



# Changes in the women's labor market and education and their impacts on marriage and inequality: evidence from Brazil

Sergio Firpo<sup>1</sup> · Lorena Hakak<sup>2</sup>

Received: 28 January 2019 / Accepted: 20 May 2021 / Published online: 21 June 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

From 1992 to 2014, Brazil experienced a decline in income inequality along with a significant increase in schooling level, though the latter was more pronounced among women. Brazil also experienced a decline in returns to education, whereas an opposite trend was observed in several developed countries and China. In this paper, we evaluate the effects of educational, marital, and labor market factors on the income inequality of married couples. We also analyze how changes in educational assortative mating affect their income. Our findings suggest that changes in educational marital sorting parameters had a small but statistically significant effect on household income inequality. We show that growth in female labor force participation and a decrease in the gender wage gap explain part of the decline of the Gini coefficient. Educational factors also explain a part of that decline. Nevertheless, the main driver of the reduction in income inequality among couples appears to be the overall decrease in the returns to education.

---

We are indebted to Luis Araujo, Emerson Marcal, Daniel Monte, David Rubin, Adhemar Villani, and seminar participants at EESP-FGV, SBE-2016, UFABC, FEA/USP, the 2017 Annual Congress of the European Economic Association, and the 2017 Latin American Meeting of the Econometric Society. Special thanks go to Lasse Eika, who generously shared his Stata Codes with us. We thank NEREUS (The University of Sao Paulo Regional and Urban Economics Lab - <http://www.usp.br/nereus/>) for computational support. Financial support of CNPq-Brazil is also acknowledged. Previous title: *Household Income Inequality and Education in the Marriage Market in Brazil: an empirical study.*

---

✉ Lorena Hakak  
lhakak@usp.br

Sergio Firpo  
firpo@insper.edu.br

<sup>1</sup> Insper, Rua Quata 300, CEP: 04546-042 Sao Paulo/SP, Brazil

<sup>2</sup> FEA - Sao Paulo University, Av. Prof. Luciano Gualberto 908 FEA2 room 233, CEP: 05508-010 Sao Paulo/SP, Brazil

**Keywords** Household income inequality · Assortative mating · Marriage market · Returns to schooling · Labor market · Family economics

**JEL codes** J12 · J16 · D31 · I24 · I21

## 1 Introduction

The social norms that shape roles within households have changed in the preceding decades, with important implications for women. The decision to invest in human capital has become more attractive to women, and the division of labor within households has also changed between partners. As a consequence, women have become better educated, have fewer children, participate more in the labor market, and earn wages that have approached those of their male counterparts. Along with these changes, the amount of time spent on household work has decreased due to technological progress.

Because women are studying more on average, there has been a change in the composition of marriages with respect to education in recent years. Educational composition has not only changed how couples are formed but also affected several variables, including returns to education, household income inequality, women's contribution to household labor income, and the number of children (Becker (1991), Chiappori et al. (2009), Pencavel (2006), Charles and Luoh (2010), Goldin and Katz (2002), and Greenwood et al. (2014)).

There is a growing body of the literature investigating how the changes in the role of women have affected income inequality (Greenwood et al. (2014) and Eika et al. (2019)). With the exceptions of Pereira and Santos (2017) and Feng and Tang (2019), studies have focused on developed countries, where income inequality has increased over the past few years.

We contribute to the literature by performing a decomposition analysis based on DiNardo et al. (1996) to measure the impact of changes in the labor market on household income inequality in Brazil.<sup>1</sup> The country is an important emerging market economy where returns to education has been declining in the preceding years. Such a decline has probably been caused by the increase in the proportion of educated men and women, whereas in the USA, Norway, Denmark, Germany, the UK, and urban China the returns to education have risen.<sup>2</sup>

We also study how the impact of changes in social norms, such as the increase in the participation of married women in the labor market and more years of schooling, on average, for women (and greater than men), may have contributed to the decline in household income inequality in Brazil during the period from 1992 to 2014 in a country that experienced a 15% decline in its Gini coefficient over that period.

The goal of this paper is to investigate using Brazilian data whether household income inequality has been affected by changes in female labor force participation and the gender wage gap, as well as by returns to education, educational composition,

<sup>1</sup> In this work, household income inequality considers income of (married or cohabiting) couples.

<sup>2</sup> There was an increase in the proportion of educated men and women in some developed countries as well; however, in Brazil, this phenomenon was more pronounced.

and marital sorting parameters. The decline in household income inequality in Brazil over the recent years raises the interesting question of how these factors explain such a decline.

We analyze changes in the Brazilian household income distribution by examining the marriage market as follows. In the first part, we study how couples are formed in Brazil in relation to education, that is, educational assortative mating patterns, and how such patterns have changed over time, using the methodology of Sinkhorn and Knopp (1967). We then measure how the growth in female labor force participation and a decrease in the wage gap have affected household income inequality.

Moreover, we investigate whether assortative mating patterns might, together with returns to education and educational composition, have a significant effect on household income inequality. To assess the impacts of all these factors on household income inequality, we opt to perform a counterfactual analysis based on the decomposition methods used by Eika et al. (2019) and DiNardo et al. (1996). We then calculate the differences between the actual and counterfactual Gini coefficients. In addition, we calculate a sequential decomposition where we include all five components to analyze the overall change in the Gini coefficient.

Our main results can be summarized as follows. The role played by married women in the labor market in recent years, considering female labor force participation and the wage gap, can explain a part of the decline in the Gini coefficient. The 2014 Gini coefficient would have been higher if it were not for the observed changes in the form of increased participation of married women in the labor market and the narrowing of the gender wage gap. In addition, changes in educational assortative mating had an approximately 0.01 point effect on income inequality in 2014. This effect is small but statistically significant. Our results differ from those obtained for developed countries, where assortative mating has no effect on household income inequality.

Random matching plays an interesting role in this study. If men and women married randomly in regard to education, the Gini coefficient in 2014 would be 0.05 points lower than it actually is (0.449 instead of 0.499). This result is consistent with those for other countries, presented by Eika et al. (2019) and Feng and Tang (2019). Changes in educational composition also have an impact on the Gini coefficient. In 2014, if educational composition were kept as it was in 1992, the Gini coefficient would be 0.02 points lower (0.487 instead of 0.499). Nevertheless, returns to education play a major role in the decline of the Gini coefficient. In 2014, the Gini coefficient (0.499) was 0.04 points lower than the 1992 counterfactual value (0.534), that is, what the Gini coefficient would be if the return to education had not declined. All of these results are statistically significant.

Applying a sequential decomposition, we find a difference in the actual Gini coefficient in 1992 from 0.571 to 0.551 (almost 4%) related to all five counterfactual components combined. We interpret this result as a consequence of changes in Brazil, such as a reduction in the female wage gap, an increase in the participation of women in the labor market, an improvement in education rates, and a decline in returns to education.

The paper is divided into the following sections. The next section contains a literature review and stylized facts. In the third and fourth sections, we describe the possible determinants of income inequality and the relevant methodology. In the fifth

section, the data set is discussed, and in the next section, the results are given. In the last section, we present the conclusions.

## 2 Literature review and stylized facts

Developed and developing countries have recently exhibited different income inequality patterns. It is important to understand in greater depth how these differences can be explained by the differential effects of changes in social norms and other labor market aspects. Recent increases in the Gini coefficient in countries such as the USA, Norway, Denmark, Germany, and the UK have not been observed in less developed countries, such as most countries in Latin America, including Argentina, Chile, and Brazil. Moreover, these less developed countries have experienced a major increase in the schooling levels of their workforces in recent decades. As shown by Barro and Lee (2013), the average number of years of schooling (at age 15 or above) in less developed countries rose between 1990 and 2010 from 5.79 to 8.26, a 43% increase, whereas in developed countries, the increase was from 9.56 to 11.03, a 15% increase. The proportion of graduate students increased by 57% and that of high school graduates doubled in Latin America, whereas in developed countries the increases were by 25% and 45%, respectively. Considering the average years of schooling by gender in Latin America, educational attainment grew faster for women than for men, namely by 44% and 41%, respectively.

In addition, the Gini coefficient in Latin America was 0.04 lower in 2010 than in 2000, declining from 0.54 to 0.50.<sup>3</sup> In some countries, the decline in the Gini coefficient was even more pronounced. According to World Development Indicators, the Gini coefficient in Colombia declined from 0.577 in 2001 to 0.535 in 2014; in Brazil, it declined from 0.593 to 0.515, in Argentina from 0.533 to 0.427, and in Chile from 0.555 to 0.505.

Educational composition can have an impact on different aspects of society. One question that arises is whether changes in educational composition alter assortative mating patterns and whether these patterns have remained stable. An increase in the number of educated men and women can increase the probability of a college-educated man marrying a college-educated woman. A decline in the educational gender gap increases the probability of an educated wife marrying an educated husband even if the pattern of assortative mating has not changed. In this context, homogamous marriages could affect household income distribution.<sup>4</sup>

Moreover, the role of women in intrahousehold work and the labor force has changed over time. The rise in female labor force participation and the decline in the wage gap between husbands and wives, together with the fact that women are on average more educated than men, may also have had an impact on the household income distribution.

Over the recent decades, the Brazilian society has changed in many aspects. During the 1980s, the Brazilian economy faced hyperinflation, stabilization plans failed, and

<sup>3</sup> Article "Gini back in the bottle," published in 10/13/2012 by The Economist.

<sup>4</sup> A married couple with similar traits, including educational level, religion, ethnicity, and socioeconomic status, can be said to be in a homogamous marriage.

public finances were disorganized. This period is known as the “lost decade.” In the middle of the 1990s, after effective economic stabilization and reforms, the economic environment improved.

College accessibility has improved since the 1990s in two ways: An increase in the number of colleges has enabled an expansion of higher education vacancies, and there has been an increase in public funding for tuition and fees that alleviate credit constraints. These two actions have made college degrees more accessible to a wider population (Sampaio (2015)).

On the other hand, returns to education in Brazil have declined.<sup>5</sup> This change is a key element in explaining income inequality. In contrast, returns to education have risen in developed countries such as the USA. However, apart from the decline in the schooling premium in Brazil, the country has also experienced an important reduction in other gaps, such as gender, racial, and geographic wage gaps, as documented by Ferreira et al. (2017).

Greenwood et al. (2014) investigate whether assortative mating in terms of education can have an effect on household income inequality. The researchers analyze assortative mating and income inequality using US data covering the period from 1960 to 2005. They estimate the pattern of assortative mating for husbands' and wives' educational attributes and compare the actual Gini coefficient to a counterfactual one with random matching. The researchers observe that the difference between the actual Gini coefficient and that resulting from random matching was small in 1960, but the difference was significant in 2005. In their work, if marital sorting in 2005 were the same as in 1960, the Gini coefficient would have declined from 0.43 to 0.35.

The researchers also analyze the effect on income inequality of the change in female labor force participation. A counterfactual exercise that considers couples matched in 2005, conforming to the standardized mating pattern of 1960, shows that inequality would be reduced only if considering married female labor force participation in 2005. This finding may be evidence that the contribution of women to household income has an impact on household income inequality.

Eika et al. (2019) estimate counterfactual Gini coefficients for the USA, Norway, Denmark, Germany, and the UK, creating counterfactual scenarios built using the decomposition method developed by DiNardo et al. (1996).<sup>6</sup> The authors analyze the importance of three factors in household income inequality: returns to education, educational composition, and marital sorting parameters. The main factor responsible for the rise in couples' income inequality in the USA and Norway is the increase in returns to education in these countries.

Pereira and Santos (2017) develop a study involving random matching, assortative mating, and household income inequality in Brazil. The researchers apply the methodology of Greenwood et al. (2014) to Brazilian census data and observe that positive assortative mating explains part of the high level of income inequality in Brazil but cannot explain most of the changes over time. This is the closest study to ours even though there are important differences from our contribution. We use an annual household survey (and not decennial censuses) and investigate the role of

<sup>5</sup> See Barros et al. (2007), Menezes-Filho et al. (2007), and Ferreira et al. (2017).

<sup>6</sup> The covered period varies by country and extends until 2013.

other changes between genders, especially in education and in the labor market, in explaining changes in income inequality.<sup>7</sup>

All the changes in the economic environment, demographics, educational composition, and returns to education constitute important determinants that affect household income inequality.

### 3 Possible determinants of income inequality

We choose to analyze three different aspects, namely marriage, education and labor supply, that we consider to be important determinants of household income inequality. Following Ferreira et al. (2017), Barros et al. (2007), Menezes-Filho et al. (2007), and Eika et al. (2019), a population's education characteristics, such as educational composition and returns to education, are a key determinant of income inequality.

Becker (1991) and Browning et al. (2014) explore the relationship between marriage and the dynamics of the family in the context of labor supply and education. "Marital status is strongly correlated with the allocation of work time and the market wages that individuals receive. Thus, compared with singles, married men work more in the market and have higher wages, while married women work less in the market, receiving lower market wages. This pattern may result from two different effects."<sup>8</sup> The first entails the division of labor between spouses, and the second concerns the marriage market itself. However, the pattern of educational attainment has been changing in recent decades in many countries around the world. Women have achieved higher levels of education and entered the labor market, which had direct implications for household division of labor between spouses.

There has been a change in intrafamily bargaining due to changes in technology such as the birth control pill and home appliances' improvements as well as changes in law such as an improved maternity leave rule and payments allowing women to share childbearing and household responsibilities with men. Marriage and education can be better described as complementary rather than mutually exclusive. Wives are becoming less specialized in household work, whereas husbands are spending more time in household work, and both husbands and wives have entered the job market.<sup>9</sup>

Therefore, marital sorting parameters, the wage gap and female labor supply together with educational composition and returns to education are possible candidates for explaining household income inequality.

### 4 Methodology

In this section, we describe the methodology used in the paper to estimate the marital sorting parameters and the counterfactual scenarios for household income inequality.

<sup>7</sup> Data for China are discussed in Feng and Tang (2019).

<sup>8</sup> Browning et al. (2014).

<sup>9</sup> Barbosa (2019) presents data showing that the hours spent per week by women in household work decreased (31 to 24 hours) between 2001 and 2015 in Brazil, whereas among men it rose from 5 to 6 hours a week.

