm, because it should not experience significant variations with respect to the number of quantiles, particularly in the first check.
- Results from experiments also differ substantially because the imposition of random mating causes an **increment** in inequality for period 4, however if we also allow variations in female labor participation, it cancels out this increase. This result is also valid with a standardized contingency table.
- Finally, a reduction in the number of groups makes all variations smaller, while an increase in groups, causes the exact opposite, this means that results not only are not robust, but also an increasing function of the number of groups.

8 Discussion

The outcome of our simulations and robustness checks carried out in previous sections are in stark contrast with respect to the four indicators of Assortative Mating we estimated earlier. The latter show that this phenomenon can explain up to six points of income inequality, as measured by the Gini coefficient, while all indicators show an insignificant increase in the strength of Assortative Mating over the period.

As mentioned earlier, Harmenberg, 2014 shows that the method used in this paper tends to exaggerate the impact of Assortative Mating on the Gini coefficient compared to other methods, we improve his explanation showing that this is actually a consequence of the grouping strategy: the higher the number of groups, the higher the impact estimated.

There are other likely causes for these discrepancies, but their importance is relative to what assumption we think is violated:

Time span : perhaps the most important drawback of our data is the time span we use (21 years) vis-à-vis the one used by Greenwood et al. (45 years), the starting year was chosen because data from years prior to 1992 has serious issues related to income measurement due to the high inflation from the 1980s, a much smaller sample due to the surveying of less urban areas and problems with survey frequencies. As mentioned, we could not use IPUMS data since the Argentine Population Census does not survey household income data. Finally, in order to increase our sample we joined together five consecutive years into two periods (1992-1996 and 2008-20012.)

Method : both Hryshko, Juhn, and McCue (2014) and Eika, Mogstad, and Zafar (2014) found a much smaller impact of Assortative Mating, even with a much larger time span than this paper, this also questions Greenwood's method as the likely cause for the discrepancies.

Nature of Assortative Mating : changes in mating patterns are expected to be slow, that is, year to year variations in any indicator should not be very large (that is the rationale for using census data with a 10-year frequency). Even taking frequency into account, we see that mating patterns in Argentina do not change substantially, it could be argued that mating had already a large degree of sorting prior to 1992, so only increased slightly, therefore, its contribution to inequality is not substantial.

9 Concluding remarks

The purpose of this paper was to assess the importance of Assortative Mating on income inequality in Argentina between 1992 and 2012 because this topic has received very little attention not only in Argentina, but also in Latin America. The evidence we found suggests that Assortative Mating plays a minor role in the determination of income inequality, two findings on this paper support this conclusion:

- we used four indicators of Assortative Mating used by different authors, all of them show little variation in the period 1992-2012.
- a simulation exercise was performed with household survey data, despite finding evidence of a significant contribution of AM to inequality, inequality measured with the Gini coefficient falls by more than 3 points after imposing random mating, but these results are not robust to variations in the number of groups.

Evidence supporting Greenwood's hypothesis of a large impact of sorting on household income inequality is, on our view, little, for the reasons pointed out before, they contradict the results from all simple Assortative Mating indicators and counterfactual experiments are not robust to modifications in the number of groups used. Taken as a whole, this findings weaken the results of Greenwood et al. (2014), while supporting those of Eika, Mogstad, and Zafar (2014) and Hryshko, Juhn, and McCue (2014), who postulate a much smaller contribution of Assortative Mating on income inequality.

These seemingly discouraging findings (at least from a researcher's point of view) are positive if we think of them from a policymaker's point of view because, in case its impact was large, it could not be mitigated with any public policy in a direct way, unlike other sources of inequality.

There is still plenty of room to continue research in this area, since we still need to assess what would happen if we apply this method to data from other Latin American countries or, if we use a different decomposition method.

References

Beccaria, Luis and Fernando Groisman (2008). "Informalidad y Pobreza en Argentina". In: *Investigación Económica* 67.266.

Becker, Gary (1991). *A Treatise on the Family*. English. 1st ed. Cambridge, Massachussets: Harvard University Press. ISBN: 0-674-90699-3.

Bergés, Miriam (2011). "Escala de equivalencias en el consumo para Argentina". PhD thesis. Universidad Nacional de La Plata.

