Spouses' Income Association and Inequality: A Non-Linear Perspective

Shoshana Grossbard
Lucia Mangiavacchi
William Nilsson
Luca Piccoli

Working Paper

2019-076

12/2019

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Spouses' income association and inequality: A non-linear perspective[1]

Shoshana Grossbard
San Diego State University, HCEO and IZA. (E-mail: sgrossba@sdsu.edu)

Lucia Mangiavacchi
University of Perugia, Italy, and IZA

William Nilsson
University of the Balearic Islands, Spain

Luca Piccoli
University of Trento, Italy, and IZA

12/10/2019

Abstract

We analyze the association between spouses' incomes using a rank-rank specification that takes non-linearities along both spouses' income distribution into account. We also document that the relationships between income and labor force participation and income and couple formation are non-linear. Using simulations, we then analyze how changes in spouses' rank-dependence structure, labor force participation and couple formation contribute to the upsurge of income inequality in the U.S between 1973 and 2013. We find that an increased tendency towards positive sorting contributed substantially to the rise in inequality among dual-earner couples, but contributed little to overall inequality across households. When considering all households, the factor accounting most for the increased inequality during this period is an increased tendency for individual men and women to remain single.

JEL codes: J12, J22, D31

Keywords: income inequality, assortative mating, female labor supply, marriage, fertility, rank dependence

The authors thank the participants in the 2019 Workshop on Intergenerational Mobility, Gender and Family Formation in the Long Run in Oslo, the 3rd Applied Microeconomics Workshop in Bozen-Bolzano in 2019, the 2018 and 2019 meetings of the Society of Economics of the HOusehold (SEHO) in Paris and Lisbon, the 2018 IZA World Labor Conference in Berlin, and a seminar at the Department of Economics (University of Florence).

1 Introduction

Household income inequality has increased sharply in the United States since the mid-1970s (Piketty and Saez, 2003; Saez and Zucman, 2016). In terms of the most commonly used measure of inequality, the Gini coefficient, inequality increased by more than 30% over the period 1973-2013, climbing from about .45 to about 0.6 (see Figure 1).[1] One of the commonly proposed explanations for this dramatic increase in inequality was inspired by Gary Becker's (1973, 1974) theory of marriage and attributes the increase to a rise in positive sorting by spouses' income. A rise in positive assortative mating has been documented in the U.S. since the 1970s both in terms of spouses' earnings (Burtless, 1999; Schwartz, 2010; Gonalons-Pons and Schwartz, 2017) and education (Schwartz and Mare, 2005; Greenwood et al., 2014; Eika et al., 2019). Does such increased assortative mating account for increased inequality? According to recent papers increased assortative mating has had a relatively small impact on growth in inequality: just about 3% according to Hryshko et al. (2017) and about 5% according to both Dupuy and Weber (2019) and Eika et al. (2019).[2] Earlier studies had attributed more of the change in inequality to increased assortative mating: 10% according to Larrimore (2014), 13% according to Burtless (1999) and 25% according to Hyslop (2001).

Alternatively, a secular positive trend in female labor force participation could have led to increased inequality. According to Greenwood et al. (2014), Gonalons-Pons and Schwartz (2017) and Pestel (2017) the rise in female labor force participation is the main channel linking positive assortative mating to inequality. A third explanation attributes the rise in income inequality in the U.S. to the rise in single-parent households, especially among women with low education (McLanahan, 2004; Western et al., 2008). Burtless (1999) estimates that one-fifth to one-quarter of the jump in inequality in the period 1979 to 1996 was due to the sharp increase in the proportion of single-headed families. Changes in marriage over the period 1960 to 2005 accounted for about

[1] Based on data from the World Inequality Database available at https://wid.world. To give an idea of the magnitude of such an increase, according to Aaberge (1997) it is equivalent to a reform having introduced a lump sum tax of 30% of the average income and a transfer to each family equal to 30% of its actual income. For example, considering an average income of 32,800$ a family with 10,000$ of income would pay 9,500$ and receive 3,000$, while a family with 100,000$ of income would pay 9,500$ and receive 30,000$. The poor family suffered a loss of 6,500$, more than 68% of its income, while the rich family would have a gain of 20,500$, or about 20.5% of its income.

[2] Hryshko et al. (2017) consider wage-based assortative mating and years 1980 to 2009; Dupuy and Weber (2019) and Eika et al. (2019) consider educational assortative mating and the period 1962 to 2013.

18% of the rise in income inequality according to Greenwood et al. (2016) and to about 22.5% of the increase according to Larrimore (2014). Most of the above-mentioned literature has examined changes over the entire income distribution, overlooking some important non-linearities. Over the period under study a disproportionate part of the inequality action occurred at the tails of the income distribution. Figure 1 shows that the share of income owned by households at the top 10% of the income distribution increased from 35% to 45%, while the share of income owned by households at the bottom 50% of the distribution shrunk from about 20% to 13%.

In this paper we first document that assortative mating by income, the relationship between income and labor force participation, and the association between income and couple formation are of a non-linear nature in both 1973, our starting point, and 2013, the last year we examine.[3] We then investigate how changes in these non-linear associations help account for increased inequality over the period 1973 to 2013.

This paper's first methodological innovation is a new way of evaluating the non-linear impact of changes in each factor (sorting by income, labor force participation and couple formation) on changes in income inequality over time: Monte-Carlo-like simulations of the 2013 income distribution using one-to-one matching coupled with regression predictions that capture non-linearities found in 1973. This method differs substantially from the widely used semiparametric decomposition method proposed by DiNardo et al. (1996), which is based on reweighting specific sample groups to obtain counterfactual distributions.

A second methodological innovation consists of assessing rank dependence of spouses' income using an extension of the rank dependence method that Dahl and DeLeire (2008) and Chetty et al. (2014) implemented in their studies of intergenerational mobility. While the rank dependence method focuses on the mean correlation, we propose to apply the same technique to different quantiles of the distribution. The resulting visual improvement highlights the non-linear rank dependence structure simultaneously along both spouses' income distributions.

We analyze three increasingly general samples extracted from the March Current Population Survey (CPS) data in 1973 and 2013: dual-earner couples, all couples, and couples and singles of working age. We find that in the case of dual-earner couples the increase in spouses' income association accounts for almost 26% of the increase in inequality measured in terms of Gini

[3] The following studies take into account that changes in women's labor force participation varied across husbands' income distribution: Cancian and Reed (1998) and Bredemeier and Juessen (2013).

coefficient, which is in line with early findings by Burtless (1999) and Hyslop (2001). In contrast, when non-working spouses and singles are also included in the analysis, replicating the 1973 spouses' income association patterns for 2013 accounts for only about 3.5% of the increase in inequality over the period. This is in line with recent findings by Hryshko et al. (2017), Dupuy and Weber (2019) and Eika et al. (2019). When inequality is measured in terms of income shares, our results suggest that the increased spouses' income association has been particularly relevant to the relative worsening of the condition of the bottom 50% of the income distribution, contributing to almost 26% of the decrease in the income share of the poorest dual-earner couples. In contrast, increased spouses' income association contributed only 15% of the increase in the share of the top 10%.

When we consider couples and singles, assuming that the probability of couple formation (including marriage) in 2013 stayed the same as in 1973 accounts for about 13% of the increase in inequality, which is more than the 7% that Eika et al. (2013) accounted for. Again, a large portion of the change can be attributed to changes in the lower part of the income distribution. In line with Greenwood et al. (2016), we find that the increase in female labor force participation had a negligible impact on the increase in inequality.

In interpreting these results, the reader should be aware that the simulations performed here are of a descriptive nature and are not meant to indicate causal relationships. The rest of the paper is organized as follows. Section 2 describes the empirical strategy, including data details and methods. Section 3 presents the results and Section 4 concludes.

2 Data and methods

2.1 Data and sample selection

We use data from the 1974 and 2014 March Annual Social and Economic Supplement (ASEC) regularly added to the Current Population Survey (CPS), a monthly survey of labor force participation carried out by the Minnesota Population Center (Flood et al., 2017). Although the survey's main purpose is to collect information on employment, it also collects information on the demographic status of the population, including age, sex, race, marital status, educational attainment, and family structure. The March income supplement to the CPS contains information about income earned in the previous calendar year. We start our analysis in 1973, a symbolic date chosen as a tribute to the first economics article on income sorting in marriage: Gary Becker's (1973) theory of marriage. Our analysis ends in 2013, exactly 40 years later.[4]

For each person in the sample who is 15 or older, the data include information on the amount of money income that individuals earned in the preceding calendar year, including wage and salary income, business and farm income. Observations with negative incomes are excluded.[5] To compute single household equivalent household income we use number of members belonging to the main family as a group of two persons or more (one of whom is the householder) residing together and related by birth, marriage,[6] or adoption. We then apply the square root equivalence scale to obtain the equivalent income.[7] We select individuals between 30 and 55 years old living in couples or in single households, with or without children.[8] In 1973 couples are all married couples.[9] In 2013 couples also include non-marital cohabitants, for by then the CPS records non-marital cohabitation. According to the 2010 census 5.3% of all couples were unmarried. Counting them as singles would have likely biased our simulation results. Our decision not to distinguish between married and unmarried cohabiting couples in 2013 is also rooted in the fact that in 2013 about eleven U.S. States recognized unmarried couples as common-law marriages (see Grossbard and Vernon 2015).

In 1974 (for 1973 incomes) the final sample of couples and singles consists of 18,899 households (41.7% of the original sample), the sample of all couples of 14,372 observations, and the sample of dual-earner couples of 7459 observations. For 2014 (and 2013 incomes) we have 35,214 households (47.5% of the original sample) including both couples and singles, 22,761 couples, and 15908 dual-earner couples.

[4] The CPS is only available since 1962. A robustness analysis conducted with different starting and ending years leads to little variation in the results.

[5] In each sample they are less than 1%, so their exclusion does not affect the results.

[6] Marriage or non-marital cohabitation.

[7] Compared to other popular equivalence scales (e.g. the OECD or the OECD-modified scales), which give different weights to adults and children, the square root scale allows us to simplify calculations of inequality measures in our simulations, where counterfactual scenarios imply variations in household composition, both in terms of presence or not of a spouse and by number of children. The choice of the square root scale is common for U.S. studies (Burkhauser et al., 2012; Larrimore, 2014). As a robustness check we also perform simulations using the two extreme scales, i.e. per-capita income (each family member is given the same weight) and household income (no equivalence scale is applied). We interpret the corresponding results as upper and lower bounds of the effect of income distribution attributable to changes in the family structure.

[8] Selection based on individual characteristics, means that complex households are not dropped from the data, but only the main family within the household is used. This has implications for the computation of household income as contributions provided by other members of the household are not included in the analysis.

[9] Unmarried couples were not identified in 1973, as such status was not recorded in the CPS, but estimates suggest that the prevalence of unmarried couples was limited to about only 1% of all couples (Fitch et al., 2005).