#### 4.1 Educational assortative mating

The first part of our empirical study is based on techniques introduced by Eika et al. (2019). The authors estimate a ratio of the joint probability of marriage and education and that of a random marriage in education, thereby computing the marital sorting parameter. This measure,  $s_{ij}$ , is the ratio of the joint probability that a man with educational level  $j$  is married to a woman with educational level  $i$ , and the probability under random matching of educational levels:

$$s_{ij} = \frac{\Pr(Wife = i \cap Husband = j)}{\Pr(Wife = i) \cdot \Pr(Husband = j)}. \quad (1)$$

The marital sorting parameter shown in expression (1) is a measure of independence in the decision to get married. If  $s_{ij}$  is equal to one, the decision to get married is independent of education. If  $s_{ij}$  is greater (or less) than one, the probability of a married couple having the same levels of schooling ( $i = j$ ) is more (or less) than that of a random marriage in terms of educational levels. This is known as positive (or negative) assortative mating.<sup>10</sup>

#### 4.2 Decomposition method

To create our counterfactual household income inequality scenarios, we use the decomposition method proposed by DiNardo et al. (1996).<sup>11</sup> In the next three subsections, we produce counterfactual scenarios considering five possible changes (one change at a time) in the following attributes: female labor force participation, male–female wage gap, marital sorting parameters or random matching, returns to education and educational composition. We denote these attributes by  $z$ . We use counterfactual scenarios to analyze the impacts of these variables on couples' income distribution.<sup>12</sup>

The income distribution is given by

$$F_y(y|t) = \int_{z \in \Omega_z} F_{y,z}(y|z, t_z = t) dF_z(z|t_z = t). \quad (2)$$

The counterfactual density is given by

$$\begin{aligned} F_y(y|t_y = t, t_z = b) &= \int F_{y,z}(y|z, t_z = b) dF_z(z|t_z = b) \\ &= \int F_{y,z}(y|z, t_z = b) \Psi_z(z) dF_z(z|t_z = t). \end{aligned} \quad (3)$$

<sup>10</sup> The levels of schooling are described in the Data section.

<sup>11</sup> Household income inequality here refers to couples' income inequality, and our sample is composed only of couples (married or cohabiting). The decompositions developed in this study disregard the effects of the general equilibrium.

<sup>12</sup> We follow the notation used in DiNardo et al. (1996).

The reweighting function  $\Psi_z(z)$  is defined as

$$\Psi_z(z) = \frac{dF_z(z|t_z = b)}{dF_z(z|t_z = t)}, \quad (4)$$

where  $t_z$  denotes the attributes  $z$  in a certain year,  $b$  is the base year,  $\Omega_z$  is the domain of the definition of an individual attribute,  $F_y(y|t)$  is the couples' income distribution,  $F_{y,z}(y|z, t_z = t)$  is the conditional distribution of couples' income on  $z$ , and  $dF_z(z|t)$  is the marginal distribution of the spouses' attributes  $z$  in year  $t$ .<sup>13</sup>

#### 4.2.1 Female labor force participation

We construct a reweighting function for female labor force participation as follows:

$$F_y(y|t) = \int \int F_{y,l,e_{ij}}(y|l, e_{ij}, t_{e_{ij}} = t, t_l = b; z_1 = t) \Psi_{l|e_{ij}} dF_l(l|e_{ij}, t_{l|e_{ij}} = t) dF_{e_{ij}}(t_{e_{ij}} = t), \quad (5)$$

where  $e_{ij}$  denotes a marriage of a woman with educational level  $j$  with a man with educational level  $i$ ,  $l$  is assumed to be 1 if a woman works and zero otherwise, and  $F_{y,l,z_1}(y|l, z_1, t_{z_1} = t, t_l = b)$  is the conditional distribution of income on  $l$  and  $z_1$  (which is a vector of all attributes but female labor force participation and spouses' level of education).

Suppose that we want to estimate couples' income inequality in year  $t$  but that female labor force participation is known in year  $b$ . Then, our counterfactual will follow equation (5), where  $\Psi_l$  is a reweighting function defined as follows:

$$\Psi_{l|e_{ij}}(l|e_{ij}, t_{l|e_{ij}}) = \frac{dF_l(l|e_{ij}, t_{l|e_{ij}} = b)}{dF_l(l|e_{ij}, t_{l|e_{ij}} = t)}. \quad (6)$$

To obtain the reweighting function  $\Psi_l$ , we calculate the following equations:

$$\Psi_{l|e_{ij}}(l|e_{ij}, t_{l|e_{ij}}) = l * \frac{\Pr(l = 1|e_{ij}, t_{l|e_{ij}} = b)}{\Pr(l = 1|e_{ij}, t_{l|e_{ij}} = t)}, \quad (7)$$

and

$$\Psi_{l|e_{ij}}(l|e_{ij}, t_{l|e_{ij}}) = (1 - l) * \frac{\Pr(l = 0|e_{ij}, t_{l|e_{ij}} = b)}{\Pr(l = 0|e_{ij}, t_{l|e_{ij}} = t)}. \quad (8)$$

In this exercise, we estimate what couples' income inequality would have been if female labor force participation were fixed in year  $b$  instead of year  $t$ . The difference between the counterfactual exercise and the actual one represents the impact that female labor force participation has on couples' income inequality.

<sup>13</sup> From now on, we omit the integration domain.



### 4.2.2 Female wage gap

To estimate the dummy wage gap, we perform the following steps.

First, we calculate the average female and male wages ( $\gamma_f, \gamma_m$ ) considering wives' and husbands' levels of education. Then, we estimate the ratio of average female and male wages for all educational level combinations:

$$\gamma(wage|e_{ij}, t = b) = \frac{\gamma_f(wage|e_{ij}, t_w = b)}{\gamma_m(wage|e_{ij}, t_w = b)}. \quad (9)$$

Next, we calculate  $\vartheta$ , which is a product of  $\gamma(wage|e_{ij}, t = b)$  and husbands' wages determined considering all couples and all educational level combinations for years  $t$  and  $b$ :

$$\vartheta = \gamma(wage|e_{ij}, t = b) * husband\_salary(wage|e_{ij}, t_w = t). \quad (10)$$

We consider a dummy wage gap variable  $\omega$ , assumed to be 1 if a woman with educational level  $i$  earns a wage greater than  $\vartheta$  and equal to 0 otherwise. We repeat the process for year  $b$ , calculating equation (10) with  $t_w = b$ . To construct the counterfactual scenario, we follow equations (5)-(8), considering our dummy variable  $\omega$  instead of  $w$ .

### 4.2.3 Marital sorting parameters, educational composition, and returns to education

In this section, we will present other counterfactual scenarios based on Eika et al. (2019). We create counterfactual scenarios considering four possible changes (with one change occurring at a time) in returns to education, marital sorting parameters, random matching and educational composition. The distribution of income in year  $t$  can be written as

$$F_y(y|t) = \int F_{y,s}(y|s, t_s = b; z_1 = t) \Psi_s dF_s(s|t_s = t), \quad (11)$$

where  $F_{y,s}(y|s, t_s = b; z_1 = t)$  is the conditional distribution of income on the marital sorting parameter. We allow all variables<sup>14</sup> to change over time represented by  $z_1$ , and hold the marital sorting parameter fixed, considering a base year, to create each counterfactual scenario. To analyze the random matching counterfactual scenario, we consider the marital sorting parameter equal to one.

The reweighting function  $\Psi_s(s)$  is defined as

$$\Psi_s(s) = \frac{dF_s(s|t_s = b)}{dF_s(s|t_s = t)}. \quad (12)$$

To obtain the educational composition counterfactual, we replace  $s$  by  $e_{ij}$  in equation (11). In this case, we fix the educational composition of men and women in year  $b$  and the other attributes in year  $t$ .

<sup>14</sup> Female labor force participation, female wage gap, educational composition and returns to education.

$$F_y(y|t) = \int F_{y,e_{ij}}(y|e_{ij}, t_{e_{ij}} = b; z_1 = t) \Psi_{e_{ij}} dF_{e_{ij}}(e_{ij}|t_{e_{ij}} = t). \quad (13)$$

The reweighting function  $\Psi_{e_{ij}}(e_{ij})$  is defined as

$$\Psi_{e_{ij}}(e_{ij}) = \frac{dF_{e_{ij}}(e_{ij}|t_{e_{ij}} = b)}{dF_{e_{ij}}(e_{ij}|t_{e_{ij}} = t)}. \quad (14)$$

The reweighting function is the ratio of the proportion of couples that would have been formed with educational levels  $x$  if the marginal distributions of the educational composition of men and women were formed in  $b$ , marital sorting were performed in  $t_s = t$  considering income's levels in year  $t_y = t$ . For example, consider an increase in the proportion of men and women with undergraduate (or graduate) degrees in the population. The greater this increase, the greater is the probability of forming randomly couples with this level of education. In this case, we want to analyze what is the impact on income inequality in year  $t$  when fixing this probability in year  $b$ .

We can rewrite equation (14) as

$$\Psi'_{e_{ij}}(x|t_{e_{ij}}, t_s, t_y) = \frac{P'_x(t_{e_{ij}} = b, t_s = t)}{P_x(t_y = t)}. \quad (15)$$

To obtain the reweighting function  $\Psi'_{e_{ij}}$  and then the counterfactual household income distribution, we need to follow several steps of Eika et al. (2019). In particular, we need to estimate  $P'_x(t_{e_{ij}} = b, t_s = t)$ .<sup>15</sup>

The distribution of income in year  $t$  is derived as

$$F_y(y|t) = \int F_{y,e_{ij}}(y|e_{ij}, t_y = t; z_1 = t) \Psi_y dF_{e_{ij}}(t_{e_{ij}} = t), \quad (16)$$

where  $F_{y,e_{ij}}(y|e_{ij}, t_{e_{ij}} = t; z_1 = t)$  is the conditional distribution of income for couples with level of education  $e_{ij}$  in year  $t$ . Moreover,  $F_{e_{ij}}(t_{e_{ij}} = t)$  is the joint distribution of spouses' level of education in year  $t$ . The expression (16) represents the counterfactual income distribution when considering  $t_y = b$  and the term  $\Psi_y(y)$  is the reweighting function for the return to education counterfactual.<sup>16</sup> This counterfactual measures the mean difference in income between couples with a certain level of education  $i, j$  in relation to the base level. In this case, we want to analyze what is the impact on income inequality in year  $t$  when fixing this mean difference in year  $b$ .

#### 4.2.4 Sequential decomposition using counterfactuals

Following DiNardo et al. (1996), we perform a sequential decomposition analysis to estimate the difference in income inequality measures between 1992 and 2014.

<sup>15</sup> We follow Eika et al. (2019)'s algorithm to estimate  $P'_x$  as shown in "Appendix B". This approach is different from DiNardo et al. (1996) who estimate the reweighting function  $\Psi_x$  using a probit model.

<sup>16</sup> Returns to education are evaluated in terms of couples' education level unlike what we usually find in the literature, that is, by analyzing returns to education at the individual level.

Consider the income distribution represented by equation (17):

$$F_y(y|t) = F(y; t_s = t, t_{eij} = t, t_y = t, t_w = t, t_l = t). \quad (17)$$

Next, consider the following sequential decomposition based on the analysis of the change in couples' income between 1992 and 2014:

$$\begin{aligned} & F_y(y|t = 2014) - F_y(y|t = 1992) \\ &= F(y; t_s = t, t_y = t, t_{eij} = t, t_w = t, t_l = t) \\ &\quad - F(y; t_s = b, t_y = t, t_{eij} = t, t_w = t, t_l = t) \\ &\quad + F(y; t_s = b, t_y = t, t_{eij} = t, t_w = t, t_l = t) \\ &\quad - F(y; t_s = b, t_y = b, t_{eij} = t, t_w = t, t_l = t) \\ &\quad + F(y; t_s = b, t_y = b, t_{eij} = t, t_w = t, t_l = t) \\ &\quad - F(y; t_s = b, t_y = b, t_{eij} = b, t_w = t, t_l = t) \\ &\quad + F(y; t_s = b, t_y = b, t_{eij} = b, t_w = t, t_l = t) \\ &\quad - F(y; t_s = b, t_y = b, t_{eij} = b, t_w = b, t_l = t) \\ &\quad + F(y; t_s = b, t_y = b, t_{eij} = b, t_w = b, t_l = t) \\ &\quad - F(y; t_s = b, t_y = b, t_{eij} = b, t_w = b, t_l = b). \end{aligned} \quad (18)$$

Using sequential decomposition, we intend to analyze the five counterfactuals we have presented in Sects. 4.2.1 to 4.2.3, adding one at a time. Doing so will let us obtain the total effect due to these components on household income inequality. The sum of the effects of all components has to be equal to the change in income inequality between 1992 and 2014.

The partial effects are not invariant to the order in which each component is added. The sequence is first generated using one factor; then, we add one more, that is, we obtain the joint effect of two factors. This process continues until all factors have been included. Thus, we add a sequential decomposition approach that simultaneously takes into account all explanatory factors.

In this work, we chose to start the counterfactual analysis with the marriage variable because we analyze couple's income, followed by the education variables and finally the labor market variables. An important observation is that the same value of the Gini coefficient is ultimately obtained regardless of the order.<sup>17</sup>

As highlighted by DiNardo et al. (1996), one disadvantage of sequential decomposition is that the "effect" we find in a specific component "generally depends on the order of decomposition." Therefore, to avoid enhancing the impact of marital variables, we also generate the decomposition in the reverse order. All in all, the net effect after including all five counterfactuals components is observed in the Gini coefficient.

<sup>17</sup> DiNardo et al. (1996) began its sequential decomposition with institutional and labor markets factors, which are the main components they wanted to analyze.

## 5 Data

We use the Brazilian National Household Sample Survey (PNAD) for years 1992, 1999, 2001, 2005, 2009, and 2014.<sup>18</sup> We analyze couples (married and cohabiting) with husbands and wives aged between 26 and 60. We set the lower bound to 26 years of age because by that age a person has completed studies and the upper bound to 60 years of age because at that age a person can retire.

