Watching in your partner's pocket before saying "Yes!" Assortative mating and income inequality*

Carlo V. Fiorio† and Stefano Verzillo‡

February 2, 2017

Abstract

Standard measures of assortative mating by income levels of spouses at the year of wedding are flawed by endogeneity issues. By using a unique administrative data on a large region in Italy we provide the first measure of assortative mating on labor income levels. The administrative nature of our data (tax records) and the modeling choice based on percentile groups reduce measurement error to the minimum. The availability of labor income for both spouses up to three years before wedding is used to account for the simultaneity bias.

Results provide evidence that top income women are much more likely to get married to top income men. In particular, 13% of women belonging to the top 1% of their income distribution get married to a man belonging to the top 1% of grooms' income distribution.

Counterfactual analysis on the effect of assortative mating by income levels on income inequality suggests that if love was the unique driver of marriage and falling in love was a randomly allocated event, the distribution of family income could largely vary, even reducing the Gini index by half.

JEL codes: J12,J41
Keywords: assortative mating, income, administrative data

*The information and views set out in this paper are those of the authors and do not reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

†University of Milan and Irvapp-FBK. Address: Department of Economics, Management and Quantitative Methods. Via Conservatorio, 7. 20122 Milan. email: carlo.fiorio@unimi.it

‡Joint Research Centre, European Commission and CRISP - Interuniversity Research Centre on Public Services, University of Milan-Bicocca

1 Introduction

The theoretical literature of assortative mating in the marriage market developed extensively since the seminar paper by Becker (1973), predicting perfect assortativeness (for a comprehensive survey, see Chiappori and Salanié, 2016). However, the empirical literature documented strong homogamy mostly looking at educational achievements of spouses, which typically complete before wedding (e.g., see Schwartz and Mare, 2005; Chiappori, Oreffice, and Quintana-Domeque, 2012; Greenwood, Guner, Kocharkov, and Santos, 2014). The measure of assortative mating by level of incomes is a daunting task for several reasons.

Individual income contribution to the couple income is either measured in surveys or unavailable in administrative tax records in countries where spouses fill in their tax forms jointly. Survey incomes are however often flawed with measurement error, which tends to increase for high levels of income given the long top-tail of typical income distributions (Burkhauser, Feng, Jenkins, and Larrimore (2012)). Moreover, assortative mating measured at the wedding year is likely to be biased due to simultaneity. In fact, positive assortative mating observed at the wedding year might be due to the fact that spouses are similar because of marriage and not that they marry because they have similar income. The simultaneity bias cannot be estimated with standard panel data, such as the PSID, because of the way in which panels are constructed, following individuals in a married couple only as spouses join the sample after marrying someone who was previously in the panel.

In this paper we use a unique data set collecting the whole set of tax forms for the population of residents in Lombardy, a major region in Italy, with a population of 10 million people, over the period 2007-2011. This allows us to identify a set of over $434,000$ of newly married couples and to measure the size of assortative mating by labor income, dealing with measurement error and simultaneity biases.