Table 1 reports some descriptive statistics for the three samples for 1973 and 2013. For the sample of dual-earner couples the mean equivalent income increased by 36% in the period, while the Gini coefficient increased by 28% ramping up from 0.279 to 0.357. The top 10% income share increased by 26.5%, from 0.219 to 0.277, while the bottom 50% income share decreased by almost 15%, from 0.309 to 0.263. Considering all couples, income increased by slightly less than 30%; inequality as measured by the Gini coefficient, started from a higher level of 0.319 and increased by almost 33%, to reach 0.424. For this sample the top 10% income rose from 0.236 to 0.306, about 30%, while the bottom 50% declined from 0.284 to 0.216, a reduction of 24%. In the most comprehensive sample, which also includes singles, the average income increased by just less than 20%, but the Gini coefficient which starts at 0.379 still had a big increase of 26.6%, to 0.48. The top 10% share on income increased by more than 28% and the bottom 50% share of income shrank by more than 27%, down to 0.175 from 0.241. Changes in the magnitude of inequality for the general population as measured by the Gini, are in line with our estimates, but the changes for the general population are somewhat smaller when measured in terms of the top 10% and bottom 50% income shares (according to WID.World data and reported in Figure 1).

2.2 Rank dependence analysis

Studies of spouses' income association have generally focused on correlation measures over the entire income distribution (Hyslop, 2001; Schwartz, 2010; Hryshko et al., 2017; Gonalons-Pons and Schwartz, 2017; Pestel, 2017). This leads them to overlook possible non-linearities along individual earnings distributions. Such non-linearities have been shown to be important, e.g. by Bredemeier and Juessen (2013) who found that from the 1970s to the 2000s the most pronounced increase in US wives' hours spent in the market occurred for wives of high-wage men.

To study how income association in couples varies along men's and women's income distributions we adopt the binning technique, an approach developed in the literature on intergenerational income mobility (Dahl and DeLeire, 2008; Chetty et al., 2014; Bratberg et al., 2017). These articles use bins based on parents' income ranks and for each of the bins calculate the average income rank of their children. There are two main advantages of using bin means based on income rank instead of correlation coefficients. First, a correlation coefficient can be zero, not capturing a non-linear dependence structure such as positive sorting in the lower part of the distribution and negative sorting in the higher part. Second, binning by rank can easily incorporate individuals with zero income.[10] At the same time the binning technique also has some disadvantages: it produces discontinuous results that are similar to a scatter plot and it allows non-linear patterns to emerge only along the dimension of one variable. For instance, in their applications the average rank of children varies non-linearly with parent's rank, but within each bin all children are averaged out. One possible solution is to compute several quantiles of the child distribution within each bin of parents' income rank. But this would reduce graphical clarity: dots' overcrowding could impede the detection of relevant patterns. Instead, we propose a visual improvement that consists of plotting nonlinear functions that smoothen the information embedded in binned means and quantiles.

First, we analyze the rank dependence[11] structure of spouses' incomes using a continuous mean function of one spouse's income rank with a polynomial of the partner's rank on the right-hand side, i.e.

$$r_{w,i} = \alpha + \sum_{k=1}^{K} \beta_k \, r_{h,i}^{k} + e_i, \qquad (1)$$

where $r_{w,i}$ is wife's income rank for family i, and $r_{h,i}$ is her husband's rank. Husband's and wife's income ranks can be swapped. Estimating equation (1) by Ordinary Least Square produces a *continuous mean function*, as parameters are estimated in order to minimize the (squared) deviation from the conditional mean. The mean function is a plot of $\hat{r}_{w,i} = \hat{\alpha} + \sum_{k=1}^{K} \hat{\beta}_k \, r_{h,i}^{k}$. The flexibility introduced with a polynomial allows the mean function to illustrate a nonlinear dependence pattern along the husband's income rank distribution.

Second, we estimate equation (1) with quantile regressions (Cameron and Trivedi, 2005). In contrast to OLS, estimators of quantile regressions minimize the (absolute) deviation from a given quantile of the dependent variable distribution. Such estimators allow us to produce a series of non-linear functions that are centered on different quantiles of income rank on the left-hand side of equation (1). These functions show how a wife (husband) positioned at a given quantile of her (his) income rank distribution depends non-linearly on their husband's (wife's) income rank. These *rank dependence curves* allow us to detect different rank dependence structures at different positions of each spouse's income rank distribution.

[10] The correlation coefficient can, of course, also be calculated including zeros, but for zero-inflated distributions it should be computed using appropriate techniques, otherwise non-uniform concentrations of zeros would imply possibly biased estimates.

[11] The technique's name 'rank dependence' originated because it was about how children's income possibly depend on those of parents. In our case, we focus on spouses and the term 'rank association' is more appropriate than 'rank dependence'. We use both terms.

2.3 Labor force participation and marriage.

We want to establish whether there are non-linearities in the relationships between spouses' income rank, labor force participation, and marriage.

We estimate the probability of not having a spouse and spouse's labor force participation (defined as probability of having a spouse with zero-income from work, business or farm) as functions of the logarithm of own income using local linear nonparametric regressions (Fan and Gijbels, 1996)[12], i.e.

$$y_i = g(\mathbf{x}_i) + e_i. \tag{2}$$

Local-linear regression estimates a regression similar to OLS, but a kernel function and a bandwidth assures that observations closer to the evaluation point in the distribution of the explanatory variable receives more weight. In our application, all of the values of **x** are used as evaluation points. This allows us to have an estimate of the relationship for each rank, and thus the resulting estimates capture any sort of non-linearity in the relationship between the outcome and the explanatory variable.[13]

2.4 Simulations

Existing research about the factors possibly contributing to the increased earnings inequality in the U.S. has used two main methodological approaches: variance decomposition and simulations of counterfactual scenarios that are then compared to actual inequality measurements. Inequality is often measured by Gini coefficients, and we also do so, but since changes over time may possibly be concentrated at the top or bottom of the distribution, we also use the top 10% share of income, that is the share of population income earned by families at the top 10% of the income distribution, and the bottom 50% share of income. In order to account for non-linear patterns of association between spouses' income we follow a counterfactual approach simulating the 1973 non-linear patterns for 2013 data. To better highlight the contribution of each change on inequality we use different sample selections: i) dual-earner couples only; ii) all couples, including those with non-working spouses, to account for changes in female labor force participation; and iii) all couples

[12] We apply standard statistical techniques, our goal being to produce reliable simulations, not to identify causalities. More specifically, we use the *npregress* command in Stata 15.

[13] A robustness check performed using a 4th degree polynomial shows no relevant differences in the predicted functions. Results are available upon request.

and singles, to account for changes in couple formation. We examine how income association accounts for changes in inequality in all three samples.

Our strategy is somewhat similar to that of Burtless (1999), who analyzes changes in Gini coefficient for equivalized adult income in 1996 assuming the rank association and couples' proportion of 1979. The advantage of a simulation procedure is that it makes it possible to incorporate variation along income distributions in the counterfactual situations being examined. This technique allows us to avoid imposing randomization (as commonly done in the existing literature) or a particular degree of association (such as a specified correlation coefficient) as the hypothetical reference case. The details of each step of the simulation are presented in Appendix A. Here we just present the intuition behind the simulation technique.

Since the number of observations changes substantially from 1973 to 2013 we randomly extract 200 samples from 2013, each composed of the number of observations available for 1973.[14] This allows a one-to-one match for probabilities of not being married and having a spouse with no income from work and income association by rank for both spouses' income distributions. For example, to find a counterfactual spouse's income for a male in a particular rank in 2013 we search for a man at the same rank in 1973 and find his wife's rank. We then find a counterfactual spouse who has the same rank in 2013 as the actual 1973 wife.[15] This procedure is also used when spouse's income includes imputed values.

The procedure used to simulate the probability of having a spouse with zero earnings and that of being in couple is different and instead of using one-to-one matching we apply predictions based on the regression models proposed in Section 2.2.

Each of these counterfactual distributions can be included separately or jointly, given that in each simulation estimated probabilities and income association can come from either 1973 or 2013. For example, if only income association is to be evaluated, the sorting is simulated according to the 1973 patterns, but the estimations for 2013 are used to predict having a zero-earning spouse and not being in couple. This allows us to evaluate how each single change in income association, household composition and labor force participation contributes separately to earnings inequality in 2013. When comparing the simulated income distribution in 2013 to that of 1973 we compare

[14] This ensures that the results are not driven by sample selection and maintains the statistical properties of the results, which are the average of the 200 repetitions.

[15] For example, to assign a counterfactual spouse to a husband ranked #2654 in males' income distribution in 2013, we find a man ranked in exactly the same position in 1973 and find the rank of his wife, say #1896. The counterfactual wife for the 2013 husband ranked #2654 is a wife ranked #1896 in 2013.

the Gini and income shares of the reduced sample in 2013 to that of the actual 1973 sample and assume that the same change observed for the reduced sample (averaged across the 200 repetitions) applies to the actual Gini and income shares.

3. Results

Section 3.1 presents results related to the evolution of husbands' and wives' income ranks association (dependence) for heterosexual couples, the probability of having a spouse not working outside the home, and the probability of being single. Section 3.2 focuses on how such changes account for the evolution of inequality among couples and singles based on a simulation analysis.

3.1. *Changes between 1973 and 2013 as a function of income rank*

3.1.1 Association between spouses' income rank.

Figure 2 and Figure 3 present measures of the association between spouses' income distributions for samples of U.S. married couples in 1973 and 2013. Each graph presents the three different measures of such association mentioned in Section 2: (1) binned means (gray dots); (2) mean function (solid black line), and (3) rank dependence curves (solid gray lines). Vertical lines separate the 'rich' from the 'poor'. It is evident from the figures that spouses' income association patterns are highly non-linear in both 1973 and 2013. The slope indicating rank dependence or spousal income association often changes as a function of one spouse's income rank.

Figure 2 (Figure 3) plots income dependence patterns as a function of men's (women's) income rank. In 1973 (Figure 2, left panel) the overall correlation coefficient between spouses' incomes is close to zero (-0.04) even though the spouses' income ranks are associated either positively or negatively at most income ranks. The interpretation of the rank dependence curves is as follows. The y-axis is the female rank in a global sense, i.e. it indicates the rank positions in the income distributions of all female spouses. The curves indicate a local ranking, each curve representing a decile. Consider D6 and men at rank 90 in the male distribution as an example. We see that if his spouse is locally on decile 6, i.e. has a rank of 60 among the subset of spouses married to men ranked around 90, this corresponds to a global ranking of about 45. The negative slope of the curve indicates that an increased male rank is associated with a reduced global spouse rank, while the local rank is maintained at 60. This indicates a negative dependence structure. Now, consider D9 and men at rank 90. A female spouse who is locally at rank 90 is also globally at rank 90. The slope is negative, but its magnitude is much lower than at D6, implying less of a negative dependence structure. Rank dependence curves show that for rich men there is a negative association between the spouses' income. Negative sorting is apparent for men with an above-average income rank: the higher their rank the lower their wife's rank is likely to be. This negative association characterizes all couples with rich men but is least pronounced for rich men married to the richest women (women at the top female income rank, the ninth decile D9). In contrast, for poor men, especially those in the bottom two deciles, the slopes of the rank dependence curves are positive, indicating positive sorting. As men move up in income rank they are more likely to have a wife at a higher income rank. Note that the rank of women at the bottom 40% (D1 to D4) is not correlated with husband's rank as all these women are out of the labor force and have zero income.