This dataset does not include data on marital status. Therefore, we use data available in the questionnaires about the residents and form couples (either married or cohabiting) within every family unit. We consider couples where one person is the head of household and the other is a spouse or a partner of the opposite sex.

Educational levels are divided into five mutually exclusive groups. The first group comprises individuals who are illiterate (have less than one year of education); the second consists of those with elementary school education (4–5 years of schooling); the third consists of those with middle school education (8–9 years of schooling); and the fourth of those who dropped out of and those who graduated from high school (10–12 years of schooling). The last group comprises individuals with at least some postsecondary education regardless of whether they earned an undergraduate degree, M.A., Ph.D., or no degree at all (more than 12 years of schooling).

All variables are averaged by using household sampling weights, except for years 1992 to 1995 when we use individual sampling weights. For 2014, we have a sample with 50,739 couples.<sup>19</sup>

### 5.1 Descriptive statistics

Table 1 shows the percentage of homogamous couples at each level of schooling. We observe a change in the educational composition of couples in Brazil from 1992 to 2014. In 1992, 64% of couples were formed by illiterate spouses or those with only elementary schooling; 14% of couples had middle school education; 13% had high school-level education; and 9% had some college education. By 2014, these percentages had changed to 20%, 20%, 39%, and 21%, respectively. The proportion of assortative couples decreased from 54% to 47%. The proportion of uneducated assortative couples declined, whereas that of educated couples rose to 21% of all assortative couples in 2014.<sup>20</sup> Among the entire sample, the percentage of educated couples increased from 4.5% in 1992 to 10% in 2014, as shown in Fig. 1.

The decline in the proportion of assortative couples in recent decades may be evidence of the change in educational composition of the population in Brazil. In 1992, almost 70% and 71% of men and women, respectively, had less than 5 years of schooling, while in 2014 this proportion was 21% for women and 27% for men, as

<sup>18</sup> The microdata and questionnaires are available at [www.ibge.gov.br](http://www.ibge.gov.br).

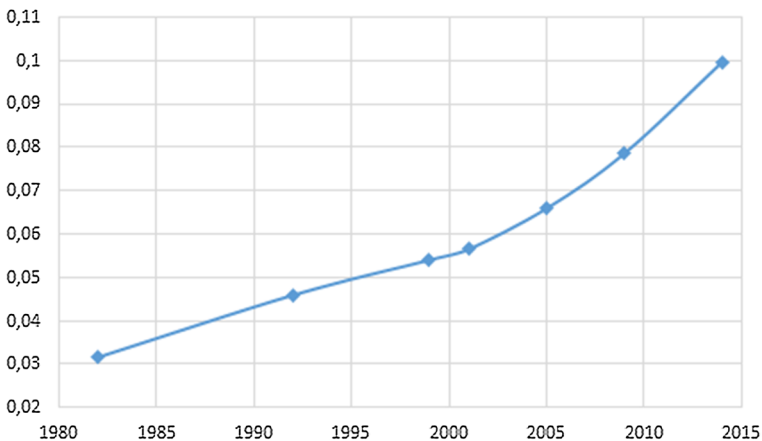
<sup>19</sup> Between 1992 and 1995, only individual sampling weights were available. “Appendix D” provides a more complete description of the dataset.

<sup>20</sup> We regard individuals with less than 5 years of schooling as uneducated and those with more than 12 years of schooling as educated.

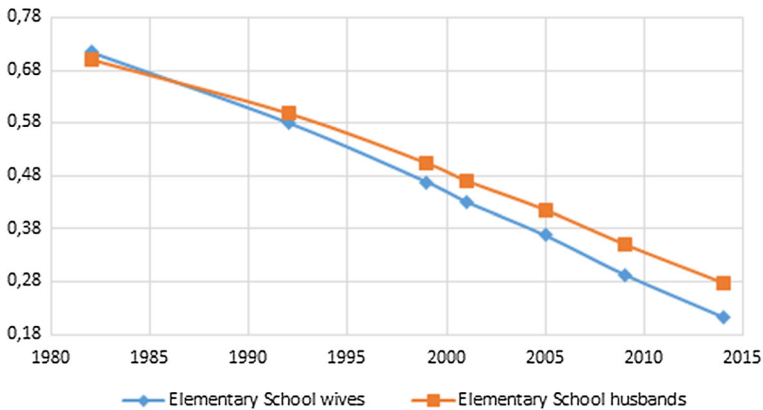
**Table 1** Couples with the same level of schooling, 1992 and 2014 (PNAD)

Level of schooling	Couples	Percentage of assortative couples	Percentage of total couples
	1992		
Illiterate	3535	0.17	0.09
Elementary school	9902	0.47	0.25
Middle school	3005	0.14	0.08
High school	2661	0.13	0.07
College and graduate school	1842	0.09	0.05
Total	20,945	100%	54%
	2014		
Illiterate	1187	0.05	0.02
Elementary school	3553	0.15	0.07
Middle school	4748	0.20	0.09
High school	9248	0.39	0.18
College and graduate school	5000	0.21	0.10
Total	23,736	100%	47%

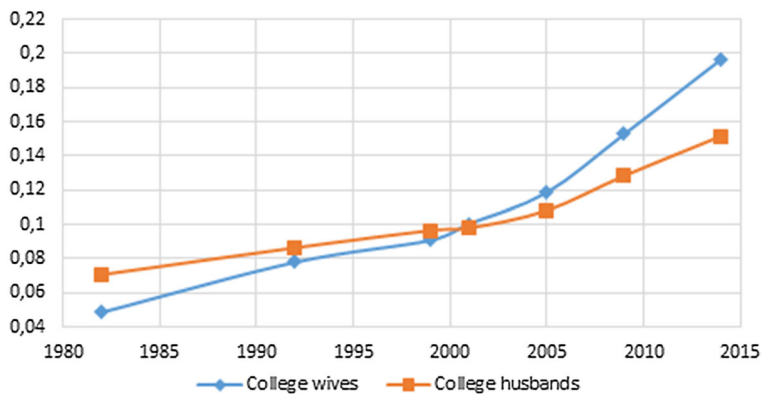
The third column shows the percentage of assortative couples by level of schooling, and the fourth column presents the percentage of assortative couples in the whole sample. The total (aggregate) number of couples reported in this table represents 54% (respectively, 47%) of the total number of couples in 1992 (respectively, 2014). In both cases, we use sampling weights. The age of spouses considered here is between 26 and 60 years. In this table, we use sample weights



**Fig. 1** Proportion of couples in which both spouses are educated (PNAD). Notes: Here, we consider individuals with more than 12 years of schooling to be educated. This figure illustrates the proportion of all educated couples among all couples in the sample



**Fig. 2** Proportion of uneducated men and women (PNAD). Note: We consider individuals with less than 5 years of schooling to be uneducated

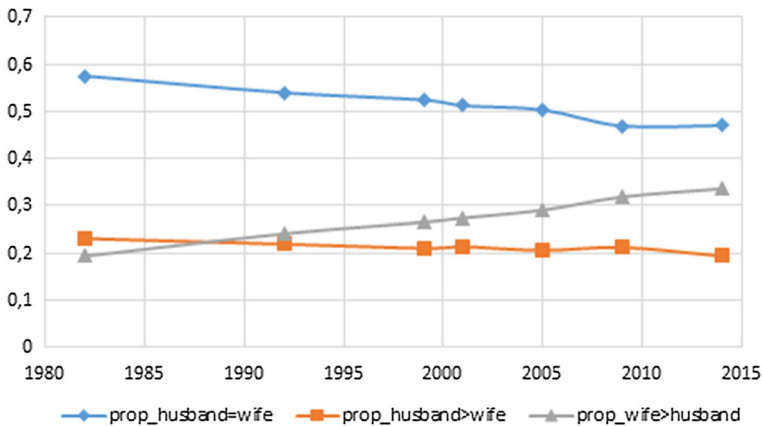


**Fig. 3** Proportion of educated men and women (PNAD). Note: We consider individuals with more than 12 years of schooling to be educated

shown in Fig. 2. Before 2014, the chance of finding an assortative uneducated couple was very high compared to that at present.

The number of educated people grew in the preceding decades, but the rise has been faster for women, as shown in Fig. 3. The proportion of women with a college degree increased from 5 to 20% and that for men rose from 7 to 15% in the same period. Presently, there are more educated women than men, and as shown in Fig. 4, the percentage of assortative couples decreased from 58% in 1992 to 47% in 2014. This result may indicate that women are marrying down, implying an increase in the number of mixed marriages in relation to education.

Figure 4 displays the proportion of couples with the same level of schooling, and those of couples with husbands more educated than wives, and wives more educated than husbands. The proportion of couples with wives more educated than husbands grew from approximately 20% in 1982 to 34% in 2014, but the fraction of couples with husbands more educated than wives decreased from 23% to 20% in the same



**Fig. 4** Proportion of educational attainment of wives and husbands (PNAD). Notes: This figure displays the proportions of couples with the same level of schooling, couples with husbands with more years of schooling than their wives', and couples with wives with more years of schooling than their husbands'

years. Changes in the aggregate number of educated women and men indeed have an impact on the decision involving who marries whom, and the marital surplus.

In Table 2, we observe many changes in marital patterns between 1992 and 2014. The mean age of spouses grew by 4% for husbands and 6% for wives. The proportion of couples in the population decreased from 66% to 60%. The proportion of labor force participation of husbands and wives changed, especially for women from 54% to 67%. The proportion of married and single women in the labor market also increased from 58% to 68%. There was a marginal decrease in the male labor force participation. As to education, there was a greater increase in female labor force participation in the higher levels of education (Table 3). In a comparison of data of this table for single and married women, a pattern similar to that shown in Table 4 is observed. The number of hours spent on household work by wives who participate in the labor force is 24.7 hours per week, while the respective number for husbands was 10.1 hours in 2014.

Table 5 shows the ratio of wives' and husbands' average wages for all educational levels. In all combinations, this ratio increased from 1992 to 2014 (for all levels but one). During the preceding decades, the female labor force has grown, and the difference in the average wages earned by men and women has decreased. However, considering the ratios for the same educational levels, we observe that wives earn almost half of what their husbands do, on average. Moreover, the number of hours spent on household work by wives who participate in the labor force is more than double that spent by their husbands.

Table 6 displays the results of a regression of the logarithm of income for educational levels (conditional on potential experience) estimated by ordinary least squares (OLS). It provides evidence of the decrease in returns to education in Brazil from 1992 to 2014.<sup>21</sup> This result differs from those for the USA, Norway, Denmark, Germany, and

<sup>21</sup> The purpose of this regression is just to illustrate that the returns to education decreased from 1992 to 2014, but we do not use these coefficients in this work. We calculate the return to education according to equation 9. Additionally, this is what we hold fixed when we estimate the counterfactual.

**Table 2** Summary statistics in 1992 and 2014 (PNAD)

Sample means	1992				2014			
	Women	Men	Wives	Husbands	Women	Men	Wives	Husbands
Years of schooling	5.43	5.50	5.49	5.39	8.97	8.25	9.03	8.14
College and graduate school	8.5%	8.6%	7.8%	8.6%	21%	16%	20%	15%
Age	39.59	39.46	38.33	41.58	41.72	41.45	40.47	43.29
Yearly income (R\$)	6354	19021	5036	21419	13599	23874	12496	26776
Labor force part.	0.58	0.93	0.54	0.94	0.68	0.90	0.67	0.93
Number of children	2.25	2.22	2.46		1.35	1.30	1.45	
Women/men without children	0.14	0.15	0.10		0.24	0.28	0.22	
Proportion of couples in the population	0.66				0.60			
Observations	62754	57070	39277	39277	89139	80262	50739	50739

(1) These statistics include all 26- to 60-year-old men and women. Columns with wives and husbands consider only married (and cohabiting) people and columns “men” and “women” consider single and married (and cohabiting) people. Yearly income is in 2014 reais (ipeadata.gov.br)

**Table 3** Female labor force participation of married women by level of education (PNAD)

	1992		2014	
	FLFP	Proportion of schooling	FLFP	Proportion of schooling
Illiterate	0.53	0.16	0.53	0.05
Elementary school	0.52	0.42	0.58	0.16
Middle school	0.49	0.19	0.61	0.23
High school	0.58	0.15	0.68	0.36
College and graduate school	0.78	0.08	0.83	0.20
Total%	54%	100%	67%	100%

Notes: FLFP means female labor force participation. The ages considered here are between 26 and 60. In columns 2 and 4, we report the proportion of women who work by level of schooling, and the last row (total) shows the proportion of female labor force participation. This sample includes only married women. In this table, we use sample weights

**Table 4** Female labor force participation by level of education (PNAD): single and married women

	1992		2014	
	FLFP	Proportion of schooling	FLFP	Proportion of schooling
Illiterate	0.53	0.19	0.46	0.07
Elementary school	0.54	0.39	0.56	0.16
Middle school	0.55	0.18	0.62	0.21
High school	0.67	0.15	0.71	0.35
College and graduate school	0.84	0.08	0.82	0.21
Total%	58%	100%	68%	100%

FLFP means female labor force participation. The ages considered here are between 26 and 60. This sample includes single and married women. In this table, we use sample weights. In columns 2 and 4, we report the proportion of women who work by level of schooling, and the last row (total) shows the proportion of female labor force participation



**Table 5** Ratio of average wages of wives and husbands at all educational levels ( $\gamma$ ) (PNAD, 1992 and 2014)

Education level—wife	Education level—husband	Ratio $\gamma$ in 1992	Ratio $\gamma$ in 2014
Illiterate	Illiterate	0.15	0.53
Illiterate	Elementary school	0.17	0.41
Illiterate	Middle school	0.21	0.39
Illiterate	High school	0.12	0.37
Illiterate	Some college or above	0.16	0.36
Elementary school	Illiterate	0.19	0.48
Elementary school	Elementary school	0.15	0.40
Elementary school	Middle school	0.14	0.37
Elementary school	High school	0.12	0.27
Elementary school	Some college or above	0.08	0.26
Middle school	Illiterate	0.23	0.49
Middle school	Elementary school	0.22	0.41
Middle school	Middle school	0.18	0.38
Middle school	High school	0.14	0.33
Middle school	Some college or above	0.11	0.20
High school	Illiterate	0.52	0.53
High school	Elementary school	0.39	0.46
High school	Middle school	0.32	0.46
High school	High school	0.26	0.41
High school	Some college or above	0.14	0.27
Some college or above	Illiterate	1.82	1.09
Some college or above	Elementary school	0.84	0.86
Some college or above	Middle school	0.75	0.79
Some college or above	High school	0.57	0.71
Some college or above	Some college or above	0.35	0.52

The age of couples considered here is between 26 and 60

the UK that are reported by Eika et al. (2019). In addition, Fig. 5 displays the difference in income inequality from 1992 to 2014 in Brazil using couples' income.