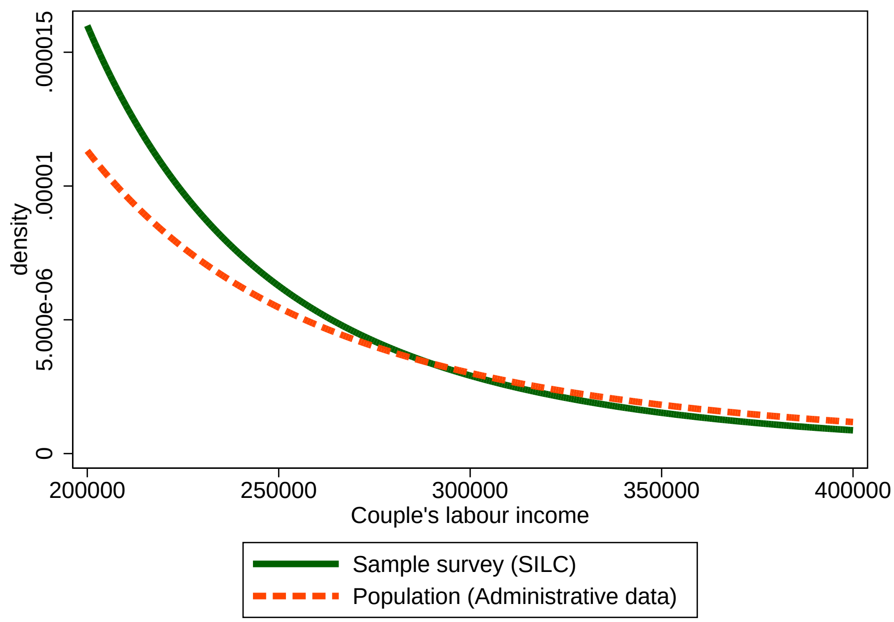
The use of administrative data is particularly interesting for the analysis of assortative mating as patterns of assortative mating could differ for different levels of income and survey data are often affected by top coding and under-representation of top incomes. For assessing the severity of the under-representation of top incomes, we selected all couples with labor income above €200,000 residing in Lombardy over the period 2008-2011 from our administrative sample and from the EU Statistics on Income and Living Conditions (EU-SILC), which is the largest representative sample of Italian households produced by Eurostat. Hence, we estimate the average Pareto (Type I) parametric distribution using both dataset. Figure 1 shows that the top tail using the population dataset decades at a much slower rate using the survey data confirming that survey data would provide a biased picture of assortative mating at the top.

The structure of the paper is as follows. Section 2 will describe in more details the data and present some descriptive statistics of the dataset we use. In Section 3 we will introduce the assortative mating ratio used to describe assortative mating by income groups and represent it in an effective three dimensional picture. In Section 4 we provide an estimate of an empirical model of assortative mating on income and deal with endogeneity issues due to simultaneity. In Section 5we assess the role of assortative on income inequality and finally in Section 6 we conclude and discuss results.

2 The data

Administrative data generally provides better estimate of top-income shares due to the fact that families at the very top of the income distribution are usually under-represented by most of the household surveys. In fact, the existence of

Figure 1: Average Pareto (Type I) parametric distribution of couple's labour income above 200.000 euro, over the period 2007-2011.



thin tails at the top leads to large sampling variability which may cause disproportional participation rates of top-income families and consequently may produce downward biased estimates. For this reason tax income records are considered by the literature the best sources to study inequality issues at the top of the families' income distribution.

In this paper we explore the full administrative dataset of tax forms of one of the largest Italian regions over the period 2007-2011. It is the richest region of Italy, counting about 10 million residents (as large as Portugal and twice the size of Denmark). Our dataset is based on confidential administrative data[1] and includes information on approximately 45 millions of tax records referred to 9 millions of individuals for each available year, of whom 6.7 millions of taxpayers and 2.1 millions of dependent children and spouses.

Tax records allow the exact identification of tax units, comprising both spouses and fiscally dependent children and other relatives. All married taxpayers are obliged to declare the tax code of their spouse even if the latter is not fiscally dependent (i.e. she/he is earning a yearly gross income larger than € $2,840$, which is the maximum level of income allowed for being exempted to fill in a tax record). Unfortunately, data does not allow an identification of households (i.e. the group of people sharing the same dwelling and part of costs) if they are made by more than one tax unit, as the exact address of residence is not known to us.

The representative husband of a new-couple in the observed period (2008-2011) is 47 years old with a 88% of chances to earn around $20,000$ euros (gross) yearly as an employee and a 7% of the chances to earn around 800 euros (gross) yearly as self-employee (see Table 2 and 3).