In 2013 non-linearities continue to dominate the rank dependence patterns organized by male income rank (Figure 2, right panel). If we repeat the analysis for men in rank 90, we find that a female spouse locally at rank 90 (D9) is found globally at about rank 95, and the slope is positive. An improved male rank would increase the spouse's global ranking, even though the local position is kept at 90. For D6, we find a global rank of almost 70, but the slope is quite flat: an increased male rank would not improve that ranking. For D3, indicating a spouse's local ranking of 30, and men at rank 90, the spouse's global rank is about 25. While the local rank is kept at 30, the slope of the curve is negative, and an improved male rank would reduce the global rank even further. These examples show that the complexity of the dependence goes far beyond what can be captured by a correlation coefficient.

Rank dependence curves have positive slopes, indicating positive income sorting, in the case of (a) relatively poor men with an income rank to the left of the 60th percentile, regardless of women's rank, and (b) couples with relatively 'rich' husbands and wives (i.e. in the 6th decile or higher). However, there is heterogeneity in the case of rich men: slopes of the rank dependence curves are negative for rich men married to women with relatively low income (between the 3d and 5th deciles), indicating negative sorting. It thus appears that in 2013, from the perspective of men's income rank, there is positive sorting for a majority of couples, with the exception of rich men in couple with women from relatively low-income ranks.

Overall, during the forty years separating the two panels in Figure 2 for most groups there was a switch from negative to positive sorting at most income ranks. The only group characterized by negative income sorting in both years are rich men in couple with women in the lower deciles of their distribution. The only group characterized by positive sorting in both years are couples with men in the bottom 20 percent of the distribution.

Figure 3 is similar to Figure 2 but takes female rank as its starting point and places it on the horizontal axis. In 1973 the continuous mean function has a U-shape with a slightly negative association for low to medium female income ranks (up to about the 60th percentile). The association then turns strongly positive for higher female ranks. This indicates that at the mean there was slightly negative sorting for women at low income ranks and strongly positive sorting for those at high income ranks. This positive sorting at higher female income ranks is compatible with the flat line for D9 in Figure 2. Conditioning on female can thus produce results very different from those obtained conditioning on male rank.[16]

The quantile rank dependence functions defined by women's rank also vary depending on whether women are 'rich' or 'poor.' For 1973 in Figure 3 a 'rich' woman is defined as one belonging to the 60th percentile or higher. It is apparent from Figure 3 that in 1973 U-shaped curves prevail at all ranks of the male income distribution except for the 1st decile (D1), where the positive spouses' income association at top female rank is weak and the rank dependence curve is more like a line. This indicates that most men in the lowest income decile have low earnings and that this varies little with their wives' income rank.

By 2013 the right panel in Figure 3 displays a mean function that no longer has a U-shape. On average, spousal income association is close to zero but this hides heterogeneity in spousal income association. Positive sorting kicks in at the 40th percentile of the female income distribution. For female ranks above the 60th percentile there is positive sorting at all male ranks in both 1973 and 2013, except for the lowest decile of men. The positive slopes on the right side of the 40th percentile line in Figure 3 (2013) are compatible with the positive slopes on the right side of the 60th percentile line in Figure 2 (2013) for most income ranks.

The overall increased prevalence of positive sorting over time that we document was also found in most of the previous literature (e.g. Burtless, 1999; Schwartz, 2010; Gonalons-Pons and Schwartz, 2017; Eika et al., 2019). However, our results also indicate some negative sorting.

[16] For example, the further to the right in Figure 3, the more selectively we condition on a high female rank, and D9 indicates the 9th decile of the spouses to that more selective group, while the y-axis indicates the position in the overall male distribution. In Figure 2, on the other hand, the higher rank of the males makes the corresponding selection stronger for males. Evaluating D9 for male spouses at female rank of 90, is certainly not the same as evaluating D9 for female spouses at the male rank of 90, because the selected groups are very different.

In low ranges of the female income distribution, to the left of the 40th percentile, a negative slope prevails at high male income ranks, indicating negative sorting between high income men and low-income women, while the rank association curves are upward-sloping for lower male deciles in the right panel of Figure 3.

In 2013 the pockets of negative income sorting are thus concentrated among (a) couples with high male rank (P60 and up) and relatively low female rank (D3 and D4) in Figure 2, (b) women with low income ranks (P40 and below) and men above the median in Figure 3, and (c) couples with women above the 80th percentile and men in the lowest 10th percentile in Figure 3. The presence of such negative sorting among the lowest 10th percentile of men and the top 20th percentile of women in 2013 suggests a new form of specialization among couples with top-earning women and low-income men. These men may not earn much because they work part-time and are relatively more involved in household production.

3.1.2 Labor force participation.

Figure 4 (Figure 5) shows two non-linear functions at two points in time. First, the predicted probability of a married woman (man) having zero income from labor as a function of husband's (wife's) income rank is indicated by a dark curve. Second, the predicted probability of a man (woman) being single as a function of own income rank is shown in light color. The left panel is for 1973 and the right panel for 2013. A comparison of the darker curves on the left and right panel in Figure 4 indicates that the relationship between wife's probability of having zero income and husband's income rank varied drastically over time. In 1973 a married woman's probability of not having a job does not change with husband's income rank until approximately the 40th percentile of male rank. After that point it increases rapidly with husband's income rank. In contrast, in 2013 the relationship between men's income rank and women's likelihood of being out of the labor force is almost flat, with two small peaks for women married to (1) men with the lowest income ranks and (2) men in the top decile. The drop in married women's likelihood of opting out of the labor force over the period was most striking for the wives of men in the top 4 deciles of the income distribution. This is consistent with recent research on opting out of the labor force being associated with women's elite education and the possible role of elite universities as facilitating marriages involving high-income men (Hersch 2013).

Figure 5 shows the same predicted probabilities as Figure 4, but with female income rank on the horizontal axis. Husbands have a very low probability of zero income and it does not vary much with wife's income rank in both years. However, the probability is slightly higher in 2013 than in 1973.

The changes in income association (rank dependence) illustrated in Figure 2 and Figure 3 are likely to be related to the changes in probability that spouses earn an income documented in Figure 4 and Figure 5. As positive sorting became more common at high male income ranks it also became less common that men with high income had wives out of the labor force. At the same time, it continues to be rare for women, even for women at the top of the female income distribution, to have husbands out of the labor force.

The existence of pockets of negative income sorting in 2013, mostly when one spouse has a high-income rank, and the upward-sloping part of the probability that married women have zero income in 2013 both suggest that specialization in household production continues to be strongly associated with gender. Men at the highest income ranks are still more likely to be married to women who earn less if they are in the labor force (due to fewer hours of work or lower wages) or don't participate in the labor force. This may be the result of a persistent gender pay gap still favoring men.[17] Alternatively, women's lower wages may be a result of their lower labor supply, and in turn this lower supply may be related to continuing demand for women's work in household production (see Grossbard-Shechtman, 1984; Grossbard, 2015).

3.1.3. Couple Formation.

Another variable that changed dramatically over this period is couple formation, which in both years mostly took the form of marriage. In 1973, in the age range we selected, 75.5% of all households were married couples.[18] By 2013 that was only the case with 61.1% of all households, including 3.2% of cohabiting couples That is also evident from Figure 4 and Figure 5 in which the light color curves show how the probability of being single varied with income rank in 1973 and 2013. In addition, Figure 4 shows that men's predicted probability of being single was a negative function of their income rank in both 1973 and 2013. On average that probability had a steeper slope in 2013 than in 1973.

Women's probability of being single as a function of their income rank is shown in Figure 5. In 1973 women at higher income ranks were less likely to be married than they were likely to be

[17] Over this period the gender pay gap shrank but is still significant: it stood at 38.1% in 1973 versus 17.9% in 2013.
[18] This figure does not include unmarried couples –about which no information is available--, but estimates suggest that they are about 1% of all couples.

in couple in 2013: at the highest income rank women had a probability of being single that was close to 40 percent. This stands in strong contrast to men's probability of being single in 1973, which is a negative function of their income rank. Women's probability of being single and its association with income change dramatically from 1973 to 2013. By 2013 women's probability of being single has a slightly negative slope: women at higher income ranks are less likely to be single. For example, the in-couple rate is approximately 76.7% for women at the top income rank (P90) in 2013, up from a marriage rate of approximately 68.1% in 1973. An opposite trend can be observed at the 20th Percentile (P20): the in-couple (marriage) rate went down over time from approximately 82.9% to approximately 72%. The probability of being single thus decreases with income rank in three of the four cases we examine: men in both 1973 and 2013 and women in 2013. This contrasts with the positive slope of the likelihood of being single as a function of income rank observed for women in 1973.

This implies that as we compare 1973 and 2013, income inequality among couples could have grown because proportionally more high-income women and fewer low-income women are in couple in 2013 than in 1973. Consequently from 1973 to 2013 inequality would have increased more at the higher income ranks such as the 80th percentile than at the lower income ranks such as the 20th percentile.

3.2 What accounted for *changes in inequality: a simulation analysis*

3.2.1 Simulation results when inequality is measured by the Gini coefficient

We have thus established that factors that could have possibly affected the evolution of income inequality among married couples in the period 1973 to 2013 --spouses' income association, spouse's likelihood of opting out of the labor force and likelihood of being single-- are related to spouses' income ranks in a non-linear way. The following simulation analysis of inequality in 2013 as a function of 1973 characteristics takes account of these non-linearities.

We perform simulations for the three different samples: dual-earner couples, all couples, including those with one or no earners, and all households (couples and singles). For dual-earner couples we perform simulation model S1 that simulates income association between husband and wife using counterfactual income rank associations from 1973 (Panel A of Table 2). In Panel B we report simulation models for a sample that includes zero-income spouses. This allows us to also include counterfactuals regarding the likelihood of opting out from the labor force. Here we report results for the following simulation models: Model S2 assuming the income rank associations of 1973 and Model S3 assuming 1973 probabilities of opting out of the labor force. In Panel C we report results for a sample that includes all households: couples and singles. Here we report results for Models S4 to S6. Each model is based on counterfactuals for a particular variable: spouses' income sorting, opting out of labor force, and couple formation.

For each sample the first two rows of Table 2 report the actual Gini coefficients in 1973 and 2013. The following rows report results based on simulation models: in column 1 the Gini coefficient; in column 2 the percentage change in Gini relative to the 2013 Gini; in column 3 the percentage change in Gini relative to the 1973 Gini; and in the last column the contribution of the simulation model to the change in Gini coefficient between 1973 and 2013.

Sorting by income. For dual-earners couples (Panel A of Table 2) the actual Gini coefficient was 0.279 in 1973 and 0.357 in 2013. Column 3 reports the percentage change in Gini between 1973 and 2013, amounting to almost 28%. The third line reports that the Gini obtained from simulating scenario S1 "assortative mating" is 0.340. This means that if the income association in 2013 would have followed the same non-linear patterns observed in 1973 the Gini would have been 4.8% lower. The Gini increase from 1973 would only be 21.8%, implying that in the case of dual-earner couples the rise in income association between 1973 and 2013 contributed 22.2% of the Gini increase over the period.