Bredemeier, Christian and Falko Juessen (2013). "Assortative mating and female labor supply". In: *Journal of Labor Economics* 31.3. URL: http://www.jstor.org/stable/10.1086/669820.

Campos-Vázquez, Raymundo, Andrés Hicapié, and Rubén Rojas-Valdéz (2012). "Family Income Inequality and the Role of Married Females' Earnings in Mexico: 1988-2010". In: *Latin American Journal of Economics* 49.1, pp. 67–98.

Cancian, Maria and Deborah Reed (1999). "The Impact of Wives' Earnings on Income Inequality: Issues and Estimates". In: *Demography* 36.2, pp. 173–184.

Cowell, Frank A. and Emmanuel Flachaire (2015). "Chapter 6 - Statistical Methods for Distributional Analysis". In: *Handbook of Income Distribution*. Ed. by Anthony B. Atkinson and François Bourguignon. Vol. 2. Handbook of Income Distribution. Elsevier, pp. 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. and Robert G. Valletta (2006). "Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour". In: *Economica* 73.289, pp. 75–98.

DiNardo, John, Nicole Fortin, and Thomas Lemieux (1996). "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach". In: *Econometrica* 64.5, pp. 1001–1044.

Eika, Lasse, Magne Mogstad, and 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.

Funes Leal, Victor (2015). "Efectos del emparejamiento selectivo sobre la distribución del ingreso en Argentina". Master's thesis. Universidad Nacional de La Plata.

Gasparini, Leonardo, Martín Cicowiez, and Walter Sosa Escudero (2013). *Pobreza y Desigualdad en América Latina. Conceptos, herramientas y aplicaciones*. 1st ed. Buenos Aires: Temas. ISBN: 978-987-1826-45-2.

Greenwood, Jeremy et al. (2014). "Marry Your Like: Assortative Mating and Income Inequality". In: *American Economic Review* 104.5, pp. 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". In: *American Economic Journal: Macroeconomics*.

Harmenberg, Karl (2014). "A Note: The Effect of Assortative Mating on Income Inequality". Institute for International Economic Studies-Stockholm University.

Hryshko, Dmytro, Chinhui Juhn, and Kristin McCue (2014). *Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important is Assortative Matching?* Tech. rep. 8729. Institute for the Study of Labor (IZA).

James, Gareth et al. (2013). *An Introduction to Statistical Learning with applications in R*. New York: Springer-Verlag.

Kendall, Maurice G. (1970). *Rank Correlation Methods*. 4th ed. Londres: Charles Griffin and Co.

Killingsworth, Mark and James Heckman (1986). "Female Labor Supply, A Survey". In: *Handbook of Labor Economics*. Ed. by Robert Layard and Orley Ashenfelter. Vol. 1. Elsevier Science Publishers, pp. 103–204.

Mare, Robert D. (1991). "Five Decades of Educational Assortative Mating". In: *American Sociological Review* 56.1, pp. 15–32.

Mosteller, Frederick (1968). "Association and Estimation in Contingency Tables". In: *Journal of the American Statistical Association,* 63.321, pp. 1–28.

Pagan, Adrian and Amman Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press.

Rao, V. M. (1969). "Two Decompositions of Concentration Ratio". In: *Journal of the Royal Statistical Society. Series A (General)* 132.3, pp. 418–425.

Schwartz, Christine R. (2010). "Earnings Inequality and the Changing Association between Spouses' Earnings". In: *American Journal of Sociology* 115.5, pp. 1524–1557.

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.

10 Annex: contingency tables

Table 6: Contingency table for 1992-1996 with random mating

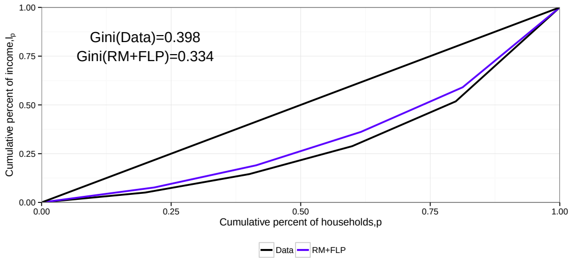
Husband	Wife					
	P-	**P**	**S-**	**S**	**U-**	**U**
P-	0.094	0.031	0.068	0.059	0.044	0.020
P	0.030	0.010	0.021	0.018	0.014	0.006
S-	0.058	0.019	0.042	0.036	0.027	0.012
S	0.063	0.021	0.045	0.039	0.029	0.014
U-	0.032	0.011	0.023	0.020	0.015	0.007
U	0.022	0.007	0.016	0.014	0.010	0.005

Source: Household surveys (EPH).