[1] Data are analyzed in collaboration with CRISP - Inter-university Research Centre on Public Services at the University of Milan-Bicocca - under the framework of a preliminary research program with the Tax and Income Department of the Lombardy Region. Tax records were recorded and treated after being made anonymous using an irreversible hashing algorithm.

Table 1: The formation of new (tax) couples in Lombardy (2008-2011)

	Year				
	2008	2009	2010	2011	Total
Existing couples	2,250,757	2,232,540	2,230,129	2,186,882	8,900,308
New couples	113,907	113,495	103,430	103,375	434,207
Total	2,364,664	2,346,035	2,333,559	2,290,257	9,334,515

Table 2: Descriptive Statistics of the Spouses by Income Source (2008-2011)

Year	2008	2009	2010	2011
Females				
People Employed	62.16%	59.14%	61.73%	60.43%
People Self-Employed	2.32%	2.23%	2.45%	2.24%
N. Obs	113,907	113,495	103,430	103,375
Males				
People Employed	88.64%	88.41%	86.98%	87.54%
People Self-Employed	7.11%	6.48%	7.02%	5.91%
N. Obs	113,907	113,495	103,430	103,375

Table 3: Descriptive Statistics of the Spouses (2008-2011)

Year	2008	2009	2010	2011
Females				
Age	44.31	44.61	44.30	43.71
	(16.61)	(16.76)	(16.30)	(15.69
N. children	0.61	0.58	0.57	0.63
	(0.98)	(0.91)	(0.88)	(0.92
Town of residence size	168,389	167,478	160,530	160,942
	(394,955)	(393,826)	(385,856)	(386,175
Labor Income	9,366.62	9012.76	9,633.32	9,444.16
	(13,285.05)	(13,951.49)	(12,222.71)	(14,227.63
Self-Employed Income	302.36	282.95	314.96	256.79
	(4,668.30)	(4,996.76)	(4,330.32)	(3,705.83
N. Obs	113,907	113,495	103,430	103,375
Males				
Age	47.37	48.33	48.27	47.47
	(16.53)	(17.06)	(16.79)	(16.06
N. children	0.61	0.58	0.57	0.63
	(0.98)	(0.91)	(0.88)	(0.92
Municipality size	154,890	156,868	153,815	150,535
	(377,802)	(380,790)	(377,238)	(372,687
Labor Income	21,004.47	20,086.90	20,674.33	20,936.84
	(39,682.06)	(34,624.05)	(41,971.63)	(71,515.44
Self-Employed Income	800.86	730.30	896.56	703.36
	(14,341.27)	(9,664.97)	(12,733.38)	(9,642.63
N. Obs	113,907	113,495	103,430	103,375

Notes: Standard deviations in parentheses.

The representative husband marries a wife who is on average 3 years younger (44), working as employee in the 61% of the cases with an average gross annual income close to $9,000$ euros and/plus a self-employee gross income of around 350 euros (in the 2.5% of the cases). The overall average number of dependent children for these new-couples in the $2008-2011$ time span is 1 (see Tables 2 and 3). In addition, the husband's income related variables (labor, employee and self-employed incomes) show larger variabilities when compared to the wives ones. However, the multivariate profile of the new spouses seems to be persistent over time without any notable difference in the observable characteristics occurred during the observed time span (even if during the economic crisis).

3 The Excessive Mating Ratio

If the determinants assortative mating were uncorrelated with income one would expect that the likelihood of a rich man getting married to a poor woman would be the same as that getting married to a rich one. Although there is a large evidence showing that assortative mating depending on educational level is strong and increasing over time (Schwartz and Mare, 2005; Greenwood, Guner, Kocharkov, and Santos, 2014), and income is highly correlated with past education choices, there is fewer evidence on assortative mating on income.