Panel B in Table 2 reports the results for all couples, regardless of whether both spouses were in the labor force or not. Not surprisingly, the actual 1973 and 2013 Gini coefficients for this sample are substantially higher than those for the sample of dual-earner couples, rising to 0.319 in 1973 and to 0.424 in 2013. For this sample the Gini thus increased by 33% over the period. When income association replicates the 1973 distribution (model S2) the estimated Gini in 2013 amounts to 0.415 which is 2.25% lower than the actual Gini. By using 1973 values of rank dependence in spouses' income we miss about 10% of the increase in Gini, which implies that changes in assortative mating by income contributed about 9% of the change in inequality measured by Gini.

Finally, Panel C reports results for all couples and singles. The inclusion of singles further increases the actual Gini coefficients, which rise to 0.379 in 1973 and 0.48 in 2013, an increase of 29.3% over the period. Simulating spouses' income association as in 1973 (model S4) the Gini is

0.475, about 1% lower than the actual Gini, implying a contribution of less than 4.5% to the increase in the Gini over the period.

To summarize our findings regarding possible effects of spouses' income association on inequality, when applied to dual-earner couples our simulation models replicating 1973 income association patterns indicate that changes in assortative mating explain an important fraction (25%) of the increase in inequality over the period. This finding is in line with earlier findings on the contribution of assortative mating to the rise in inequality such as Hyslop (2001). When applied to all couples—including those with only one earner--similar simulations only explain about 9% of the increase in inequality over the period 1973-2013. The contribution to the inequality increase is further reduced to about 4.5% when considering all couples as well as singles. This is in line with more recent findings, including Hrysko et al. (2017), Dupuy and Weber (2019) and Eika et al., (2019), and suggests that a major source of large differences in estimates lies in differences in how various researchers selected their samples. For instance, Eika et al. (2019) use a sample of all couples, whereas Hyslop (2001) use a sample of dual-earner couples. As suggested by our analysis enlarging the sample from dual-earner couples to all couples reduces the contribution of income association by almost 60%, from 22.2% to 9%. Of course, other differences in the applied methods may be relevant, but sample selection seems to be a key factor.

Labor force participation. Simulation model S3 shows that if the distribution of non-working spouses had been that of 1973 (when the proportion of dual-earner couples was considerably lower) the Gini would have increased even more than it actually did: it would have stood at 0.433 instead of 0.424. Changes in the likelihood of opting out of the labor force thus contributed -7.82% to the change in Gini coefficient over the period. That actual increases in spouses' labor force participation (mostly on the part of married women) contributed negatively to the inequality upsurge implies that increased female labor force participation had an equalizing effect. A similar result is also reported in Greenwood et al. (2014), while Greenwood et al. (2016) find no impact of female labor force participation on inequality. Simulations reported in Panel C of Table 2, using a sample that includes both couples and singles, indicate a much smaller equalizing effect (-2.1%). This is to be expected given that most singles are in the labor force.

Couple Formation. The final set of simulations (S6) analyzes the degree to which changes in the prevalence of couple households accounts for changes in inequality over time. Here we use the sample of all households. Our results suggest that the increase in the proportion of singles contributed substantially (almost 12%) to the increase in inequality measured by the Gini coefficient, making it the most important factor considered in this study. This result is in line with findings by Burtless (1999) and Greenwood et al. (2014, 2016), although we find a smaller relative contribution to the inequality increase (they found relative contributions of 19.6 and 25%, respectively).

3.2.2 Simulation results when inequality is measured by Income shares

Sorting patterns. Table 3 reports the actual and simulated shares of income of the top 10% in 2013 if assortative mating were as in 1973. For double-earner couples (Panel A) the actual income share observed in 2013 is 27.7% and it would have been 26.8% if assortative mating by income rank had been the same as in 1973. Considering that the top income share was 21.9% in 1973 implies that changes in assortative mating accounted for about 15.1% of the increase in the top 10% income share over the period. In the case of all couples (Panel B) assortative mating only explains 9.3% of the increased inequality. As for the sample of all households (Panel C) changes in assortative mating explain just 5.9% of changes in the share of income going to the top 10 percent of households.

Comparing these figures with the results of Gini simulations indicates that assortative mating by income contributes less in the case of the sample of double earner couples (it contributed 25% of change in the Gini), but it accounts for more of the change in inequality in the case of the more inclusive sample.

Simulations of assortative mating on the share of income owned by the bottom 50% of the population are reported in Table 4. It can be seen that changes in assortative mating account for about 25.8% of the drop in the share of the bottom 50 percent over the period in the case of dual-earner couples (Panel A), 8.8% of that drop in the case of all couples (Panel B) and 3.5% of the drop in the case couples and singles (Panel C).

Discussion. Over the 40 years we examine patterns of assortative mating evolved from a clear negative sorting for the top 10% of the distribution in 1973 to a more differentiated pattern in 2013, with negative sorting for low rank women and positive sorting for high rank women. Positive sorting by income at the top of men's distribution (see Figure 2) contributed relatively little to the increase in income share of the top 10% dual-earner households, as more than half of those families maintain a negative sorting. As expected, for the bottom 50% changes in sorting by spouse's income had a large impact on inequality among dual-earner couples, as income association went from almost flat to sharply positive for women in any rank position.

Compared to results for dual earners, changes in assortative mating by income accounted for less of the rise of the top 10% and of the impoverishment of the bottom 50% in the case of all couples and all households, including singles. This applies more to the bottom 50% than to the top 10%. This result can be explained as follows: the proportion of dual-earner couples is more than double among the top 10% of households (about 67.5% of all top 10% households) than among the bottom 50% (only about 22.2% of the poorest 50%). Among dual-earner families belonging to the bottom 50% of the income distribution the role of income association is strong. However, these couples are only a small fraction of the sample: the importance of assortative mating vanishes rapidly as other household types enter the sample, namely couples with only one worker and singles for which such income association does not apply by definition. In the case of top 10% dual earners, the relative importance of spouses' income association is also low because individual incomes are already very high so the second income is adding proportionately less to household income. However, as dual-earner couples are a large proportion of the sample, the relative importance of income association decreases more slowly as single-earner couples and singles are included.

Labor force participation. The contribution of increased female labor force participation to increased inequality measured in terms of income shares is in line with that contribution when the Gini coefficient is used to measure inequality (Table 3 and Table 4, panel B). Regardless of the measure of inequality for the sample of couples changes in labor force participation don't explain much of the increased inequality: had labor force participation remained at the 1973 levels the top 10% of the income distribution would have controlled even more of the income (31.2% vs 30.6%) and the bottom 50% of the income distribution would have a slightly lower share of the income (21.1% vs 21.6%). When singles are also included (Panel C in Tables 3 and 4), the results are similar but of a smaller magnitude, especially for the bottom 50% income share, as shown in Table 4, Panel C.

Couple formation. It can be seen from simulation S6 in Panel C of Table 3 that if the probability of being married in 2013 had matched that of 1973, the share of income owned by the top 10% of households would be smaller (32.3% vs 32.8%), implying that the reduced probability of being in couple contributed about 6.6% to the increase in the income share of the highest income decile. Changes in probability of being in couple played an even more important role when it comes to the evolution in the share of income going to the bottom 50% of the distribution: it can be seen from Table 4 (Panel C) had that probability remained at its 1973 level the bottom 50% would have obtained 18.4% of the income. In fact, they only obtained 17.8% in 2013, implying that changes in couple formation accounted for about 13.1% of the change in inequality measured in terms of the share going to the poorest half of the population.

3.2.3 Robustness

The results reported in Table 2 to 4 were obtained using an equivalized income computed as household income divided by the square root of household size. This is one of several possible choices and it may influence the results mainly for two reasons: (1) household size is not constant along the income distribution thus inequality measures could be affected by choice of equivalence scale; and (2) fertility patterns are not constant over time, hence the reduction in the fertility rate observed during the period under study could account for part of the inequality increase.

Equivalence scales. The use of equivalence scales is almost ubiquitous in the analysis of economic inequality, for individuals with the same income living in different-sized households are not equally well off. Thus, the use of equivalence scales helps making such individuals comparable. However, equivalence scales introduce a level of discretionality that directly affects the measurement of inequality. The literature has proposed a number of different equivalence scales based on the demographic composition of households (Lewbel and Pendakur, 2008). So far, we have used one of the simplest ones: the square root scale, which divides household income by the square root of household size. To analyze the sensitivity of the simulation results reported in Table 2 to 4 to the method by which equivalence scales are computed we report simulations performed for the Gini coefficient with the two most extreme equivalence scales: i) no equivalence at all, i.e. simulations are conducted on total household income, and ii) using the household size equivalence scale, which translates household income to per-capita income.

Results are found in Table 5 and are compared to those in Table 2. Consider the Gini coefficient for dual-earner couples in 1973: the Gini is 0.279 when using the square-root scale (Table 2), 0.249 with no scale, and .34 with the per-capita scale (Table 5). This implies that this inequality measure is 36.5% larger when using per-capita versus household income. By 2013 the differences between the two extremes is much smaller (about 11%, based on a comparison of 0.349 and 0.387), probably due to the much larger prevalence of singles and the reduced average household size. This is still a substantial variation that justifies further checks.

Likewise, the simulation results vary substantially as a function of how equivalized income is computed. For example, simulation S1 which assumes spouses sorted by income in 2013 the way they did in 1973, leads to estimated Gini coefficients ranging from 17.6% using household income to 33.7% of per-capita income. As expected, the variation drops substantially in the larger samples. The contribution to the Gini ranges from 7.9% to 10.7% in the case of all couples (Table 5, Panel B), and from 3.6% to 5.8% in the case of all households (Panel C).

Fertility patterns. Related to the choice of equivalence scale is the fact that over the period under investigation there has been a notable reduction in fertility rates. The relation between fertility and inequality seems to depend on income redistribution within the family based on individual needs and on the presence of economies of scales. Recent studies estimating the cost of children and equivalence scales based on family behavioral models (Dunbar et al., 2013; Mangiavacchi et al., 2018; Betti et al., 2020) confirm the importance of accounting for changes in family structure when analyzing inequality trends.

Figure 6 plots the average fertility rate for all households at different ranks of the male and female income distribution in 1973 and 2013. The left panel highlights that there has been about a 0.5 reduction in the average number of children at all men's income ranks. For women there has also been a large drop in fertility, but it has not been constant along the income distribution. First, the number of women with no income from work is much lower in 2013 (the horizontal line in the right panel is much shorter in 2013 than in 1973, reaching only about 20% of the sample in 2013 versus almost 50% in 1973). Second, the fertility gap is very large, almost 1 child, in the middle of the income distribution. Third, the share of rich women with very low fertility rates roughly doubled in 2013. The association between female income rank and fertility changed substantially over the forty-year period we examine. In 1973 fertility rates for working women went down with income rank at a constant rate, except for a sharp drop above the 90th percentile. In 2013 the income/fertility association appears to be more dichotomic: as women start to work, at about rank 20, there is a sharp drop in fertility. It then remains stable up to about the 80th percentile, where another huge drop happens. These shifts in the relationship between fertility profile and women's income rank suggest that when using equivalized income as a baseline variable there may be a direct effect of fertility changes on income inequality.