## 5.2 Educational assortative mating

We analyze whether there is positive educational assortative mating in married couples. Tables 7 and 8 report all possible educational level combinations in a couple. In 1992, more educated individuals were seven times as likely to be married to one another as they were to be in a possible random educational matching, but in 2014 this ratio was approximately 3.5. One probable reason for this change was the increase in the number of educated men and women, as shown in Fig. 3.

The denominator of the marital sorting parameter  $s_{ij}$  has increased because the proportion of educated men and women has grown. As a consequence, the marital

**Table 6** Impacts of levels of education on log wages in 2014 (compared to 1992)

	(2014) Husband's log wage	(2014) Wife's log wage
Potential experience	0.17*** (0.00)	0.63*** (0.00)
Potential experience—squared	−0.0042*** (0.00)	−0.018*** (0.00)
Elementary school	−0.20*** (0.00)	−1.05*** (0.00)
Middle school	−0.11* (0.05)	−1.69*** (0.00)
High school	−0.43*** (0.00)	−3.40*** (0.00)
College and graduate school	−0.48*** (0.00)	−5.52*** (0.00)
Observations	90016	90016

*p*-values in parentheses

*Notes:* This table reports pooled OLS regression of years 1992 and 2014. Yearly income is in constant terms of 2014 (ipeadata.gov.br) Each column is a separate regression. Each coefficient measures how the return to another level of schooling has changed over the twenty two-year period. Potential experience is defined as years of work reported in the research. The dependent variable is yearly log income. The regression includes year fixed effect. In order to keep the same number of couples in each year, we drop spouses with missing variables in the data (log income and potential experience). Standard errors are in parentheses. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . The age of the spouses considered here is between 26 and 60 years old

**Table 7** Marital sorting parameter for education, 1992 (PNAD)

	Wife's education		Husband's education		
	m1	m2	m3	m4	m5
w1	<b>3.12</b>	0.88	0.28	0.08	0.02
w2	0.98	<b>1.53</b>	0.73	0.34	0.11
w3	0.31	0.79	<b>2.18</b>	1.33	0.50
w4	0.14	0.38	1.23	<b>2.98</b>	2.17
w5	0.02	0.12	0.58	1.84	<b>6.82</b>

In this table, we calculate Eq. (1). The age of spouses considered here is between 26 and 60. Numbers in bold represent marital sorting as of 1992, where both spouses had the same level of education. Education levels: w1(m1)—illiterate, w2(m2)—elementary school, w3(m3)—middle school, w4(m4)—high school, and w5(m5)—college and graduate school and w: women, and m: men

sorting parameter has become lower than it was years ago for more educated partners. The chance an educated man has of marrying an educated woman compared to a random matching has become lower, probably due to an increase in the number of educated individuals in our sample, and the chance of finding an educated person randomly has become higher.

**Table 8** Marital sorting parameter for education, 2014 (PNAD)

	Wife’s education	Husband’s education				
		m1	m2	m3	m4	m5
w1		<b>5.14</b>	1.49	0.69	0.30	0.10
w2		2.05	<b>2.39</b>	0.94	0.36	0.11
w3		0.94	1.21	<b>1.66</b>	0.76	0.19
w4		0.42	0.58	0.97	<b>1.53</b>	0.76
w5		0.16	0.26	0.41	1.01	<b>3.35</b>

In this table, we calculate Eq. (1). The age of spouses considered here is between 26 and 60. Numbers in bold represent marital sorting as of 2014, where both spouses had the same level of education. Education levels: w1(m1)—illiterate, w2(m2)—elementary school, w3(m3)—middle school, w4(m4)—high school, and w5(m5)—college and graduate school and w: women, and m: men

**Table 9** Joint distribution of education in 1992: using 2014 marginal distributions (PNAD)

Wife’s education	Husband’s education					Mg Dist (wives)
	m1	m2	m3	m4	m5	
w1: Illiterate	<b>0.03</b>	0.02	0.00	0.00	0.00	0.05
w2: Elementary school	0.03	<b>0.08</b>	0.03	0.01	0.00	0.16
w3: Middle school	0.01	0.05	<b>0.10</b>	0.06	0.01	0.23
w4: High school	0.01	0.04	0.08	<b>0.19</b>	0.05	0.36
w5: College and graduate school	0.00	0.01	0.02	0.07	<b>0.09</b>	0.20
Marginal distribution (husbands)	0.08	0.19	0.25	0.33	0.15	1

In this table, we use the marginal distribution of education as of 2014. The ages considered here are between 26 and 60. Mg Dist: marginal distribution; w: women, and m: men. Numbers in bold represent the joint distribution of education as of 2014, where both spouses had the same level of education. 5) Education levels: w1(m1)—illiterate, w2(m2)—elementary school, w3(m3)—middle school, w4(m4)—high school, and w5(m5)—college and graduate school. We run 500 iterations of a resampling bootstrap and verify that the values in the standardized table are contained in the 95% confidence interval

**Table 10** Joint distribution of education in 2014: using 2014 marginal distributions (PNAD)

Wives’ education	Husbands’ education					Mg Dist (wives)
	m1	m2	m3	m4	m5	
w1	<b>0.02</b>	0.02	0.01	0.01	0.00	0.05
w2	0.03	<b>0.07</b>	0.04	0.02	0.00	0.16
w3	0.02	0.05	<b>0.09</b>	0.06	0.01	0.23
w4	0.01	0.04	0.09	<b>0.18</b>	0.04	0.36
w5	0.00	0.01	0.02	0.06	<b>0.10</b>	0.20
Mg Dist (husbands)	0.08	0.19	0.25	0.33	0.15	1

In this table, we use the marginal distribution of education as of 2014. The ages considered here are between 26 and 60. Mg Dist: marginal distribution; w: women, and m: men. Numbers in bold represent the joint distribution of education as of 2014, where both spouses had the same level of education. 5) Education levels: w1(m1)—illiterate, w2(m2)—elementary school, w3(m3)—middle school, w4(m4)—high school, and w5(m5)—college and graduate school

**Table 11** Actual and counterfactual Gini coefficients: base year 1992

Year	Actual Gini	Random matching	Marital sorting	Educ. composition	Returns to education
1992	0.571	0.520	0.572	0.572	0.571
1999	0.573	0.522	0.572	0.582	0.562
2001	0.574	0.528	0.573	0.585	0.559
2005	0.551	0.503	0.546	0.556	0.553
2009	0.520	0.469	0.516	0.516	0.544
2014	0.499	0.449	0.493	0.487	0.534

A bootstrap exercise was performed, considering the difference between the counterfactual scenarios and the actual/random Gini coefficients. The respective results are shown in Tables 13 and 14

**Table 12** Actual and counterfactual Gini coefficients: base year 2014

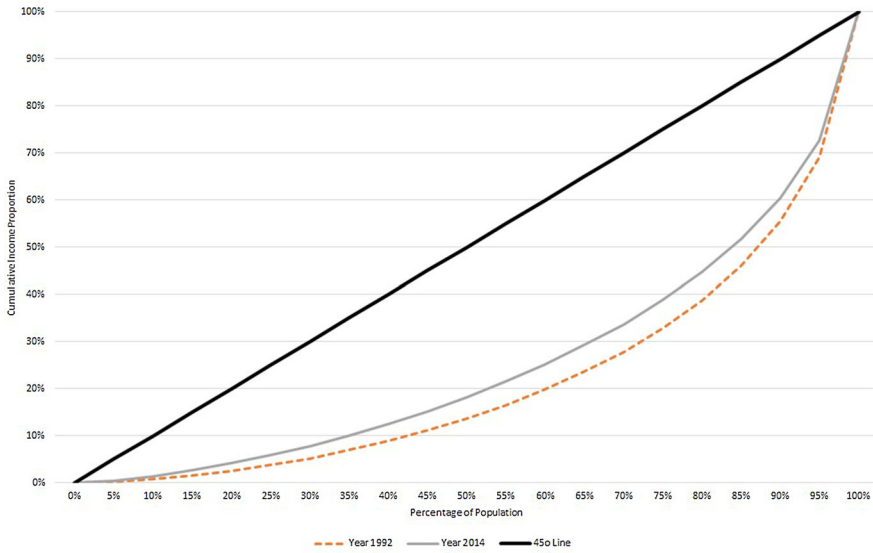
Year	Actual Gini	Random matching	Marital sorting	Educ. composition	Returns to education
1992	0.571	0.520	0.563	0.532	0.494
1999	0.573	0.522	0.567	0.546	0.493
2001	0.574	0.528	0.571	0.555	0.493
2005	0.551	0.503	0.546	0.544	0.495
2009	0.520	0.469	0.519	0.519	0.497
2014	0.499	0.449	0.500	0.500	0.499

A bootstrap exercise was performed, considering the difference between the counterfactual scenarios and the actual/random Gini coefficients. The respective results are shown in Tables 13 and 14

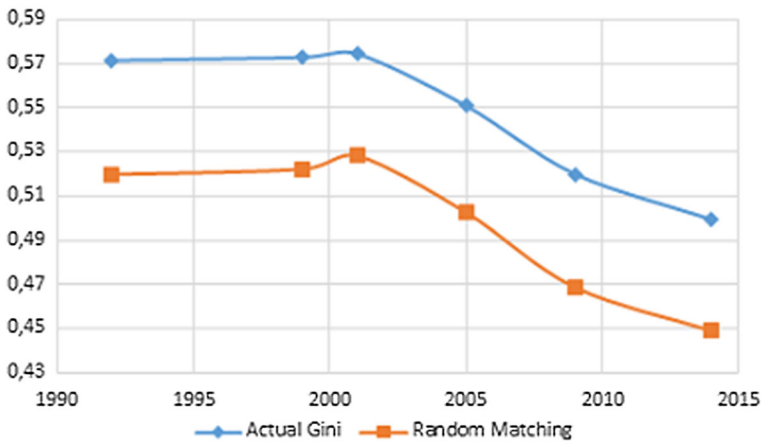
**Table 13** Bootstrap: difference between actual and counterfactual Gini coefficients (base year: 1992)

Year	Diff Actual & RM	Diff Actual & MS	Diff Actual & EC	Diff Actual & RE
1992	<b>0.052</b> (0.046;0.058)	0.000 (-0.001;0.001)	0.000 (-0.001;0.001)	0.000 (0;0)
1999	<b>0.051</b> (0.045;0.057)	0.001 (0.002;0.004)	- <b>0.009</b> (-0.012;-0.008)	<b>0.011</b> (0.003;0.017)
2001	<b>0.046</b> (0.035;0.056)	0.001 (0.003;0.006)	- <b>0.010</b> (-0.014;-0.009)	<b>0.016</b> (0.008;0.023)
2005	<b>0.048</b> (0.031;0.062)	<b>0.004</b> (0.004;0.007)	- <b>0.005</b> (-0.01;-0.004)	- 0.002 (-0.009;0.005)
2009	<b>0.051</b> (0.048;0.055)	<b>0.004</b> (0.003;0.006)	<b>0.004</b> (0;0.005)	- <b>0.024</b> (-0.031;-0.017)
2014	<b>0.050</b> (0.047;0.054)	<b>0.006</b> (0.005;0.008)	<b>0.012</b> (0.007;0.014)	- <b>0.034</b> (-0.042;-0.027)

(1) Diff Actual & RM: difference between the actual Gini coefficient and the random matching counterfactual (RM). Diff Actual & MS: difference between the actual Gini coefficient and the marital sorting counterfactual (MS). Diff Actual & EC: difference between the actual Gini coefficient and the educational composition counterfactual (EC). Diff Actual & RE: difference between the actual Gini coefficient and the return to education counterfactual (RE). (2) The 95% confidence intervals, shown in parentheses, are based on bootstrapping with 500 replications. 3) Numbers in bold are statistically significant



**Fig. 5** Lorenz curve in 1992 and 2014. Notes: These statistics include all couples formed by men and women with 26 to 36 years old. Yearly couples' income is in 2014 reais (ipeadata.gov.br)



**Fig. 6** Brazil and household income inequality: random matching

In 1992, the proportion of educated wives was 7.8% of our sample, and educated husbands represented 8.6%. By 2014, these numbers increased to 20% and 15%, respectively.<sup>22</sup> Hence, the chance of randomly finding an educated partner in 2014 became higher than it was decades ago. The denominator of the marital sorting parameter has increased faster than the numerator. As a consequence, positive assortative mating in some educational combinations is currently less prominent.

<sup>22</sup> See Fig. 3.