The Italian tax legislation requires that an individual with a yearly gross income above € $2,840$ fills in an individual tax form. It also requires that taxpayers include the tax code of all fiscally dependent relatives (i.e. spouse, children other relatives who earn less than € $2,840$ on a yearly basis) as well as of the spouse whatever is her/his income. This feature of the Italian tax code allows us to identify couples even if one of the two spouses does not earn any income, as long as the other fills in a tax form. We identify year t as the wedding year for couple i if from time t onward the spouses fill in independent tax forms declaring to be married to the other or one spouse declare he is married to the other, who is fiscally dependent, and both spouses fill in a separate tax form (if fiscally independent) or appeared as dependent relative in the tax form of someone else in our full population of tax forms.

Once all new couples formed over the period 2008-2011 have been identified, we keep only those who got married at some point in the period 2008-2011, which we call the population of brides and grooms over the period in Lombardy. In other words, we drop from the analysis all records of residents who remained single or got married before 2008, obtaining a sample of $434,207$ new couples over the period.

Hence, out of the population of brides and grooms, we ranked individual labor income by gender and year, and divided them into 100 percentile groups (obtained by the 99 numbered points that divide the ordered set of income into 100 parts each of which contains one-hundredth of the total) and build a matrix for year t of size 100×100, where cell $c^t_{k,j}$ is the group of couples with a husband at the husband-income percentile k and the wife at wives-income percentile j, where $\{k, j = 1, ..., 100\}$. For instance, cell $c_{100,100}$ contains spouses that belonged to the top 1% of their gender-specific income distribution, cell $c_{100,50}$ contains spouses with a husband in the top 1% of grooms' income distribution and a bride whose income was between the 49-th percentile and the median.

Clearly, the probability under perfectly random mating of a husband of percentile k to get married to a wife of percentile j at year t is $1/(100 \times 100)$. The observed probability instead is the count of couples in cell $c^t_{k,j}$ over the total number of newly-formed couples, $\sum_{k,j} c^t_{k,j}$. This allows us to compute the Excessive Mating Ratio (EMR) at year t as the ratio of the observed mating probability over the theoretical probability of mating under the assumption of random mating:

$$EMR_{kj}^{t} = \frac{\text{Actual frequency of couples in cell } c_{kj}^{t}}{\text{Theoretical frequency under random mating}} = \frac{c_{kj}^{t} / \sum_{k,j} c_{k,j}^{t}}{1/(100 \times 100)}.$$

If the $EMR_{kj}^{t} = 1$ for all k, j, where $k, j = 1, 2, ..., 100$ there would be no assortative mating on income at year t and people get married for some other reason, e.g. beauty, love, randomness, which are uncorrelated to income. When the $EMR_{kj}^{t} > 1$ it means that the observed frequency of couples c_{kj} exceeds the theoretical probability of observing it. When, instead, $EMR_{kj}^{t} < 1$ it means that the observed frequency of couples c_{kj} is exceeded by the theoretical probability of observing it.

Perfect assortative mating by income groups at year t could be analyzed by looking at the EMR_{kj}^{t} for $k = j$, but the whole picture might in fact be of interest.

Figure 2 plots on a 3-D graph the average-over time excessive mating ratio, $EMR_{kj} = \sum_t EMR_{kj}^{t}/4$, for our population, where only income percentiles of both spouses above the median were considered. We pooled together all years also considering that the variation over time is very limited and considered only individual labor income above the median to be sure that $c_{kj} > 0$ for all k, j.

Figure 2 (panel - a) shows that a marriage among people in very distant percentiles (e.g. a man with median income and a woman in the top percentiles, and viceversa) is relatively unlikely to occur, whereas the homogamy by labor income percentiles is relatively more frequent. What appears as a striking pattern is the increase of homogamy by labor income as the level of income increases. In particular, men above the 90^{th} percentile are more likely to get married to a woman with income at or above the 90^{th} percentile and increasingly so. In particular, men belonging to the top 1% are about 12 times more likely to get married to a woman with income in top 1% percentile of women's distribution than if they selected a wife randomly.