To analyze how fertility changes account for the increase in inequality over the forty-year period we make use of predictions based on ordered probit regressions similar to those used for the probabilities outlined in Section 2.3. Table 6 reports simulations of the 2013 the Gini coefficient, top 10% and bottom 50% income shares if fertility patterns were the same as in 1973. In the case of dual-earner couples (Panel A) simulation S7 indicates that the simulated 2013 Gini coefficient would be 20.1% larger than the actual 2013 Gini. The simulated top 10% income share would exceed the actual one by 18.8%, and the simulated bottom 50 percent share would be 21.5% below the actual 2013 share. These are large magnitudes, comparable to the extent to which assortative mating simulations help explain changes in inequality among dual earner couples.

When looking at the other two samples--all couples and all households--similar trends are observed. For all couples the Gini increase would have been 11% larger had fertility been at its 1973 level, and for all households it would have been 12.6% larger. In contrast, associative sorting by income accounted for significantly more of the increase in inequality among dual-earner couples than among more inclusive family types (all couples and couples + singles). In the latter case, the role of fertility changes in accounting for the inequality increase over time is comparable in magnitude to that of the increase in the prevalence of singles, with effects going in opposite directions.

4. Conclusions

This paper's main goals have been to assess the degree to which changes in income inequality over the period 1973 to 2013 can be attributed to changes in spouses' sorting patterns by income, increased labor force participation or a drop in couple formation. To measure spouses' income association a new graphical tool that captures non-linearity was introduced: *rank dependence curves.* This measure of degree of income association is similar to what Chetty et al. (2014) used in an intergenerational mobility context and was designed to vary non-linearly along both husbands' and wives' income ranks. We find that spouses' income association patterns are highly non-linear in both 1973 and 2013, and that these patterns changed substantially over these 40 years. Even though positive and negative income associations co-exist in both years positive rank associations are more common in 2013 than in 1973. In addition, other factors that possibly explain the upsurge in inequality--female labor force participation and prevalence of couple households--have a non-linear relationship with income rank and also changed notably over time.

To estimate how much each of these changes contributed to the rise in income inequality in the U.S. from 1973 to 2013 we compute Monte-Carlo-like simulations capturing all the non-linear relationships that emerged in the first part of the study. We find that an increased tendency towards positive sorting contributed substantially to the rise in inequality among dual-earner couples, but not among all couples or for a sample of couples and singles. Our simulation results suggest that applying the 1973 income association patterns to dual-earner couples generates a Gini coefficient that is about 5% lower than the actual 2013 Gini. This implies that the increase in positive income association accounts for about 25% of the increase in this measure of inequality over the forty-year period. We also examine the role of growth in positive income sorting using alternative measures of inequality: the share of income going to the bottom 50% and to the top 10%. Simulations reveal that for dual-earner couples the rise in spouses' income association contributed more to the shrinking income share of the bottom 50% of the distribution than to the rising income share of the top 10%: positive spouses' sorting by income accounted for 25.8% of the decrease in the income share of the bottom 50% while accounting for only 15.1% of the increase in the income share of the top 10%.

When the same simulation models are applied to a sample that also includes non-working spouses, changes in income association patterns only account for 9% of the increase in inequality measured in terms of Gini coefficient. When singles are included too, changes in the association of husbands' and wives' income barely account for increases in inequality over time, regardless of the measure of inequality. For that sample, couples and singles, the factor that best accounts for inequality is the increased likelihood that individual men and women remain single. The decreased prevalence of couple formation--including marriage and cohabitation— contributed almost 12% of the increase in the Gini coefficient between 1973 and 2013. A decreased tendency to live in couple contributed even more to inequality among the poorest 50%: the increased tendency to remain single observed in 2013 accounts for about 13.1% of the reduction of the income share of the bottom 50% but only 6.6% of the surge in the income share of the top 10% of the income distribution. As for increased female labor force participation it does not help account for the evolution of income inequality as a whole, while it plays a limited role in accounting for the rising share of the top 10%.

A robustness analysis reveals that choice of equivalence scale has substantial effects not only on the measurement of inequality in each year but also on how the factors we considered help account for increased inequality over time based on simulation analysis. For instance, income association could account for between 17.6 to 33.7% of the increase in the Gini coefficient for dual-earner couples, depending on the choice of equivalence scale. A second sensitivity analysis simulates the reduction in fertility levels that has been observed over the 40 years. As fertility reduction was not uniform over women's income distribution, it had an equalizing effect on the evolution of inequality: had fertility rates remained at their 1973 levels the increase in the Gini coefficient throughout 2013 would have been 12.6% larger.

Overall, despite notable changes in the non-linear structure of spouses' income association over time, according to our simulation analysis it accounts for a limited portion of the increased inequality in the period 1973 to 2013. The same applies to changes in labor force participation and in the association between such participation and income rank. In contrast, increased prevalence of singlehood accounts for a relatively large share of the inequality increase, regardless of the measure of inequality. It would be helpful to see studies replicating some of our analyses using different periods or data from other countries, especially countries with good income data and different taxation methods.

Our analysis has been descriptive and our simulations do not imply causality. More evidence needs to be collected before our results based on simulation analysis can be translated into policy recommendations.

Tables and Figures

Table 1: Descriptive statistics

	1 Number of observations	2 Mean income	3 Gini coefficient	4 Top 10% share of income	5 Bottom 50% share of income
Panel A: dual-earner couples					
1973	7,459	46537.28	0.279	0.219	0.309
2013	15,980	63334.92	0.357	0.277	0.263
Percentage variation		36.1%	28.0%	26.5%	-14.9%
Panel B: all couples					
1973	14,372	41557.04	0.319	0.236	0.284
2013	22,761	53869.38	0.424	0.306	0.216
Percentage variation		29.6%	32.9%	29.7%	-23.9%
Panel C: all households					
1973	18,899	38284.39	0.379	0.256	0.241
2013	35,214	45708.18	0.480	0.328	0.175
Percentage variation		19.4%	26.6%	28.1%	-27.4%

The mean equivalent income computed applying the square root equivalence scale to family income as the sum of individual incomes from wage and salary, business and farm income.

Table 2: Gini coefficients and variations for counterfactual simulation models

	1 Gini Coefficient	2 Change ref. 1973	3 Contribution to change (1973-2013)
Panel A: dual-earner couples			
1973 Actual Gini	0.279		
2013 Actual Gini	0.357	27.99%	
Simulated Gini in 2013			
S1: with assortative mating as in 1973	0.340	21.78%	22.18%
Panel B: all couples			
1973 Actual Gini	0.319		
2013 Actual Gini	0.424	33.17%	
Simulated Gini in 2013			
S2: with assortative mating as in 1973	0.415	30.17%	9.03%
S3: with probability of a non-working spouse as in 1973	0.433	35.76%	-7.82%
Panel C: all households			
1973 Actual Gini	0.379		
2013 Actual Gini	0.480	26.66%	
Simulated Gini in 2013			
S4: with assortative mating as in 1973	0.475	25.48%	4.45%
S5: with probability of a non-working spouse as in 1973	0.482	27.23%	-2.12%
S6: with probability of being married as in 1973	0.468	23.54%	11.70%

Notes: (i) Gini coefficients are computed on the equivalized income of the reference population (ii) All Gini coefficient, except the 1973 and 2013, are simulated by Monte-Carlo techniques (see Appendix A).

Table 3: Top 10% income shares and variations for counterfactual simulation models

	1 Top 10% share of income	2 Change ref. 1973	3 Contribution to change (1973-2013)
Panel A: dual-earner couples			
1973 Actual share of income	0.219		
2013 Actual share of income	0.277	26.90%	
Simulated share in 2013			
S1: with assortative mating as in 1973	0.268	22.83%	15.14%
Panel B: all couples			
1973 Actual share of income	0.236		
2013 Actual share of income	0.306	29.48%	
Simulated share in 2013			
S2: with assortative mating as in 1973	0.299	26.71%	9.38%
S3: with probability of a non-working spouse as in 1973	0.312	32.16%	-9.08%
Panel C: all households			
1973 Actual share of income	0.256		
2013 Actual share of income	0.328	27.83%	
Simulated share in 2013			
S4: with assortative mating as in 1973	0.324	26.20%	5.88%
S5: with probability of a non-working spouse as in 1973	0.331	28.96%	-4.03%
S6: with probability of being married as in 1973	0.323	26.01%	6.57%

Notes: (i) Top 10% share of equivalent income is the share of equivalized income produced by the reference population owned by the richest 10% of the same population (ii) All income shares, except the 1973 and 2013, are simulated by Monte-Carlo techniques (see Appendix A).

Table 4: Bottom 50% income shares and variations for counterfactual simulation models

	1 Bottom 50% share of income	2 Change ref. 1973	3 Contribution to change (1973-2013)
Panel A: dual-earner couples			
1973 Actual share of income	0.309		
2013 Actual share of income	0.263	-15.10%	
Simulated share in 2013			
S1: with assortative mating as in 1973	0.275	-11.20%	25.84%
Panel B: all couples			
1973 Actual share of income	0.284		
2013 Actual share of income	0.216	-23.73%	
Simulated share in 2013			
S2: with assortative mating as in 1973	0.222	-21.65%	8.77%
S3: with probability of a non-working spouse as in 1973	0.211	-25.43%	-7.14%
Panel C: all households			
1973 Actual share of income	0.241		
2013 Actual share of income	0.175	-27.35%	
Simulated share in 2013			
S4: with assortative mating as in 1973	0.178	-26.39%	3.53%
S5: with probability of a non-working spouse as in 1973	0.175	-27.61%	-0.92%
S6: with probability of being married as in 1973	0.184	-23.76%	13.13%

Notes: (i) Bottom 50% share of income is the share of total equivalized income produced by the reference population owned by the poorest 50% of the same population (ii) All income shares, except the 1973 and 2013, are simulated by Monte-Carlo techniques (see Appendix A).

Table 5: Gini coefficients - robustness to different equivalence scales

	1	2	3	4
	Gini coefficient		Contribution to change (1973-2013)	
	No scale	Per-capita	No scale	Per-capita
Panel A: dual-earner couples				
1973 Actual Gini	0.249	0.340		
2013 Actual Gini	0.349	0.387		
Simulated Gini in 2013				
S1: with assortative mating as in 1973	0.331	0.371	17.58%	33.66%
Panel B: all couples				
1973 Actual Gini	0.296	0.371		
2013 Actual Gini	0.417	0.450		
Simulated Gini in 2013				
S2: with assortative mating as in 1973	0.408	0.441	7.94%	10.74%
S3: with probability of a non-working spouse as in 1973	0.429	0.454	-9.75%	-5.99%
Panel C: all households				
1973 Actual Gini	0.369	0.444		
2013 Actual Gini	0.493	0.508		
Simulated Gini in 2013				
S4: with assortative mating as in 1973	0.488	0.505	3.56%	5.82%
S5: with probability of a non-working spouse as in 1973	0.494	0.513	-0.93%	-6.72%
S6: with probability of being married as in 1973	0.473	0.498	15.89%	16.69%

Notes: (i) Gini coefficients are computed on the household and per-capita income of the reference population (ii) All Gini coefficient, except the 1973 and 2013, are simulated by Monte-Carlo (see Appendix A).