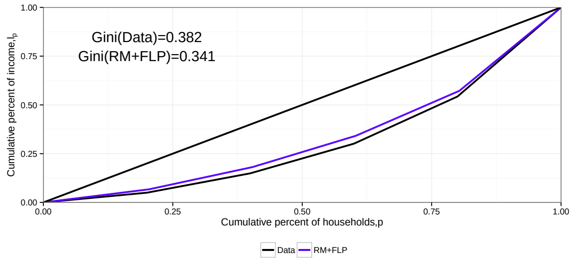
Table 7: Contingency table for 2008-2012 with random mating

Husband	Wife					
	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

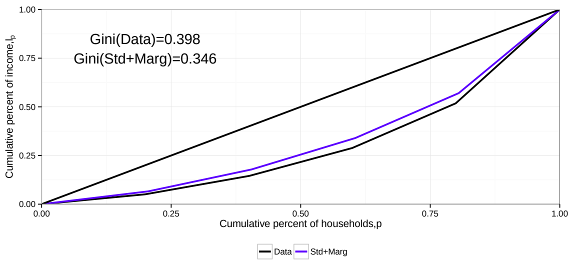
Source: Household surveys (EPH).



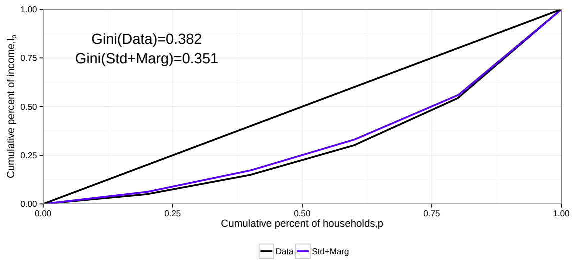
(a) Lorenz curve for experiment 2 (1992-1996)



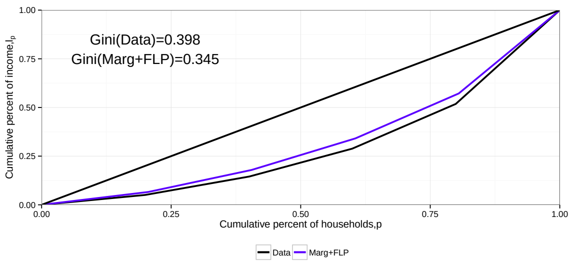
(b) Lorenz curve for experiment 2 (2008-2012)



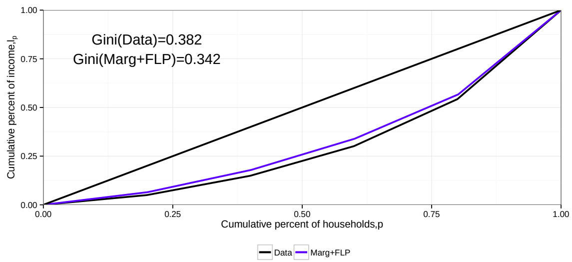
(c) Lorenz curve for experiment 3 (1992-1996)



(d) Lorenz curve for experiment 3 (2008-2012)



(e) Lorenz curve for experiment 4 (1992-1996)



(f) Lorenz curve for experiment 4 (2008-2012)

Figure 10: Lorenz curves for different counterfactual scenarios

Table 8: Contingency table for 1992-1996 with random mating and FLP for 2008-2012

Husband	Wife					
	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

Source: Household surveys (EPH).

Table 9: Contingency table for 2008-2012 with random mating and FLP for 1992-1996

Husband	Wife					
	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

Source: Household surveys (EPH).

Table 10: Standardized contingency table for 1992-1996

Husband	Wife					
	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

Source: Household surveys (EPH).

Table 11: Standardized contingency table for 2008-2012

Husband	Wife					
	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

Source: Household surveys (EPH).