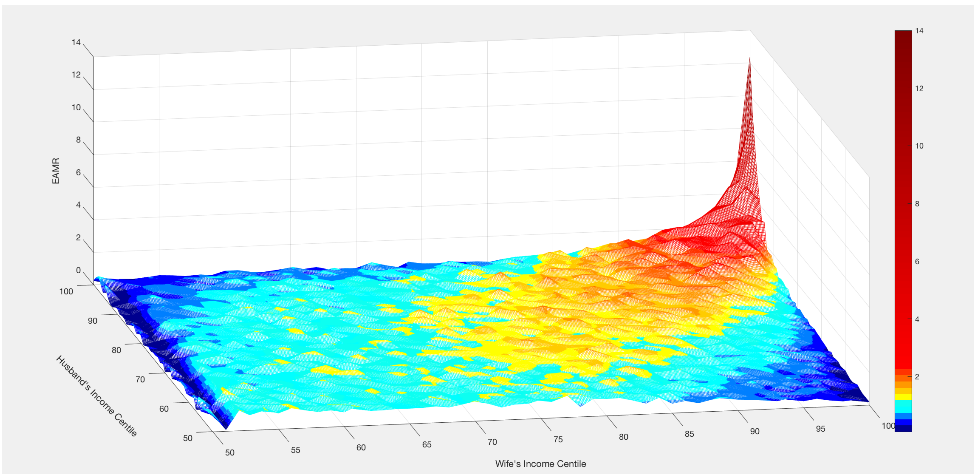
The EMR plotted in Figure 2 (panel - a) considers labor income only. This is due to the fact that Italian tax forms record labor income of all taxpayers, whereas it does record capital income only for a tiny fraction of taxpayers, as capital income taxation is more frequently fully paid at source, and real estate and building property income is very noisily reported, as it partly measure imputed income and partly effective income, and more likely it would be a better measure of socio-economic background than a measure of skill or attractiveness in the marriage market.

However, someone may argue that declared labour income might be different from true labour income as tax evasion rates in Italy is large as compared to other developed countries. If tax evasion was unevenly distributed along the income range, this might affect the assortative mating picture. Although we cannot have a reliable measure of tax evasion along income levels, we can compute the EMR as above disregarding self-employment income, which is self-reported, and consider only employment income, which is third-party reported. The picture of assortative mating by employment income only is plotted in 2 (panel - b) and it is qualitatively very similar to the one above, suggesting that the assortative mating picture would not be affected by a tax evasion bias.

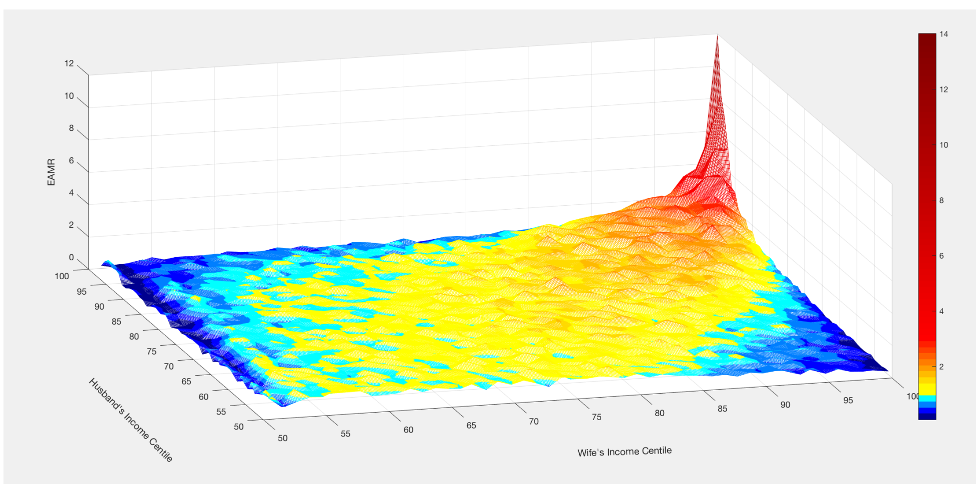
In the reminder of the paper we will analyse assortative mating on labor income regardless of its source and leave the analysis of assortative mating on employment income only to the robustness section.