Table 6: Simulation for fertility changes.

	1	2	3
	Gini coefficient	Top 10% share of income	Bottom 50% share of income
Panel A: dual-earner couples			
1973	0.279	0.219	0.309
2013	0.357	0.277	0.263
Simulated in 2013			
S7: Number of children as in 1973	0.373	0.288	0.252
Contribution to change (1973-2013)	-20.14%	-18.76%	-21.52%
Panel B: all couples			
1973	0.319	0.236	0.284
2013	0.424	0.306	0.216
Simulated in 2013			
S8: Number of children as in 1973	0.436	0.316	0.210
Contribution to change (1973-2013)	-10.97%	-14.81%	-9.48%
Panel C: all households			
1973	0.379	0.256	0.241
2013	0.480	0.328	0.175
Simulated in 2013			
S9: Number of children as in 1973	0.493	0.338	0.168
Contribution to change (1973-2013)	-12.61%	-14.46%	-11.08%

Notes: (i) Gini coefficients and income shares are computed on the equivalized income of the reference population (ii) All Gini coefficients and income shares, except the 1973 and 2013, are simulated by Monte-Carlo techniques (see Appendix A).

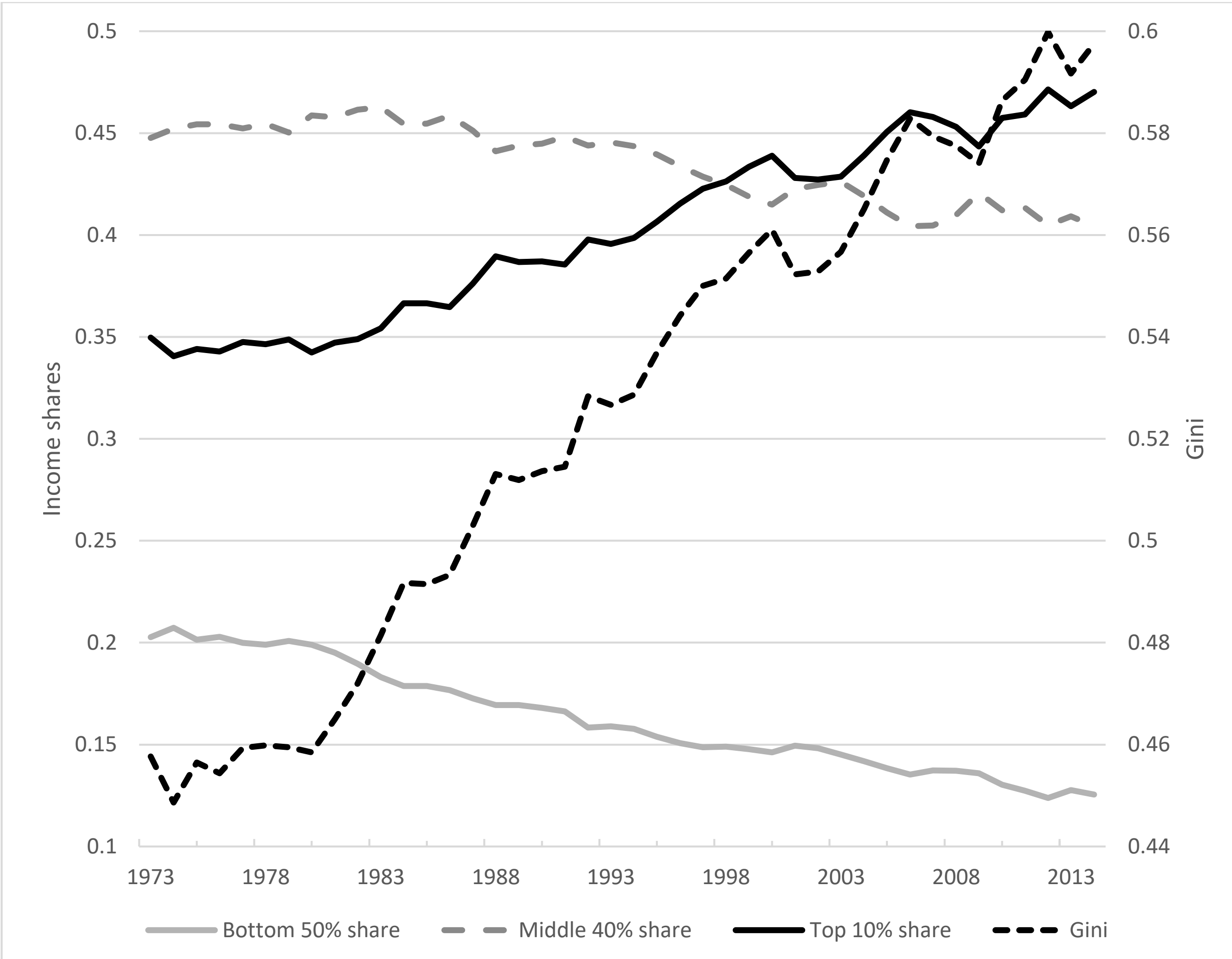

Figure 1: Evolution of the Gini coefficient and income shares in the U.S. 1973-2013. Source: World Inequality Database.

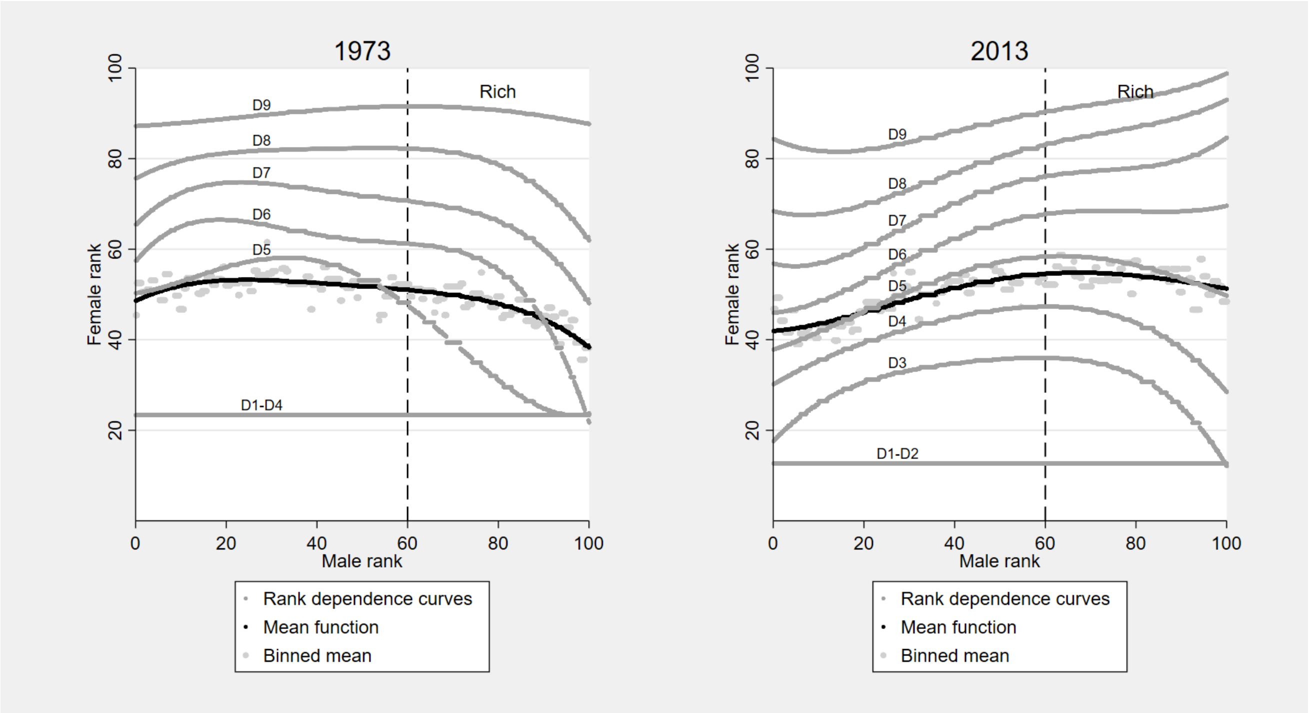

Figure 2: Rank dependence analysis for spouses' income association by male rank: binned means, mean function, and rank dependence curves.

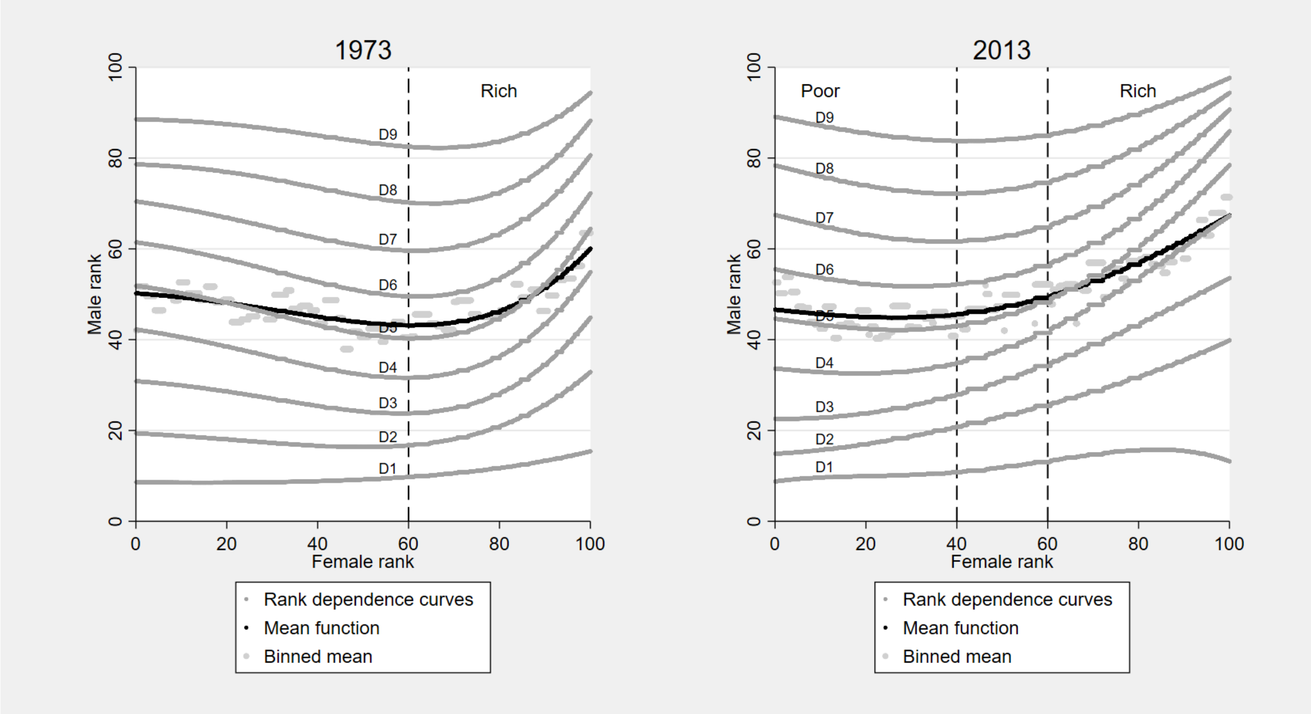

Figure 3: Rank dependence analysis for spouses' income association by female rank: binned means, mean function, and rank dependence curves.