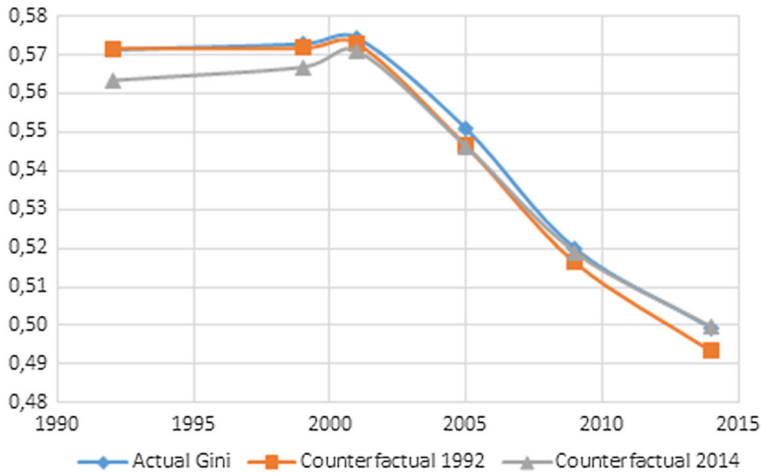


Fig. 7 Brazil and household income inequality: changes in marital sorting

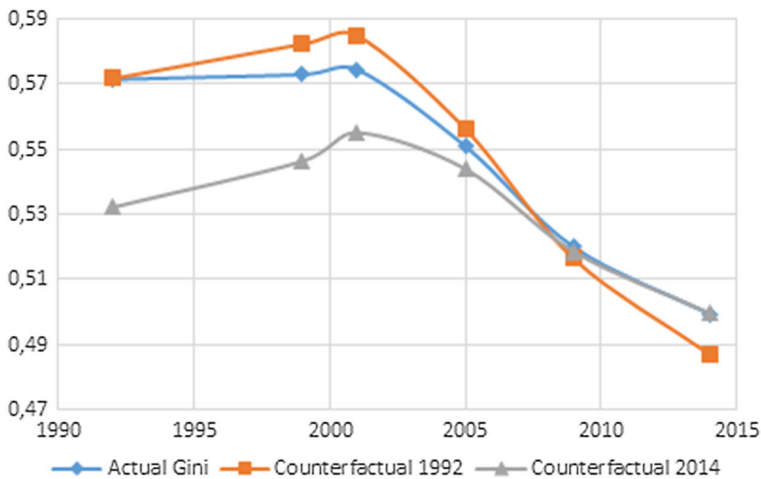


Fig. 8 Brazil and household income inequality: changes in educational composition

Another exercise we perform is to check whether the assortative mating pattern has changed over time. To this end, we fix the marginal distributions in year 2014 using the Sinkhorn–Knopp algorithm (Sinkhorn and Knopp (1967)). This algorithm scales the marginal distributions as of 1992 to year 2014, allowing the dependence structure to remain unchanged. When we calculate the odds ratios for men with educational level  $i$  and women of educational level  $j$ ; for example, using the 1992 table and the standardized table as of 1992 (using the marginal distributions of 2014), the resulting odds ratios are the same. Thus, controlling for the change in marriage composition accord-

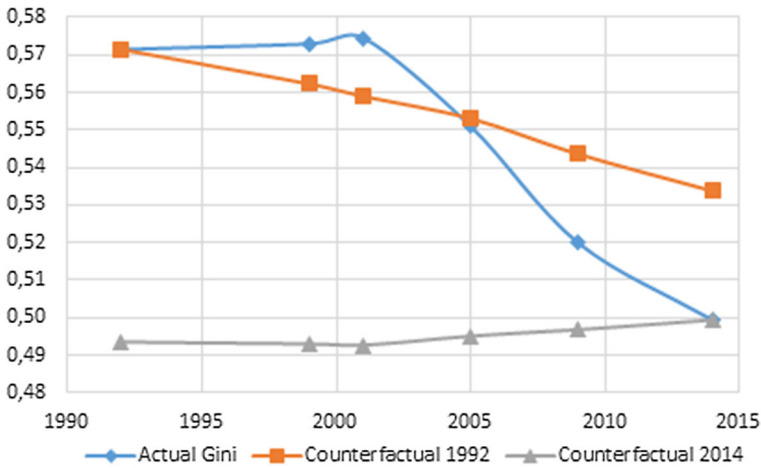


Fig. 9 Brazil and household income inequality: returns to education

Table 14 Bootstrap: difference between actual and counterfactual Gini coefficients (Base year: 2014)

Year	Diff Actual & RM	Diff Actual & MS	Diff Actual & EC	Diff Actual & RE
1992	<b>0.052</b> (0.046;0.058)	<b>0.008</b> (0.006;0.009)	<b>0.039</b> (0.035;0.041)	<b>0.078</b> (0.07;0.083)
1999	<b>0.051</b> (0.045;0.057)	<b>0.006</b> (0.005;0.008)	<b>0.027</b> (0.023;0.028)	<b>0.080</b> (0.073;0.085)
2001	<b>0.046</b> (0.035;0.056)	<b>0.003</b> (0.002;0.006)	<b>0.019</b> (0.016;0.021)	<b>0.082</b> (0.075;0.088)
2005	<b>0.048</b> (0.031;0.062)	<b>0.005</b> (0.003;0.007)	<b>0.007</b> (0.004;0.007)	<b>0.056</b> (0.049;0.061)
2009	<b>0.051</b> (0.048;0.055)	0.001 (0;0.003)	<b>0.001</b> (- 0.002;0.002)	0.023 (0.016;0.028)
2014	<b>0.050</b> (0.047;0.054)	0.000 (- 0.001;0.001)	0.000 (- 0.001;0.001)	0.000 (0;0)

See notes for Table 13

ing to education allows us to compare the results and prevent changes in schooling composition from interfering with the results of the assortative mating pattern.<sup>23</sup>

Examining the diagonals of Tables 9 and 10, we observe a decrease in assortative mating between partners with less than 12 years of schooling, an increase in that between more educated partners.<sup>24</sup> In other words, more educated couples are marrying more frequently than random educational matching would imply, which is not the case at low education levels. In addition, in almost all other combinations outside the

<sup>23</sup> See “Appendix C”.

<sup>24</sup> The difference between the values of diagonals in Tables 9 and 10 led to 30% drop in level 1, 13% drop in level 2, 9% drop in level 3 and 3% drop in level 4. In the fifth level, assortative mating increased by 6% (from 0.094 to 0.10).

**Table 15** Joint distribution of education in 1992 (PNAD)

Wife's education	Husband's education					Mg Dist (wives)
	m1	m2	m3	m4	m5	
w1	<b>0.09</b>	0.06	0.01	0.00	0.00	0.16
w2	0.07	<b>0.26</b>	0.05	0.02	0.00	0.42
w3	0.01	0.06	<b>0.07</b>	0.04	0.01	0.19
w4	0.00	0.02	0.03	<b>0.06</b>	0.03	0.15
w5	0.00	0.00	0.01	0.02	<b>0.05</b>	0.08
Marginal Distribution (husbands)	0.18	0.41	0.18	0.14	0.09	1

(1) In this table, we use the marginal distribution of education as of 1992. (2) The age of spouses considered here is between 26 and 60. (3) Mg Dist: marginal distribution; w: women, and m: men. (4) Numbers in bold represent the joint distribution of education as of 1992, where both spouses had the same level of education. (4) Education levels: w1(m1)—illiterate, w2(m2)—elementary school, w3(m3)—middle school, w4(m4)—high school, and w5(m5)—college and graduate School

**Table 16** Actual and counterfactual measures of inequality in 1992: Base year 2014

Measure	Actual	RE	Gap	Gap and RM	RM	MS	EC	FLFP	FLFP and RM
GINI	0.571	0.494	0.572	0.519	0.520	0.563	0.532	0.576	0.519
THEIL	0.651	0.500	0.657	0.519	0.522	0.631	0.532	0.670	0.521
COV	1.857	1.511	1.844	1.499	1.567	1.806	1.471	1.989	1.588
MLD	0.841	0.486	1.059	0.907	0.719	0.821	0.709	0.831	0.701

(1) RE: Return to Education counterfactual, RM: Random Matching counterfactual, Gap: Female wage gap counterfactual. Gap and RM: Female wage gap and random matching counterfactual. MS: Marital Sorting counterfactual. EC: Educational Composition counterfactual. FLFP: Female labor force participation counterfactual. Labor and RM: Female labor force participation and random matching counterfactual. (2) The age of spouses considered here is between 26 and 60. Measures of inequality: the Gini coefficient, the Theil index, the coefficient of variation and the mean log deviation

diagonal, the parameters were greater in 2014 than they were in 1992. This means that there were more mixed marriages in 2014 than in 1992. The probable reason is that the proportion of educated women was greater than that of educated men (as shown in Table 2). This change is evidence of a reduction in the educational assortative mating pattern in this period for low levels of schooling. It may also be evidence that women are marrying down.

## 6 Results

### 6.1 Household income inequality and its determinants

In this section, we estimate actual and counterfactual Gini coefficients as well as the Theil index, the mean log deviation, and the coefficient of variation. We focus the analysis on the Gini coefficient, but similar results are obtained for the other measures. In the exercises, we keep one variable (the chosen counterfactual) fixed, but



**Table 17** Actual and counterfactual measures of inequality in 2014: base year 1992

Measure	Actual	RE	Gap	Gap and RM	RM	MS	EC	FLFP	FLFP and RM
GINI	0.499	0.534	0.511	0.458	0.449	0.493	0.487	0.507	0.455
THEIL	0.495	0.534	0.517	0.407	0.390	0.480	0.487	0.507	0.400
COV	1.439	1.460	1.460	1.256	1.215	1.409	1.488	1.428	1.226
MLD	0.486	0.712	0.479	0.393	0.401	0.476	0.475	0.503	0.417

(1) RE: Return to Education counterfactual, RM: Random Matching counterfactual, Gap: Female wage gap counterfactual. Gap and RM: Female wage gap and random matching counterfactual. RM: random matching counterfactual. MS: Marital Sorting counterfactual. EC: Educational Composition counterfactual. FLFP: Female labor force participation counterfactual. Labor and RM: Female labor force participation and random matching counterfactual. (2) The age of spouses considered here is between 26 and 60. Measures of inequality: the Gini coefficient, the Theil index, the coefficient of variation and the mean log deviation

**Table 18** Summary statistics: couples aged between 26 and 36 analyzed as of 1992 and 2014 (PNAD)

Sample means of	1992		2014	
	Wives	Husbands	Wives	Husbands
Years of schooling	6.78	6.49	10.42	9.49
Age	30.06	31.75	30.56	31.86
Yearly income (R\$)	5053	19297	12374	24529
Labor force part.	0.55	0.99	0.70	0.98
Number of children	2.11		1.42	
Couples without children	0.09		0.20	
Observations	12038	12038	11925	11925

The shown statistics include all couples formed by men and women aged 26 to 36. Yearly income is in 2014 reais (ipeadata.gov.br)

the other variables vary over time. The decomposition method is used to estimate the effect of each factor on household income inequality. This can be regarded as an exercise designed to understand the effect of each factor on changes in household income inequality. We choose female labor force participation, the female wage gap, educational composition, returns to education, and, in relation to marriage and education, random matching and marital sorting parameters as explanatory variables. These variables are related to changes in schooling, labor market, and marriage with changes in social norms that mostly affect women's choices.

### 6.1.1 Household income inequality and random matching

Figure 6 shows the Gini coefficients in Brazil from 1992 to 2014. The actual Gini coefficient measures household income inequality, and the counterfactual Gini coefficient corresponds to a scenario of marriages in Brazil where husbands and wives had been randomly matched in terms of education.

**Table 19** Bootstrap: difference between measures of inequality and counterfactual one (Base year 2014)

Year: 1992	Diff Act & RE	Diff Act & Wage	Diff Act & RM	Diff Act & MS	Diff Act & EC	Diff Act & FLFP	Diff RM & Wage	Diff RM & FLFP
Gini	<b>0.078</b> (0.070;0.086)	-0.001 (-0.004;0.002)	<b>0.052</b> (0.044;0.058)	<b>0.008</b> (0.006;0.010)	<b>0.039</b> (0.034;0.042)	- <b>0.004</b> (-0.007;-0.001)	0.001 (-0.002;0.005)	0.001 (-0.005;0.007)
Theil	<b>0.151</b> (0.117;0.190)	-0.006 (-0.018;0.007)	<b>0.129</b> (0.108;0.115)	<b>0.020</b> (0.014;0.026)	<b>0.119</b> (0.101;0.138)	- <b>0.019</b> (-0.045;-0.002)	0.001 (-0.007;0.011)	0.003 (-0.005;0.016)
COV	<b>0.346</b> (0.046;0.719)	0.013 (-0.058;0.106)	<b>0.290</b> (0.201;0.394)	<b>0.052</b> (0.023;0.081)	<b>0.386</b> (0.238;0.566)	-0.131 (-0.347;0.043)	-0.020 (-0.078;0.038)	<b>0.068</b> (0.005;0.131)
MLD	<b>0.355</b> (0.327;0.381)	- <b>0.219</b> (-0.243;-0.196)	<b>0.121</b> (0.088;0.147)	<b>0.020</b> (0.007;0.028)	<b>0.132</b> (0.110;0.145)	0.009 (-0.003;0.021)	<b>0.018</b> (0.007;0.033)	- <b>0.188</b> (-0.230;-0.164)

Diff Act & RM: difference between Actual Gini and Random Matching counterfactual (RM). Diff Act & MS: difference between Actual Gini and Marital Sorting counterfactual (MS). Diff Act & EC: difference between Actual Gini and Educational Composition counterfactual (EC). Diff Act & RE: difference between Actual Gini and Return to Education counterfactual (RE). Diff Act & Wage: difference between Act Gini and Wage Gap counterfactual. Diff Act & FLFP: difference between Actual Gini and Female Labor Force Participation counterfactual. Diff RM & Wage: difference between Random Matching and Random Matching\*Wage Gap counterfactual. Diff RM & FLFP: difference between Random Matching and Random Matching\*FLBP counterfactual. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

The difference between the actual Gini coefficient (0.571) and the counterfactual one in 1992 (0.519) was 9.1%, while in 2014 the difference rose to 10%.<sup>25</sup> The counterfactuals are shown in Tables 11 and 12. These differences are statistically significant, as shown in Tables 13 and 14.<sup>26</sup> This is evidence that educational assortative mating can have an impact on household income inequality. In addition, educational composition within couples has changed in Brazil, and this probably explains why the difference between the actual and counterfactual Gini coefficients in 2014 increased relative to that in 1992. As Fig. 4 shows, the proportion of married couples in which both spouses had the same level of education decreased to 47%, and the proportion of couples in which husbands had a higher level of education that did their wives decreased to 19%, while the proportion of couples in which women had more years of schooling than did men increased to 34%, which indicates that marriages mixed with respect to educational level became more widespread.<sup>27</sup>

The difference between the counterfactual Gini coefficient with random matching and the actual Gini coefficient is negative, which is similar to the results for the USA, Norway, Denmark, Germany, and the UK obtained by Eika et al. (2019); however, there is a downward trend in household income inequality in Brazil, whereas upward trends have been observed in the listed developed countries.