Figure 2: Excessive Mating Ratio by Income at the Marriage Year

(a) Assortative mating by labor income (employment and self-employment)



(b) Assortative mating by employment income



4 The estimation of assortative mating on income

Let us denote wife's and husband's attractiveness in the marriage respectively with y_w^* and y_h^*, and assume that attractiveness is not observable. We only observe each spouse's market income, y_s, for $s = w, h$, where $y_s = y_s^* + \epsilon_s$ and ϵ_s, is an error term that incorporates a measurement error or an omitted variable in the index of attractiveness.

Now, compute the q-th income percentiles of the husbands income distribution, as $q_h := Pr[y_h < q_h] \leq q/100$ and similarly for the wives' income distribution, q_w, where $\{q = 1, ..., 100\}$. Hence, we denote with c_k^h the percentile group of husbands whose income is between the k-th and the $(k-1)$-th percentile and with I_k^h the binary indicator that takes a value equal to one if the husband belongs to group c_k^h, (i.e. $y_h \in c_k^h$) and $I_k^h = 0$ otherwise. Similarly, the binary variable I_j^w for wives.

Hence for the population of N newly formed couples, we provide an estimate of the model:

$$I_{k,i}^w = X_i^{w,h'} \beta_k + I_i^{h'} \gamma_k + u_{k,i}, \tag{1}$$

where $i = \{1, ..., N\}$, $k = \{1, ..., 100\}$, X is a matrix of both husband and wife observable characteristics including age, number of children and size of the town of residence and a year fixed effect, and I_h is the vector indicating to which quantile group w's husband belong to, i.e. $I_h' = [I_1^h, ..., I_{100}^h]$, β and γ are vector of coefficients to be estimated and $u_{k,i}$ is the residual. Model (1) is a linear probability model and we estimate it by ordinary least squares (OLS) for all wives' quantile groups, $k = \{1, ..., 100\}$.

In Figure 3 we plot the fifth-order polynomial smoothed line connecting the estimated values of the γ_k for a subset of quantiles, namely for $k = 75$, 95, 99, 100. The black solid line shows that, even controlling for a set of observable characteristics, the probability of a woman belonging to the top 1% group of the distribution of brides to get married to a man belonging to the top 1% group of the distribution of grooms is well above 10% and the probability of getting married to a man belonging to a lower percentile group quickly decreases as we move on a percentile group for lower income grooms. Table 4 reports the point estimate $\hat{\gamma}_k$ for a subset of k values, showing that the estimated value of perfect assortative mating for the top 1% groups is $\hat{\gamma}_{100} = 0.16$. Perfect assortative mating if the bride belongs to the 99^{th} quantile group (I^w_{100}) is lower, equal to 0.064, and if the bride belongs to, for instance, the 75^{th} quantile group (I^w_{75}) there is only 1.1% larger probability that she get married to a male belonging to the 75^{th} percentile group of grooms' labor income distribution.

These estimates confirm what we found in the three-dimensional plot of Figure 2, suggesting that a large degree of assortative mating, which increases with income, remains even after controlling for age, number of children at the year of wedding and town size, which are only observable characteristics in our dataset.

The probability of assortative mating as estimated in equation (1) is however possibly flawed by simultaneity bias. For instance, it could be that a groom observed at the top 1% income group is there precisely because he got married with a top 1% income woman, or viceversa. As discussed in Barban, De Cao, Oreffice, and Quintana-Domeque (2016) simultaneity bias would lead to an overestimation of assortative mating.