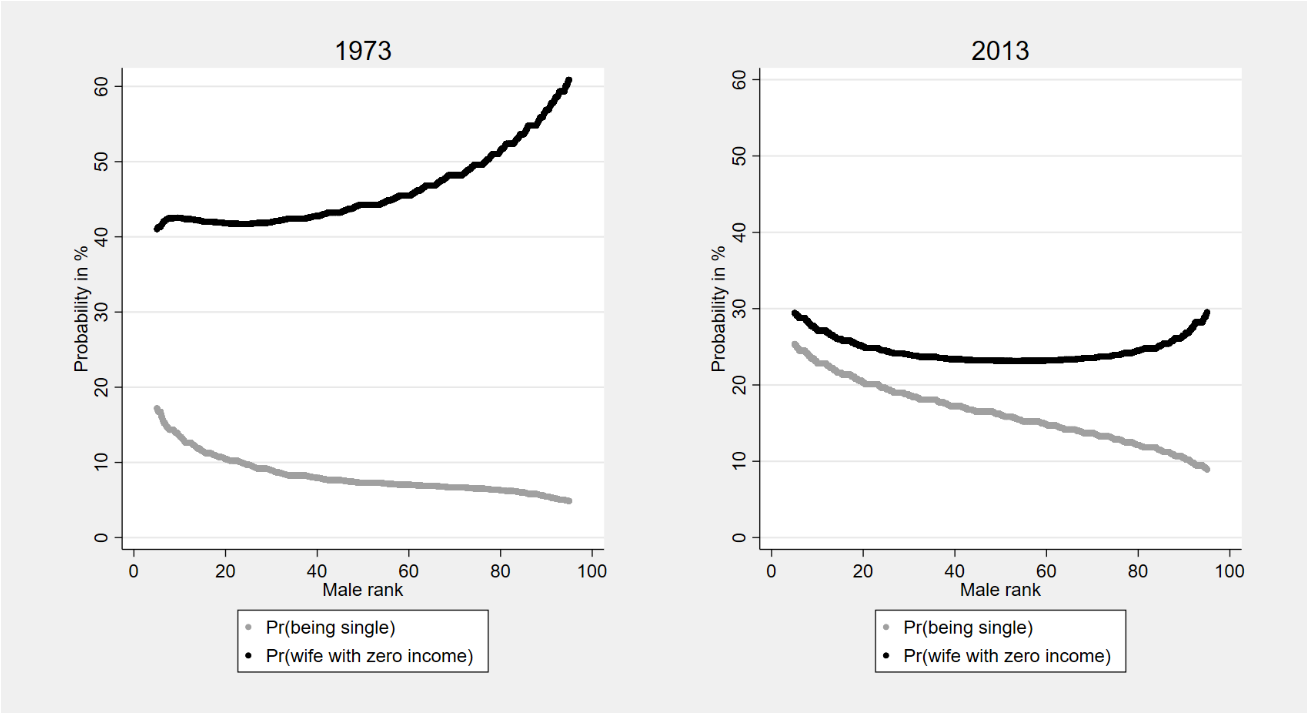

Figure 4: Male probability of being single and of having a wife with zero income

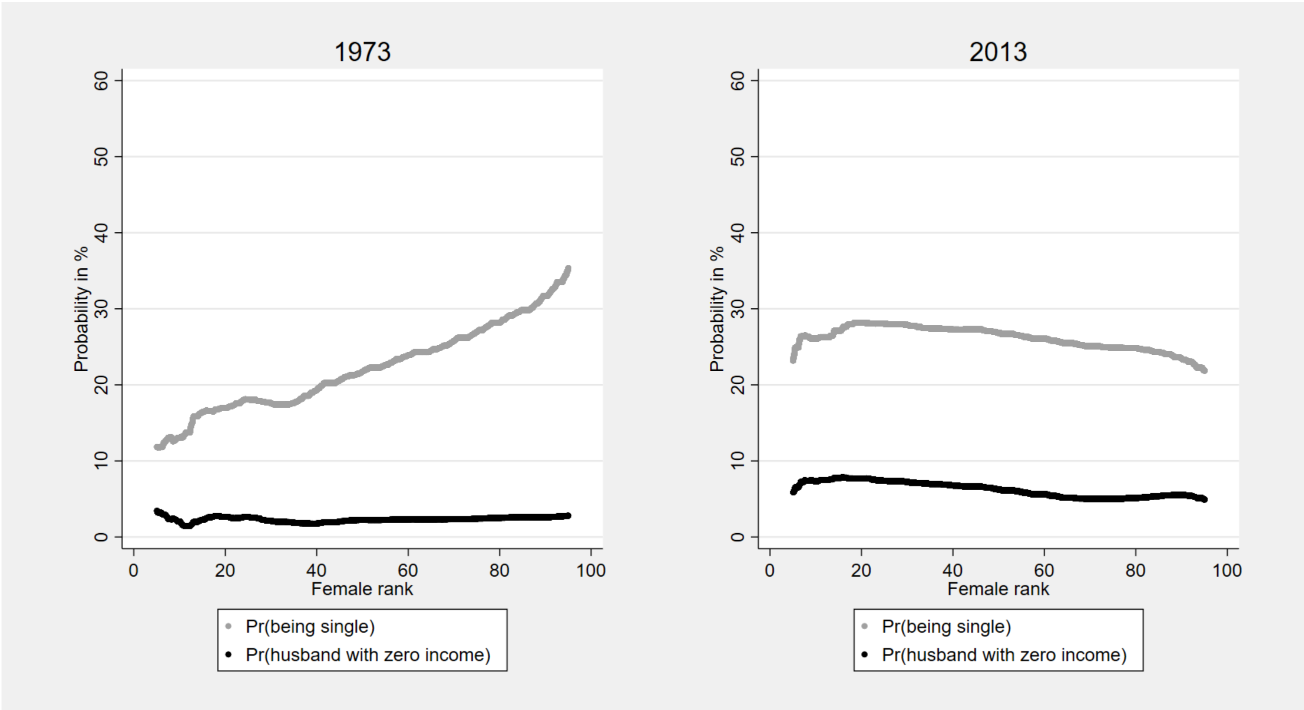

Figure 5: Female probability of being single and of having an husband with zero income

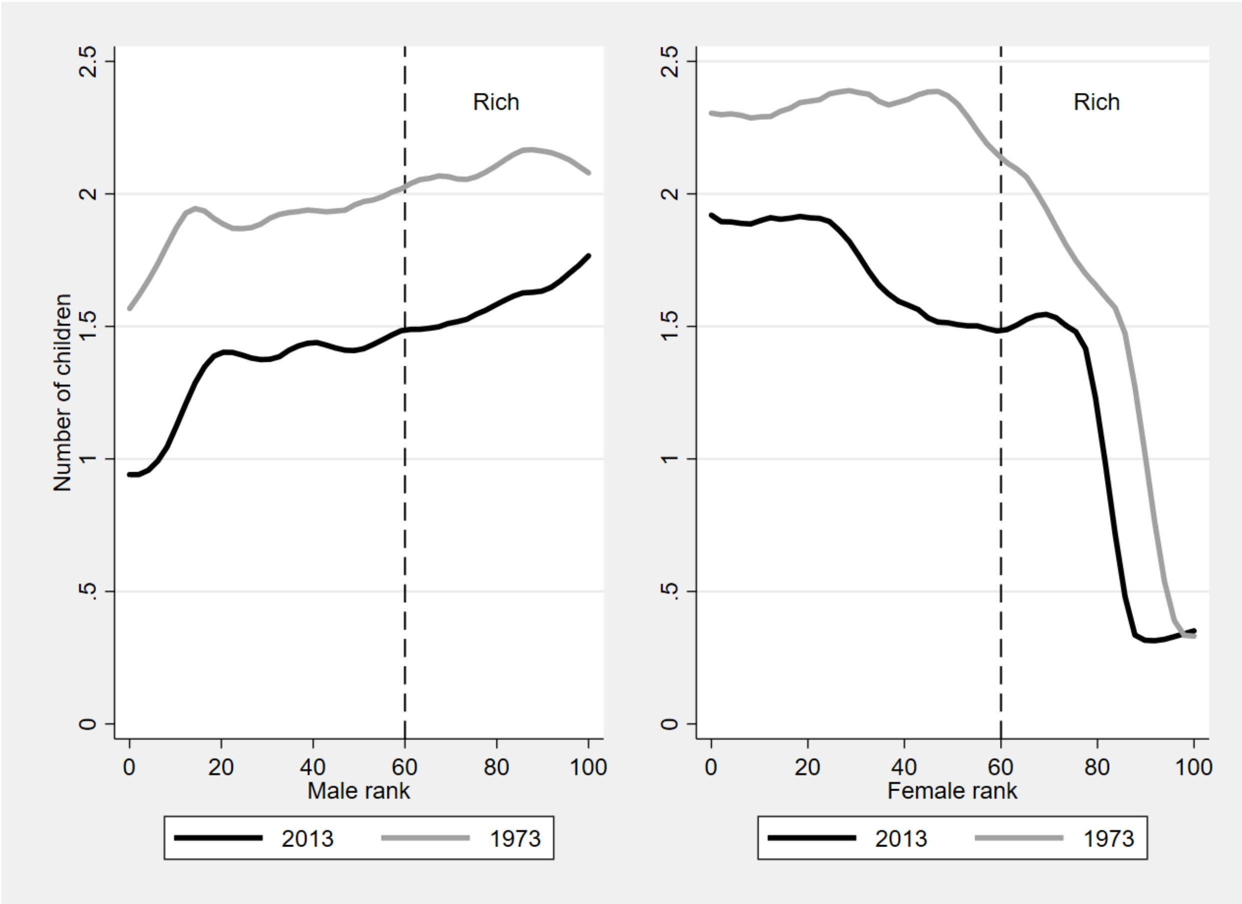

Figure 6: Fertility rates across male and female income ranks.

References

Aaberge, R. (1997). Interpretation of changes in rank-dependent measures of inequality. *Economics Letters*, 55, 215–219.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 81(4), 813–846.

Becker, G. S. (1974). A theory of marriage: Part II. *Journal of Political Economy*, 82(2, Part 2), S11–S26.

Betti, G., L. Mangiavacchi, and L. Piccoli (2020). Women and poverty: Insights from individual consumption in Albania. *Review of Economics of the Household*, forthcoming, March.