### 6.1.2 Household income inequality and female labor force participation

In this counterfactual scenario, we are interested in analyzing the impact of female labor force participation on the Gini coefficient.<sup>28</sup> We calculate the actual Gini coefficient in year  $t$  and three counterfactual scenarios. The first counterfactual uses a reweighting function that considers random matching in year  $t$  and the second is a female labor force participation counterfactual in base year  $b$ . For the last scenario, we use the reweighting function considering both random matching in year  $t$  and female labor force participation in base year  $b$  (columns RM, FLFP and, FLFP and RM, respectively, in Table 17).

The married female labor force participation in Brazil increased from 1992 to 2014, following the same trend as in the USA, Denmark, Germany, the UK, and Norway, as shown in Greenwood et al. (2014) and Eika et al. (2019). If female labor force participation in 1992 had been at the level of 2014, the Gini coefficient would have been greater at 0.576 instead of 0.571. If we set female labor force participation in 2014 to the level of 1992, the Gini coefficient would be greater than the actual one, at 0.507 instead of 0.499, a difference of 1.6%. As Table 17 shows, this result may be evidence that the rise in female labor force participation in Brazil in the preceding decades was one of the reasons the Gini coefficient declined.

To test whether the differences between the actual Gini coefficient and those of the counterfactual scenarios are significantly different from zero, we calculate

<sup>25</sup> In 2014, the actual Gini coefficient was 0.499 and the counterfactual was 0.449.

<sup>26</sup> We run 500 iterations of a resampling bootstrap to determine whether the differences between the actual Gini coefficient and the counterfactual one are significantly different from zero.

<sup>27</sup> Tables 13 and 14 show the differences between the actual Gini coefficient and the counterfactual scenarios and their confidence intervals.

<sup>28</sup> The methodology is presented in Sect. 4.2.1.

**Table 20** Bootstrap: difference between measures of inequality and counterfactual one (Base year 1992)

Year: 2014	Diff Act & RE	Diff Act & Wage	Diff Act & RM	Diff Act & MS	Diff Act & EC	Diff Act & FLFP	Diff RM & Wage	Diff RM & FLFP
Gini	-0.034 (-0.043; -0.026)	-0.012 (-0.014; -0.009)	0.050 (0.046; 0.055)	0.006 (0.004; 0.007)	0.012 (0.009; 0.016)	-0.008 (-0.011; -0.005)	-0.006 (-0.009; -0.003)	-0.009 (-0.014; -0.006)
Theil	-0.040 (-0.075; -0.007)	-0.023 (-0.033; -0.011)	0.105 (0.091; 0.121)	0.014 (0.009; 0.016)	0.008 (-0.001; 0.017)	-0.013 (-0.023; -0.003)	-0.010 (-0.017; -0.004)	-0.017 (-0.028; -0.009)
COV	-0.021 (-0.281; 0.173)	-0.021 (-0.076; 0.046)	0.224 (0.163; 0.292)	0.030 (0.017; 0.037)	-0.049 (-0.070; -0.026)	0.012 (-0.037; 0.068)	-0.011 (-0.028; 0.006)	-0.041 (-0.073; -0.010)
MLD	-0.225 (-0.252; -0.001)	0.007 (0.001; 0.014)	0.086 (0.076; 0.095)	0.010 (0.006; 0.012)	0.012 (0.003; 0.020)	-0.016 (-0.023; -0.010)	-0.016 (-0.022; -0.011)	0.008 (0.000; 0.014)

Diff Act & RM: difference between Actual Gini and Random Matching counterfactual (RM). Diff Act & MS: difference between Actual Gini and Marital Sorting counterfactual (MS). Diff Act & EC: difference between Actual Gini and Educational Composition counterfactual (EC). Diff Act & RE: difference between Actual Gini and Return to Education counterfactual (RE). Diff Act & Wage: difference between Act Gini and Wage Gap counterfactual. Diff Act & FLFP: difference between Actual Gini and Female Labor Force Participation counterfactual. Diff RM & Wage: difference between Random Matching and Random Matching\*Wage Gap counterfactual. Diff RM & FLFP: difference between Random Matching and Random Matching\*FLBP counterfactual. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

**Table 21** 1992 Sequential decomposition using counterfactuals (Base year: 2014)

Primary Order	Actual	MS	(1) + RE	(2) + EC	(3) + Gap	(4) + FLFP
Gini	<b>0.571</b> (0.565;0.578)	<b>0.563</b> (0.558;0.570)	<b>0.528</b> (0.519;0.538)	<b>0.555</b> (0.548;0.562)	<b>0.554</b> (0.547;0.561)	<b>0.551</b> (0.543;0.561)
Theil	<b>0.651</b> (0.621;0.688)	<b>0.631</b> (0.603;0.663)	<b>0.572</b> (0.528;0.641)	<b>0.595</b> (0.569;0.623)	<b>0.592</b> (0.569;0.616)	<b>0.583</b> (0.558;0.609)
COV	<b>1.857</b> (1.561;2.225)	<b>1.806</b> (1.529;2.151)	<b>1.956</b> (1.454;2.571)	<b>1.615</b> (1.438;1.848)	<b>1.558</b> (1.424;1.728)	<b>1.537</b> (1.404;1.694)
MLD	<b>0.841</b> (0.817;0.863)	<b>0.821</b> (0.799;0.843)	<b>0.777</b> (0.745;0.810)	<b>0.814</b> (0.784;0.846)	<b>1.046</b> (0.990;1.101)	<b>0.999</b> (0.943;1.053)
Reverse Order	Actual	FLFP	(1) + GAP	(2) + EC	(3) + RE	(4) + MS
Gini	<b>0.571</b> (0.565;0.578)	<b>0.575</b> (0.569;0.583)	<b>0.568</b> (0.561;0.574)	<b>0.525</b> (0.518;0.531)	<b>0.562</b> (0.555;0.569)	<b>0.551</b> (0.543;0.561)
Theil	<b>0.651</b> (0.621;0.688)	<b>0.670</b> (0.632;0.722)	<b>0.645</b> (0.613;0.682)	<b>0.513</b> (0.491;0.537)	<b>0.616</b> (0.591;0.643)	<b>0.583</b> (0.558;0.609)
COV	<b>1.857</b> (1.561;2.225)	<b>1.988</b> (0.558;0.570)	<b>1.842</b> (0.519;0.538)	<b>1.411</b> (0.548;0.562)	<b>1.667</b> (0.547;0.561)	<b>1.537</b> (1.404;1.694)
MLD	<b>0.841</b> (0.817;0.863)	<b>0.831</b> (0.811;0.853)	<b>1.009</b> (0.977;1.044)	<b>0.787</b> (0.756;0.819)	<b>1.036</b> (0.989;1.085)	<b>0.999</b> (0.943;1.053)

FLFP is female labor force participation. MS is marital sorting, RE is returns to education, EC is educational composition and GAP is female wage gap. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

bootstrapped standard errors; the results are presented in Tables 19 and 20. These differences are statistically different from zero.

Moreover, we perform the same procedures for the Gini coefficient with random matching and the other counterfactual scenarios. The results are shown in the last columns of Tables 16 and 17. This exercise is important because by using a random matching counterfactual together with the female labor force counterfactual,<sup>29</sup> we eliminate the effect of (educational) marital sorting in the analysis, and the difference estimated in the last columns of the tables is due to this effect. In 2014, the counterfactual was 0.455, whereas the random matching-based Gini coefficient was 0.449. This result shows that without the increase in female participation in the labor market that occurred in the country during the analyzed period, the Gini index would be greater than it really is. The results are also statistically significant in 2014 but not for 1992.<sup>30</sup>

The effects we observe for Brazil with this counterfactual for 2014 are similar to what has been observed in the USA. According to Greenwood et al. (2014), when comparing the random matching counterfactual in 2005 with the female labor force

<sup>29</sup> The respective results are labeled with FLFP\*RM (female labor force participation\*random matching, where the former refers to FLFP from base year 1992 (or 2014)).

<sup>30</sup> We run 500 iterations of a resampling bootstrap to determine whether the differences between the actual Gini coefficient and the counterfactual one are significantly different from zero.

**Table 22** 1992 Sequential decomposition using counterfactuals with random matching (Base year: 2014)

Primary Order	Actual	RM	(1) + RE	(2) + EC	(3) + Gap	(4) + FLFP
Gini	<b>0.571</b> (0.565;0.578)	<b>0.520</b> (0.512;0.527)	<b>0.517</b> (0.506;0.528)	<b>0.518</b> (0.505;0.533)	<b>0.519</b> (0.501;0.544)	<b>0.521</b> (0.569;0.583)
Theil	<b>0.651</b> (0.621;0.688)	<b>0.522</b> (0.496;0.550)	<b>0.533</b> (0.490;0.594)	<b>0.507</b> (0.475;0.535)	<b>0.510</b> (0.477;0.543)	<b>0.512</b> (0.468;0.562)
COV	<b>1.857</b> (1.561;2.225)	<b>1.567</b> (1.333;1.833)	<b>1.744</b> (1.341;2.262)	<b>1.433</b> (1.275;1.635)	<b>1.396</b> (1.266;1.538)	<b>1.389</b> (1.250;1.506)
MLD	<b>0.841</b> (0.817;0.863)	<b>0.719</b> (0.688;0.755)	<b>0.745</b> (0.713;0.780)	<b>0.711</b> (0.674;0.743)	<b>0.910</b> (0.847;0.939)	<b>0.870</b> (0.802;0.939)
Reverse Order	Actual	FLFP	(1) + GAP	(2) + EC	(3) + RE	(4) + RM
Gini	<b>0.571</b> (0.565;0.578)	<b>0.575</b> (0.512;0.527)	<b>0.568</b> (0.506;0.528)	<b>0.525</b> (0.505;0.533)	<b>0.562</b> (0.501;0.544)	<b>0.521</b> (0.569;0.583)
Theil	<b>0.651</b> (0.621;0.688)	<b>0.670</b> (0.632;0.722)	<b>0.645</b> (0.613;0.682)	<b>0.513</b> (0.491;0.537)	<b>0.616</b> (0.591;0.643)	<b>0.512</b> (0.468;0.562)
COV	<b>1.857</b> (1.561;2.225)	<b>1.988</b> (1.600;2.502)	<b>1.842</b> (1.557;2.234)	<b>1.411</b> (1.263;1.629)	<b>1.667</b> (1.481;1.902)	<b>1.389</b> (1.250;1.506)
MLD	<b>0.841</b> (0.817;0.863)	<b>0.831</b> (0.811;0.853)	<b>1.009</b> (0.977;1.044)	<b>0.787</b> (0.756;0.819)	<b>1.036</b> (0.989;1.085)	<b>0.870</b> (0.802;0.939)

FLFP is female labor force participation. RM is random matching, RE is returns to education, EC is educational composition and GAP is female wage gap. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

participation (considering random matching) counterfactual in 1960, the respective Gini coefficients are 0.34 and 0.44. It means that the increase in the married female labor force smothered the rising trend in the US Gini coefficient.

### 6.1.3 Household income inequality and female wage gap

The wage gap between husbands and wives has shrunk in the preceding decades and has had an effect on household income inequality. As shown in Table 17, if in 2014 the wage gap between husbands and wives had been at the 1992 level, the Gini coefficient would have been greater, at 0.511 instead of 0.499. Thus, the decrease in female wage gap is one of the reasons for the decreased Gini coefficient. The results are statistically significant for 2014 but not for 1992.

The random matching counterfactual for 2014 would be 0.449. Considering a wage gap between husbands and wives at the level of 1992 and random matching in 2014 (wage gap\*RM), the Gini coefficient would be 0.458.

These results suggest that the decrease in the wage gap in the preceding decades has contributed to a reduction of the Gini coefficient, showing the importance of the gender wage gap for income inequality.

**Table 23** 2014 sequential decomposition using counterfactuals (Base year: 1992)

Primary Order	Actual	MS	(1) + RE	(2) + EC	(3) + Gap	(4) + FLFP
Gini	<b>0.499</b> (0.494;0.505)	<b>0.493</b> (0.487;0.497)	<b>0.470</b> (0.461;0.478)	<b>0.484</b> (0.475;0.492)	<b>0.494</b> (0.484;0.504)	<b>0.494</b> (0.484;0.503)
Theil	<b>0.495</b> (0.474;0.518)	<b>0.480</b> (0.460;0.500)	<b>0.439</b> (0.410;0.470)	<b>0.470</b> (0.439;0.505)	<b>0.492</b> (0.460;0.532)	<b>0.491</b> (0.460;0.529)
COV	<b>1.439</b> (1.337;1.575)	<b>1.409</b> (1.307;1.534)	<b>1.358</b> (1.212;1.523)	<b>1.445</b> (1.280;1.641)	<b>1.499</b> (1.326;1.696)	<b>1.489</b> (1.327;1.676)
MLD	<b>0.486</b> (0.475;0.498)	<b>0.476</b> (0.464;0.485)	<b>0.428</b> (0.410;0.443)	<b>0.466</b> (0.446;0.483)	<b>0.457</b> (0.438;0.475)	<b>0.457</b> (0.437;0.476)
Reverse Order	Actual	FLFP	(1) + GAP	(2) + EC	(3) + RE	(4) + MS
Gini	<b>0.499</b> (0.494;0.505)	<b>0.507</b> (0.503;0.512)	<b>0.511</b> (0.506;0.516)	<b>0.487</b> (0.481;0.494)	<b>0.502</b> (0.494;0.511)	<b>0.494</b> (0.484;0.503)
Theil	<b>0.495</b> (0.474;0.518)	<b>0.507</b> (0.492;0.524)	<b>0.519</b> (0.501;0.536)	<b>0.496</b> (0.478;0.518)	<b>0.508</b> (0.481;0.540)	<b>0.491</b> (0.460;0.529)
COV	<b>1.439</b> (1.337;1.575)	<b>1.428</b> (1.362;1.513)	<b>1.460</b> (1.384;1.541)	<b>1.524</b> (1.439;1.629)	<b>1.503</b> (1.370;1.659)	<b>1.489</b> (1.327;1.676)
MLD	<b>0.486</b> (0.475;0.498)	<b>0.503</b> (0.493;0.513)	<b>0.481</b> (0.471;0.491)	<b>0.440</b> (0.428;0.454)	<b>0.472</b> (0.456;0.489)	<b>0.457</b> (0.437;0.476)

FLFP is female labor force participation. MS is marital sorting, RE is returns to education, EC is educational composition and GAP is female wage gap. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

We calculate the differences between the actual Gini coefficient and counterfactual scenarios to test whether these differences are significantly different from zero, as shown in Tables 19 and 20. The results are statistically significant for 2014.

Analyzing column 4 in Table 17, we observe that if the wage gap in 2014 had been as it was in 1992, the Gini coefficient would have been greater (0.458 instead of 0.449). This result highlights the role played by the wage gap counterfactual together with random matching. We eliminate, as in subsection 6.1.2, the effect of (educational) marital sorting in the analysis.<sup>31</sup> This result may indicate that a lower wage gap between men and women has a significant effect on the Gini coefficient, which means that public policies that encourage a reduction in the wage gap may promote a reduction in income inequality. The results are also statistically significant.

Our results are consistent with those of Firpo and Reis (2007), where the researchers observed an effect on income inequality of a reduction in wage gap associated with race, gender, location and formal work status.

<sup>31</sup> The respective results are labeled with WG\*RM (wage gap\*random matching, where the former refers to the wage gap from base year 1992).

**Table 24** 2014 sequential decomposition using counterfactuals with random matching (Base year: 1992)

Primary Order	Actual	RM	(1) + RE	(2) + EC	(3) + Gap	(4) + FLFP
Gini	<b>0.499</b> (0.494;0.505)	<b>0.449</b> (0.443;0.454)	<b>0.438</b> (0.430;0.447)	<b>0.438</b> (0.425;0.451)	<b>0.447</b> (0.431;0.468)	<b>0.452</b> (0.430;0.486)
Theil	<b>0.495</b> (0.474;0.518)	<b>0.390</b> (0.375;0.406)	<b>0.372</b> (0.351;0.398)	<b>0.371</b> (0.345;0.396)	<b>0.389</b> (0.351;0.428)	<b>0.399</b> (0.354;0.459)
COV	<b>1.439</b> (1.337;1.575)	<b>1.215</b> (1.142;1.299)	<b>1.207</b> (1.094;1.344)	<b>1.192</b> (1.085;1.312)	<b>1.240</b> (1.121;1.366)	<b>1.250</b> (1.125;1.374)
MLD	<b>0.486</b> (0.475;0.498)	<b>0.401</b> (0.389;0.411)	<b>0.383</b> (0.366;0.399)	<b>0.401</b> (0.365;0.442)	<b>0.387</b> (0.354;0.425)	<b>0.396</b> (0.355;0.450)
Reverse Order	Actual	FLFP	(1) + GAP	(2) + EC	(3) + RE	(4) + RM
Gini	<b>0.499</b> (0.494;0.505)	<b>0.507</b> (0.503;0.512)	<b>0.511</b> (0.506;0.516)	<b>0.487</b> (0.481;0.494)	<b>0.502</b> (0.494;0.511)	<b>0.452</b> (0.430;0.486)
Theil	<b>0.495</b> (0.474;0.518)	<b>0.507</b> (0.375;0.406)	<b>0.519</b> (0.351;0.398)	<b>0.496</b> (0.345;0.396)	<b>0.508</b> (0.351;0.428)	<b>0.399</b> (0.354;0.459)
COV	<b>1.439</b> (1.337;1.575)	<b>1.428</b> (1.362;1.513)	<b>1.460</b> (1.384;1.541)	<b>1.524</b> (1.439;1.629)	<b>1.503</b> (1.370;1.659)	<b>1.250</b> (1.125;1.374)
MLD	<b>0.486</b> (0.475;0.498)	<b>0.503</b> (0.493;0.513)	<b>0.481</b> (0.471;0.491)	<b>0.440</b> (0.428;0.454)	<b>0.472</b> (0.456;0.489)	<b>0.396</b> (0.355;0.450)

FLFP is female labor force participation. RM is random matching, RE is returns to education, EC is educational composition and GAP is female wage gap. Measures of Inequality: Gini, Theil index, coefficient of variation and mean log deviation. A 95% Confidence Interval is shown in parentheses and they are based on bootstrapping with 500 replications

**Table 25** Actual and counterfactual measures of inequality in 1992 determined using data for couples aged between 26 and 36: Base year 2014

Measure	Actual	RE	Gap	Gap and RM	RM	MS	EC	FLFP	FLFP and RM
Gini	0.540	0.470	0.541	0.498	0.497	0.551	0.520	0.541	0.484
THEIL	0.549	0.445	0.553	0.477	0.474	0.588	0.490	0.554	0.436
COV	1.395	1.391	1.392	1.296	1.313	1.498	1.249	1.416	1.211
MLD	0.768	0.421	1.037	0.962	0.681	0.802	0.700	0.762	0.617

FLFP is female labor force participation. MS is marital sorting. RE is return to education. RM is random matching, EC is educational composition and GAP is female wage gap. Measures of inequality: the Gini coefficient, the Theil index, the coefficient of variation and the mean log deviation

### 6.1.4 Household income inequality and marital sorting

Figure 7 displays the actual and the counterfactual Gini coefficients with marital sorting measured in 1992 and 2014. That is, we use the probability of marriage distribution as it was in 1992 and 2014, allowing couples' income and husbands' and wives' educational levels to vary over the years, to calculate the counterfactual Gini coefficient.



**Table 26** Actual and counterfactual measures of inequality in 2014 determined using data for couples aged between 26 and 36: Base year 1992

Measure	Actual	RE	Gap	Gap and RM	RM	MS	EC	FLFP	FLFP and RM
Gini	0.472	0.495	0.481	0.430	0.429	0.475	0.467	0.479	0.426
THEIL	0.435	0.439	0.462	0.357	0.350	0.440	0.441	0.450	0.348
COV	1.321	1.153	1.433	1.207	1.139	1.335	1.384	1.352	1.150
MLD	0.426	0.626	0.417	0.340	0.360	0.432	0.416	0.442	0.357

FLFP is female labor force participation. MS is marital sorting. RE is return to education. RM is random matching, EC is educational composition and GAP is female wage gap. Measures of inequality: the Gini coefficient, the Theil index, the coefficient of variation and the mean log deviation

The difference between the 2014 counterfactual (0.563) determined using the marital sorting as it was in 1992 and the actual Gini coefficient (0.571) in 1992 was 1.4%; however, this small difference decreased in the following decade, and there was hardly any difference at the end of the period. In 2014, the 1992 counterfactual (0.493) determined using the marital sorting as it was in 2014 was 1.2% smaller than the actual Gini coefficient (0.499).

The 2014 counterfactual is lower than the actual Gini coefficient in the 1990s. This result is probably due to the change in the pattern of assortative mating between 1992 and 2014. In 2014, there was a reduction in positive assortative mating among individuals with lower levels of schooling, while there was an increase among the more educated. Hence, if in 1992 the levels of assortative mating had been as in 2014, there would have been more marriages between more educated individuals and more mixed marriages involving individuals with lower levels of schooling; thus, couples' income inequality would have been lower.<sup>32</sup>

Using the assortative mating pattern of 1992 in 2014 gives us a different result. The marital pattern in 1992 was less assortative among the more educated. In this case, this pattern leads the 1992 counterfactual (in 2014) to be smaller than the actual Gini coefficient as of 2014.

There is evidence of a small change in the assortative pattern in Brazil in the past few decades. The pattern of assortative mating has a small but statistically significant impact on household income inequality, as shown in Tables 13 and 14. In the USA, as presented in Eika et al. (2019), and in urban China, examined by Feng and Tang (2019), there has been no change in this pattern.

### 6.1.5 Household income inequality and educational composition

Figure 8 shows the actual and counterfactual Gini coefficients of educational composition measured in 1992 and 2014. That is, we use the distributions of educational composition prevailing in 1992 and 2014 while allowing the couples' income and marital sorting to vary in these years to calculate the counterfactual Gini coefficient.

In 1992, the actual Gini coefficient (0.571) and the counterfactual one (0.532), determined while considering 2014 as the base year, differed by 7.3%. Before 2005,

<sup>32</sup> See Tables 9 and 10.

the 2014 counterfactual Gini coefficient was below the actual one, suggesting that if the educational distributions of husbands and wives had been those prevailing in 2014, household income inequality would have been lower than it actually was. The educational distribution as of 2014 is made up of individuals with more average years of schooling than that in 1992, and considering that the other factors vary, the effect is that the counterfactual of 2014 would be smaller than the actual Gini coefficient of the 1990s. The trend we observe is similar to the results in Menezes-Filho et al. (2007) for men aged between 24 and 56 in Brazil.

After 2005, this pattern changed. The 1992 Gini coefficient is lower than the actual one. This difference is probably due to the homogeneous educational composition of the Brazilian society in 1992.<sup>33</sup> The difference between the actual Gini coefficient (0.499) and the 1992 counterfactual (0.487) is approximately 2.5%. These differences are statistically significant, as shown in Tables 13 and 14.

In this exercise, we hold educational composition fixed, and allow the returns to education, marital sorting parameters, and the other factors to vary over time. This difference between the actual educational composition and the counterfactual affects household income inequality, as there is more (or less) heterogeneity among couples. In Brazil, in the most recent years, the proportion of educated couples rose significantly, while the number of uneducated couples decreased.

The results for Brazil are different from those for the USA, as analyzed by Eika et al. (2019), probably because educational composition in Brazil evolved differently from that in the USA. In the USA, the counterfactual in 1962 shows that household income inequality would have been greater than the actual Gini coefficient, and the counterfactual in 2013 shows that household income inequality would have been lower than the actual Gini coefficient. In Brazil, if the educational composition had not changed since 1992, the actual Gini coefficient would be lower. In this case, a lower Gini coefficient does not mean an improvement in education because the educational composition in Brazil in 1992 involved almost 70% uneducated men and 71% uneducated women. By 2014, this proportion fell to 27% for men and 21% for women (Fig. 2).

### 6.1.6 Household income inequality and returns to education

Figure 9 shows the actual Gini coefficient and the counterfactual for returns to education measured in 1992 and 2014. That is, we keep the distribution of returns to education as it was in 1992 and 2014, allowing couples' educational composition, marital sorting of husbands and wives, and the other factors to vary in these years, to calculate the counterfactual Gini coefficient.

As Fig. 9 shows, there is a difference between the actual Gini coefficient and the 1992 and 2014 counterfactual scenarios. The difference between the actual Gini coefficient (0.571) and the 2014 counterfactual (0.494) was approximately 16% in 1992. Part of this difference can be explained by the decline in returns to education: As Table 6 shows, one additional level of schooling in 2014 yielded less than it did in 1992.<sup>34</sup> In addition, as Figs. 2 and 3 show, educational composition changed between 1992

<sup>33</sup> In our sample, almost 70% of men and 71% of women had less than 5 years of schooling in 1992.

<sup>34</sup> The purpose of this observation is to illustrate this trend rather than to construct the counterfactual.

and 2014. The difference between the 1992 counterfactual (0.534) and the actual Gini coefficient (0.499) was 6.6% in 2014. Despite the change in educational composition each year, the returns to education were different in 2014 from those in 1992, so a significant impact on household income inequality is observed when considering returns to education fixed at the 2014 level. These differences are statistically significant, as shown in Tables 13 and 14.

The returns to education in the USA have increased, as shown in Eika et al. (2019), unlike in Brazil. In the USA, the counterfactual in 2013 shows that household income inequality would have been greater than the actual Gini coefficient, and the counterfactual in 1962 shows that household income inequality would have been lower than the actual Gini coefficient. It seems that returns to education play an important role in household income inequality.

### 6.1.7 Household income inequality and sequential decomposition

Applying the sequential decomposition method of DiNardo et al. (1996), as shown in Table 21, we observe a difference in the actual Gini coefficient in 1992 from 0.571 to 0.551 (almost 4%) related to all five counterfactual components studied in this paper. In this case, we estimate the density with all factors together, which allows us to observe the effects of the changes that have occurred since 1992 and that have had a positive impact in reducing income inequality.

The components' effects differ depending on their order in the sequence; however, the final outcome is the same (it is invariant to the order of factors). The advantage of applying sequential decomposition is that we can observe the impact of each of the factors in the reduction (or increase) of the Gini index and, in addition we observe the total effect of the decomposition on the Gini index considering all the effects together.

Therefore, we can assume that all of these five components are important to the decrease in household income inequality in Brazil. The educational composition and returns to education appear to be the only factors acting in the opposite direction. These counterfactual components alone have important effects on income distribution, but are probably correlated, which may lead one to partially offset the effect of the other.

Nevertheless, examining Table 23, we observe that the labor components as in 1992 increase the Gini coefficient, whereas educational composition as in 1992 has the opposite effect. The overall result in 2014 is a decline from 0.499 to 0.494.

In Tables 22 and 24, random matching plays a major role in the decrease in the Gini coefficient.

### 6.2 Robustness check

As a robustness check, we consider other measures of inequality, such as the Theil index, the mean log deviation, and the coefficient of variation. These inequality measures change in the same direction as does the Gini coefficient for both separate components and the sequential decomposition. The respective results are shown in Tables 16, 17, and 21 to 24. The exception is the mean log deviation and sequen-

tial decomposition (see Table 21); in this case, they are greater after combining all counterfactual.

Moreover, we perform the analysis for a subsample formed by couples aged between 26 and 36. The respective descriptive statistics are shown in Table 18. The respective results for measures of inequality are shown in Tables 25 and 26. The Gini coefficient for this subsample is smaller than that for the sample formed by couples aged between 26 and 60. Nevertheless, the counterfactual results follow the same trend as does the full sample.

## 7 Conclusions

In this paper, we apply the decomposition method of DiNardo et al. (1996) to estimate the effects of changes in the wage gap and female labor force participation on household income inequality in Brazil from 1992 to 2014 using the PNAD survey.

We find that the decrease in the wage gap and the rise in female labor force participation in recent years have contributed to a decline in the Gini coefficient. These results indicate the importance of changes in the female role in intrahousehold labor allocation in reducing household income inequality.