Bratberg, E., J. Davis, B. Mazumder, M. Nybom, D. D. Schnitzlein, and K. Vaage (2017). A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US. *The Scandinavian Journal of Economics*, 119(1), 72–101.

Bredemeier, C. and F. Juessen (2013). Assortative mating and female labor supply. *Journal of Labor Economics*, 31(3), 603–631.

Burkhauser, R. V., J. Larrimore, K. I. Simon, et al. (2012). A" Second Opinion" on the Economic Health of the American Middle Class. *National Tax Journal*, 65(1), 7–32.

Burtless, G. (1999). Effects of growing wage disparities and changing family composition on the US income distribution. *European Economic Review*, 43(4), 853–865.

Cameron, A.C., and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press

Cancian, M. and D. Reed (1998). Assessing the effects of wives' earnings on family income inequality. *The Review of Economics and Statistics*, 80(1), 73–79.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, 129(4), 1553–1623.

Dahl, M. W. and T. DeLeire (2008). The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data. Institute for Research on Poverty, Discussion Paper No. 1342-08, University of Wisconsin–Madison.

DiNardo, J., N. M. Fortin and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5), 1001-1044.

Dunbar, G., A. Lewbel, and K. Pendakur (2013). Children's resources in collective households: Identification, estimation, and an application to child poverty in Malawi. *The American Economic Review*, 103(1), 438–71.

Dupuy, A. and S. Weber (2019). Marital patterns and income inequality. Available at SSRN: https://ssrn.com/abstract=3156484 or http://dx.doi.org/10.2139/ssrn.3156484.

Eika, L., M. Mogstad, and B. Zafar (2019). Educational assortative mating and household income inequality. *Journal of Political Economy,* 127(6), forthcoming.

Fan, J., and I. Gijbels. 1996. Local Polynomial Modelling and Its Applications. London: Chapman & Hall.

Fitch C. A., and S. Ruggles (2005). The Rise of Cohabitation in the United States: New Historical Estimates. Minnesota Population Center, University of Minnesota. Prepared for the annual meeting of the Population Association of America, Philadelphia, March 31-April 2, 2005.

Flood, S., M. King, S. Ruggles, and J. R. Warren (2017). Integrated Public Use Microdata Series, Current Population Survey: Version 5.0. Technical report, Minneapolis: University of Minnesota.

Gonalons-Pons, P. and C. R. Schwartz (2017). Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage? *Demography*, 54(3), 985–1005.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2014). Marry your like: Assortative mating and income inequality. *The American Economic Review*, 104(5), 348–353.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics*, 8(1), 1–41.

Grossbard, S. (2015). *The marriage motive: A price theory of marriage*. Springer.

Grossbard, S. and V. Vernon (2015). Common Law Marriage and male/female convergence in labor supply and time use *Research in Labor Economics*, *Volume on Gender Convergence in the Labor Market*, 41: 143-175.

Grossbard-Shechtman, A. (1984). A theory of allocation of time in markets for labour and marriage. *The Economic Journal*, 94(376), 863–882.

Hersch, J. (2013). Opting out among women with elite education. *Review of Economics of the Household*, 11(4), 469–506.

Hryshko, D., C. Juhn, and K. McCue (2017). Trends in earnings inequality and earnings instability among U.S. couples: How important is assortative matching? *Labour Economics*, 48(Supplement C), 168–182.

Hyslop, D. R. (2001). Rising US earnings inequality and family labor supply: The covariance structure of intrafamily earnings. *American Economic Review*, 91(4), 755–777.

Larrimore, J. (2014) Accounting for United States household income inequality: The changing importance of household structure and male and female labor earnings inequality. *The Review of Income and Wealth*, 60(4), 683–701.

Lewbel, A., Pendakur K. (2008) Equivalence Scales. In: Palgrave Macmillan (eds) *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, London.

Mangiavacchi, L., F. Perali, and L. Piccoli (2018). Intrahousehold Distribution in Migrant-sending Families. *Journal of Demographic Economics*, 84, 107–148.

McLanahan, S. (2004). Diverging destinies: How children are faring under the second demographic transition. *Demography*, 41(4), 607–627.

Pestel, N. (2017). Marital Sorting, Inequality and the Role of Female Labour Supply: Evidence from East and West Germany. *Economica*, 84(333), 104–127.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1), 1–41.

Saez, E. and G. Zucman (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *The Quarterly Journal of Economics*, 131(2), 519–578.

Schwartz, C. R. (2010). Earnings inequality and the changing association between spouses' earnings. *American Journal of Sociology*, 115(5), 1524–1557.

Schwartz, C. R. and R. D. Mare (2005). Trends in educational assortative marriage from 1940 to 2003. *Demography*, 42(4), 621–646.

Western, B., D. Bloome, and C. Percheski (2008). Inequality among American families with children, 1975 to 2005. *American Sociological Review*, 73(6), 903–920.

Appendix A: Detailed Steps for Simulations

In the following list we detail steps of the simulations generally described in Section 2.4.

1. Choose the sample for the income year 2013, which can be found in the survey from 2014. The sample includes males with female spouse, female with male spouse, single male and single female. The income variable is recoded to obtain a variable for male income and another for income of a female spouse if she is present. If she is not present the income is zero. In the same way, a single female has the variable male spousal earnings coded to zero because no male spouse is present.

2. Separate the sample into two parts; one for male singles and male with a present spouse and another including females that are singles.

3. Use the male sample to estimate the probability to have a spouse. This is done with a local linear nonparametric regression with the command *npregress* in Stata v15. The logarithm of male income is used as the only explanatory variable. The model is used to obtain the predicted probability of having a spouse *at each rank* of the earnings distributions of the males. While we use a nonparametric regression technique a Probit model with a polynomial in log-income would produce similar results.

4. Calculate the proportion of males with zero earnings that has a spouse. This is the predicted probability for males with zero earnings (that was not included in step 3).

5. Use the male sample to estimate the probability to have a spouse with zero income, given that a spouse was present. Again a nonparametric regression is specified with log-income as the only explanatory variable. The model is used to obtain the predicted probability of having a spouse with zero earnings *at each rank* of the earnings distribution of the males, given that a spouse was present.

6. For males with zero earnings, calculate the proportion with a spouse with zero income given that a spouse was present. This is the predicted probability for males with zero earnings (that was not included in step 5).

7. Repeat step 1-6 for the income year 1973.

8. Choose sample that was made in 2) that includes single males and males with spouses and, within this group, make a random selection to obtain a sample-size equal to what is found for 1973 (for single males and males with spouses). In Table 7 we show the one-to-one match in a table where an illustrative example is included for observations $1, \dots, n$.

9. Choose a random sample from the sample of female singles. The size of the samples is selected to replicate the proportion of female singles that was found in 2013. This sample is added to the sample

obtained in 8; in Table 7, k is selected so that the proportion of single female remains the same as in 2013.

10. Mimic the situation in 2013 by using the probabilities found in (3-4) and (5-6) to recode the spousal earnings to be 0. That is, for each individual the spousal income is coded to zero if the predicted probability of not having a spouse exceeds a random uniform draw. The same is done for the probability to have spouse with zero earnings. In the table we include the mimic according to: $inc(Sp)_1^{mimic} = inc(Sp)_1(r = r_1^{13}) \times 1(P_a > p(Sp)_1^{13}) \times 1(P_b > p(0)_1^{13})$. $1(P_{\cdot} > p(\cdot)_1^{13})$ is equal to one when the inequality is true, i.e. the random draw is higher than the probability, and zero otherwise.

    With this procedure it is possible that the random draws indicate positive earnings but the actual situation was to not have a present spouse, or to have a spouse with zero wage. These cases must be replaced with imputed earnings. $inc(Sp)_1(r = r_1^{13})$ is the spousal income that is found in the rank r_1^{13}, i.e. the actual spousal income, or the income that corresponds to the rank for the imposed value.

11. To perform the imputation all men are sorted into 100 groups according to their position in the earning distribution. For each male a spousal wage is randomly chosen from the working spouses in his group. Notice that the imputation is only done for the cases where the probabilities indicate having a spouse with positive earnings, but the actual case was being single or having a spouse with zero earnings. $r_{4,*}^{73}$ is included in the example to show that some ranks are due to imputed income.

12. Calculate the Gini coefficient for simulated counterfactual situations using the 'sgini' Stata custom program[1] . In Table 7 we include income association as in 1973. In our example; $inc(Sp)_1^{cf} = inc(Sp)_1(r = r_1^{73}) \times 1(P_a > p(Sp)_1^{13}) \times 1(P_b > p(0)_1^{13})$

    In $inc(Sp)_1^{cf}$ we have only counterfactual assortative mating as in 1973, but we can, for example, easily change $p(Sp)_1^{13}$ for $p(Sp)_1^{73}$ to use the counterfactual situation with the probability to have a spouse as in 1973. Notice that if the counterfactual situation includes $p(Sp)_1^{73}$, l is used instead of k so that the proportion of single female is keep the same as in 1973.

[1] Van Kerm, P. (2009) "sgini – Generalized Gini and Concentration coefficients (with factor decomposition) in Stata," CEPS/INSTEAD, Differdange, Luxembourg.

13. Replicate steps 8-15 200 times. The total income is divided by the square root of number of family members and the Gini coefficient is calculated on the corresponding equivalized income. The average from the 200 replications is finally used.

Table 7: Example

1973				2013					Mimic	Counterfactual Assortative 1973
Male rank	Sp rank	P(Sp)	P(zero)	Male income	Male rank	Sp rank	P(Sp)	P(zero)	Sp income	Sp income
1	r_1^{73}	$p(Sp)_1^{73}$	$p(0)_1^{73}$	inc_1	1	r_1^{13}	$p(Sp)_1^{13}$	$p(0)_1^{13}$	$inc(Sp)_1^{mimic}$	$inc(Sp)_1^{cf}$
2	r_2^{73}	$p(Sp)_2^{73}$	$p(0)_2^{73}$	inc_2	2	r_2^{13}	$p(Sp)_2^{13}$	$p(0)_2^{13}$	inc_2^{mimic}	inc_2^{cf}
3	r_3^{73}	$p(Sp)_3^{73}$	$p(0)_3^{73}$	inc_3	3	$r_{3,*}^{13}$	$p(Sp)_3^{13}$	$p(0)_3^{13}$	inc_3^{mimic}	inc_3^{cf}
4	$r_{4,*}^{73}$	$p(Sp)_4^{73}$	$p(0)_4^{73}$	inc_4	4	r_4^{13}	$p(Sp)_4^{13}$	$p(0)_4^{13}$	inc_4^{mimic}	inc_4^{cf}
…	…	…	…		…	…	…	…	…	…
n	r_n^{73}	$p(Sp)_n^{73}$	$p(0)_n^{73}$	inc_n	n	r_n^{13}	$p(Sp)_n^{13}$	$p(0)_n^{13}$	inc_n^{mimic}	inc_n^{cf}
				0					$inc(Single)_1$	$inc(Single)_1$
				0					$inc(Single)_2$	$inc(Single)_2$
				…					…	…
				0					$inc(Single)_k$	$inc(Single)_l$