We calculate couples' educational composition and investigate how changes in this variable impact educational assortative mating during this period. In addition, we investigate the impacts on Brazilian household income inequality due to changes in educational assortative mating patterns. We implement a decomposition method based on Eika et al. (2019), in which we can assess the impacts of returns to education, educational composition, and marital sorting parameters on household income distribution. We also assess the statistical significance of these impacts.

The proportion of more educated couples increased in Brazil from 1992 to 2014, while the proportion of less educated couples declined. In addition, there was an increase in the number of mixed marriages at the lower levels of schooling. As the proportion of educated women increased faster than that of educated men, women are marrying down. The effect of these changes can be observed not only in marital sorting parameters but also in household income inequality. Our results suggest that if marriages were formed randomly, household income inequality would be approximately 0.05 points lower (0.44) than the actual Gini coefficient (0.49) in 2014. In addition, there is evidence that more positive assortative mating occurred in the Brazilian society between husbands and wives with more than 12 years of schooling, whereas it decreased among couples where spouses have less than 12 years of schooling.

These two opposite effects probably affect household income inequality in different directions, preventing it from declining more than it has in recent years. Moreover, the counterfactual Gini coefficient (0.49) in which the marital sorting parameters are fixed as of a base year is slightly different, by 0.01 point, from the actual Gini coefficient of 0.50. This is evidence that the pattern of assortative mating has changed and that it has a small but statistically significant effect on household income inequality, unlike the result in the USA.<sup>35</sup>

<sup>35</sup> In the US case, it appears that marital sorting has no effect on income inequality.

We also observe that the increase in the proportion of educated husbands and wives and the decrease in the proportion of uneducated husbands and wives led to a change in the educational composition of couples between 1992 and 2014. In our exercise, we estimate our counterfactual in 2014 (with the educational composition of 1992) and compare it to the actual Gini coefficient (0.499). The former (0.487) is lower by 0.013 points. This may be evidence of how homogeneous educational composition was in 1992 because most couples were formed by spouses with less than 5 years of schooling. In 2014, there were more educated couples than in 1992, and the proportion of uneducated couples declined sharply, leading to a more heterogeneous educational composition.

In Brazil, the returns to education have decreased in recent years, probably due to the increase in the proportion of educated men and women. In the Brazilian case, if we estimate the Gini coefficient in 1992 with the returns to education fixed as of year 2014, we obtain a lower value (0.494) than the actual one (0.571). The difference was almost 0.08 points in 1992. In other mentioned countries, the results evolve in the opposite direction, as described in Eika et al. (2019) and Feng and Tang (2019).

This result confirms that the decrease in the return to education in Brazil played an important role in reducing inequality as highlighted in the Brazilian literature on the subject. However, this work contributes to highlight that other factors also contributed to the decrease in income inequality.

## Declarations

**Conflicts of interest** This study is based on a chapter of Lorena Hakak's doctoral dissertation at Sao Paulo School of Economics. The author received financial support from the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq-Brazil) to pursue her doctoral studies. The authors declare that they have no conflicts of interest.

## A Appendix

### A.1 The steps to obtain the reweighting function $P'_x(t_{ej} = b, t_s = t)$ :

1) We draw one man from the husbands' marginal distribution of education and one woman from the wives' marginal distribution of education in  $t = 1992$ . With probability

$$\Pr(Wife = i, t = 1992) \cdot \Pr(Husband = j, t = 1992), \quad (19)$$

we obtain a man and a woman with educational levels  $i, j$ .

2) Considering the product of marginal distributions and the marital sorting parameter  $s_{ij}(t)$ , we decide whether the couple gets married. Hence, to construct the counterfactual we need to draw a man and woman from the marginal educational distributions of men and women and estimate  $s_{ij}(t)$ . The probability of them getting married is

$$P' = \Pr(Wife = i, t = b) \cdot \Pr(Husband = j, t = b) \cdot s_{ij}(t).$$

3) If they get married, we draw<sup>36</sup> them from the marginal distributions of education and measure the probabilities in equation (19) again without that couple. We need to calculate the marginal distributions in every iteration. Then, we repeat the process until all couples have been formed.

## B Appendix

### B.1 standardized contingency table

To analyze whether positive assortative mating in Brazil increased from 1992 to 2014, we need to compare the joint distributions from these years. To this end, we standardize both joint distribution tables, considering the same marginal distributions. We use the Sinkhorn–Knopp algorithm that allows us to iterate over columns and rows, preserving the dependent relationship between the joint distribution and the marginal distributions.<sup>37</sup>

### B.2 Sinkhorn–Knopp Algorithm

We perform the following steps to execute the algorithm:

- 1- Divide the husbands' marginal educational distribution as of 1992 (or 2014) by the marginal distribution as of the year used to standardize the table. We obtain a weight for each education level.
- 2- Divide the joint distributions in each row by these weights.
- 3- Divide the wives' marginal educational distribution as of 1992 (or 2014) by the marginal distribution as of the year used to standardize the table. We obtain a weight for each education level.
- 4- Divide the joint distributions in each column by these weights.
- 5- Repeat steps 1-4 until the desired marginal distributions are obtained.

Tables 9 and 10 show the standardized table for 1992, obtained using the marginal distributions as of 2014 and the actual joint distribution as of 2014. In Table 9, we estimate the joint distribution as of 1992, using the marginal educational distributions as of 2014, and hold the dependence structure pattern of the joint distribution constant. We assess this analysis by calculating the odds ratio of the joint distribution as of 1992 (using the marginal educational distributions as of 2014) and the joint distribution as of 1992 in Table 15, and observe that it remains unchanged in both cases (Sinkhorn and Knopp (1967), Tan et al. (2004)).

<sup>36</sup> To draw means to remove the couple from the sample.

<sup>37</sup> This algorithm was used by Greenwood et al. (2014).

## C Appendix

In this section, we describe in more detail the variables used in the empirical exercises. We use household sampling weights to construct the variables.<sup>38</sup>

### C.1 family variables

#### C.1.1 Family ID

In Brazil, it is possible to have more than one family in the same household. To avoid counting two families as one, we create an identifier for each family. In the sample, we keep families constituted by couples of one man and one woman. Other family types (families with only one head of household or same-sex couples) are excluded from the sample. To construct the family ID, we use the following variables from the Brazilian National Household Sample Survey (PNAD): control number (v0102), serial number (v0103), the number associated with the household member (v0301), the number associated with the family (v0403), and status within the family (v0402 equal to 1 and 2).

#### C.1.2 Number of children

We add up the number of children in every family using the family ID, gender (v0302), age (v8005) between 0 and 17, and status within the family (v0402 equal to 3).

#### C.1.3 Age

We analyze spouses aged between 26 and 60. Single individuals, when included, are also between 26 and 60 years of age.

#### C.1.4 Levels of education

We construct a “years of schooling” variable to represent 0 to 17 years of schooling and aggregate values of this variable into five groups. In other words, individuals are grouped by the number of years of schooling into five mutually exclusive groups. The first group consists of illiterate (less than one year of education) individuals; the second contains those with elementary school education (4–5 years of schooling); the third comprises those with middle school (8–9 years of schooling) education; and the fourth group contains those who dropped out of and those who graduated from high school (10–12 years of schooling). The last group consists of individuals who had at least

<sup>38</sup> 1) The codes presented in this section originate from the 2014 Brazilian National Household Sample Survey (PNAD). In the other years, we match the variables used in 2014 with their respective codes or similar variables. 2) The household sampling weights' code is v4732. 3) The microdata and questionnaires are available at <https://www.ibge.gov.br/en/np-statistics/social/population/20620-summary-of-indicators-pnad2.html?t=microdados>. For data before 1999, the microdata are available at <https://loja.ibge.gov.br/pesquisa-nacional-por-amostra-de-domicilios-1987-a-1999-microdados.html>. Additional information and questionnaires are available at <http://www.econ.puc-rio.br/datazoom/english/pnad.html>.

some postsecondary education regardless of whether they earned an undergraduate degree or M.A. or Ph.D. degrees (more than 12 years of schooling).

We use the following codes to form this variable: the code for school type and educational stage (v6003), grade attended (v0605), elementary school duration (v6030), highest educational stage attended (v6007), last grade completed of the educational stage attended previously (v0610), elementary school duration (v6070), and being able to read and write (v0601).

### C.1.5 Couple's income

This variable is the sum of individual monthly incomes of the husband and wife. We use the variable monthly income from all sources for individuals aged 10 or above (v4720).

## C.2 Labor counterfactual variables

### C.2.1 Female labor force participation

We develop the female labor force participation counterfactual using the reweighting function shown in equations (7) and (8) and described in Sect. 4.2.1. The reweighting function is the ratio of the proportion of working women in the base year and the actual year of interest for each of 25 combinations of male and female educational levels. Dummy variable  $w$  is 0 if variable v4704 is two; alternatively,  $w$  is 1 if v4704 is one.

### C.2.2 Female wage gap

We develop the female wage gap counterfactual reweighting functions shown in equations (9) and (10) and following equations (7) and (8). They are described in Sects. 4.2.2 and 4.2.1, respectively.

In the first step, we calculate  $\gamma$ , which is the mean for each of 25 combinations of male and female educational levels, of men's and women's incomes (v4720). Next, for every woman, we calculate  $\vartheta$ , which is the product of  $\gamma$  and the income of the woman's husband.

We then create a dummy variable  $\omega$ , which takes the value of 1 if the wife's income is greater than  $\vartheta$  and is zero otherwise.

The reweighting function is the ratio of the proportion of women with  $\omega$  equal to one (or equal to zero) and the total number of women, in the base year and the actual year of interest for each of 25 combinations of male and female educational levels.



### C.3 Marriage counterfactual variables

#### C.3.1 Marital sorting parameter

To construct the reweighting function, we follow equation (15) and steps (1) to (3) described in Sect. 4.2.3, setting  $t_{ij} = t$  and  $t_s = b$ . The methodology for  $s_{ij}$  is described in Sect. 4.1.

#### C.3.2 Random matching

To construct the reweighting function, we follow equation (15), setting  $t_{ij}=t$ , and steps (1) to (3) described in Sect. 4.2.3. In this case, we set  $s_{ij} = 1$ .<sup>39</sup>

### C.4 Educational counterfactual variables

#### C.4.1 Educational composition

To construct the reweighting function, we follow equation (15) and steps (1) to (3) described in Sect. 4.2.3.

#### C.4.2 Returns to education

We use the income distribution in year  $b$ . The reweighting function being calculated is the ratio of the joint distribution of the couples' education in year  $t$  and year  $b$  for all education levels. We then calculate the income distribution using this reweighting function to evaluate the impact of returns to education on the distribution.

## References

- Barbosa ALNdH (2019) Tendências na alocação do tempo no Brasil: trabalho e lazer. *Revista Brasileira de Estudos de População* (35)
- Barro RJ, Lee JW (2013) A new data set of educational attainment in the world, 1950–2010. *J Develop Econ* 104:184–198
- Barros RPd, Franco S, Mendonça RSPd (2007) A recente queda da desigualdade de renda e o acelerado progresso educacional brasileiro da última década. Working Paper 1304, IPEA, Brasília
- Becker G (1991) *A Treatise on the family*. Harvard University Press, Cambridge
- Browning M, Chiappori P, Weiss Y (2014) *Econ Family*. Cambridge University Press, Cambridge
- Charles KK, Luoh MC (2010) Male incarceration, the marriage market, and female outcomes. *Rev Econ Stat* 92(3):614–627
- Chiappori P-A, Iyigun M, Weiss Y (2009) Investment in schooling and the marriage market. *Am Econ Rev* 99(5):1689–1713
- DiNardo J, Fortin NM, Lemieux T (1996) Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica* 64(5):1001–1044
- Eika L, Mogstad M, Zafar B (2019) Educational assortative mating and household income inequality. *J Political Econ* 127(6):2795–2835
- Feng S, Tang G (2019) Accounting for urban China's rising income inequality: the roles of labor market, human capital, and marriage market factors. *Econ Inquiry* 57(2):997–1015

<sup>39</sup> The methodology for  $s_{ij}$  is described in Sect. 4.1.

- Ferreira FHG, Firpo S, Messina J (2017) Ageing Poorly? Accounting for the decline in earnings inequality in Brazil, 1995–2012. IZA Discussion Papers, No. 10656, Institute of Labor Economics (IZA), Bonn
- Firpo S, Reis MC (2007) O salário mínimo e a queda recente da desigualdade no Brasil, Vol. 2. IPEA: in: Barros, R.P., M.N. Foguel and G. Ulyssea (eds) *Desigualdade de Renda no Brasil: Uma Análise da Queda Recente*, IPEA, Brasília
- Goldin C, Katz LF (2002) The power of the pill: oral contraceptives and women's career and marriage decisions. *J Political Econ* 110(4):730–770
- Greenwood J, Guner N, Kocharkov G, Santos C (2014) Marry your like: assortative mating and income inequality. *Am Econ Rev* 104(5):348–53
- Menezes-Filho N, Fernandes R, Picchetti P (2007) Educação e queda recente da desigualdade no Brasil, Vol. 2, Chapt. 25, pp. 285–304. in: Barros, R.P., M.N. Foguel and G. Ulyssea (eds) *Desigualdade de Renda no Brasil: Uma Análise da Queda Recente*, IPEA, Brasília
- Pencavel J (2006) A life cycle perspective on changes in earnings inequality among married men and women. *Rev Econ Stat* 88(2):232–242
- Pereira L, Santos C (2017) Casamentos seletivos e desigualdade de renda no Brasil. *Revista Brasileira de Economia* 71(3):361–377
- Sampaio H (2015) Higher education in Brazil: Stratification in the privatization of enrollment, Mitigating inequality: higher education research, policy, and practice in an era of massification and stratification. *Advances in education in diverse communities*, Vol. 11. Emerald Group Publishing Limited, Bingley, pp. 53–81
- Sinkhorn R, Knopp P (1967) Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J Math* 21(2):343–348
- Tan P-N, Kumar V, Srivastava J (2004) Selecting the right objective measure for association analysis. *Inform Syst* 29(4):293–313

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.