Given the panel structure of our data, which provides us with incomes of all brides and grooms, over the period $2007 - 2011$, provided they both resided in Lombardy, we can estimate the following model for a generic couple i:

$$I_{k,i}^{w,t-3} = X_i^{w,h,t-3'}\beta_k + I_i^{h,t-3'}\gamma_k + u_{k,i}, \quad (2)$$

where $I_{k,i}^{w,t-3}$ is a binary variable indicating that in couple i a woman who got married at time t belonged to brides' income group k three years before wedding $(t-3)$, I_h^{t-3} is the vector indicating to which quantile group her husband belong to three years before they got married, i.e. $I_i^{h,t-3,'} = [I_{1,i}^{h,t-3}, ..., I_{100,i}^{h,t-3}]$, where $t = 2008, ..., 2011$, $k = \{1, ..., 100\}$.

In Figure 4 we report the estimated vector of coefficients $\hat{\gamma}_{w=k}$ for some selected values of $k = \{75, 95, 99, 100\}$ and contrasting this with Figure 3 one can notice that the pattern of assortative mating by income groups is broadly confirmed even dealing with the simultaneity bias. Table 5 reports the estimation results and their standard errors, showing that even after dealing with endogeneity due to simultaneity, assortative mating remains much higher at top income groups, at 13.3% for top 1% brides and top 1% grooms. Contrasting Table 5 with Table 3, it also shows that the simultaneity bias accounts for an upward bias by 2.7 percentage points, which is about 20% of the unbiased estimate.

5 Effects of assortative mating on inequality

Eventually we address the question whether this peculiar assortative mating has an effect on standard inequality measures. Assortative mating by education increases income inequality, as pointed out by Greenwood, Guner, Kocharkov, and Santos (2014). Here we assess by how much assortative mating by income percentile groups also affects inequality, which is of particular interest given the highly positive assortative mating at top income levels.

Here we consider four inequality measures, the Gini, which is a well-known statistical index, and the Atkinson indices, which are developed starting with an

Figure 3: The probability of mating a husband by percentile groups (smoothed lines)

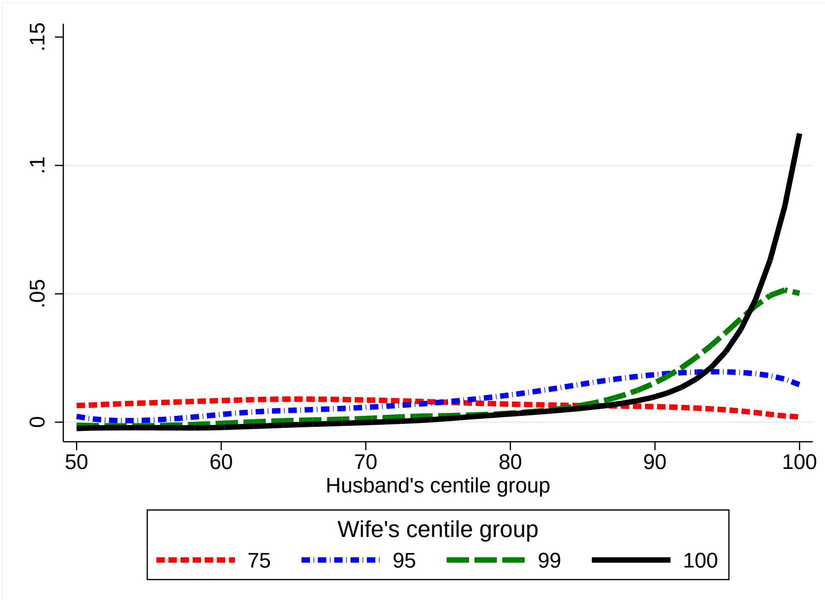


Table 4: OLS estimates of the probability of a woman belonging to percentile I_k^w to get married to a man belonging to percentile I_j^h, for some, selected k and for $j = \{50, ..., 1000\}$, at the year of wedding.

	I_{75}^h		I_{95}^h	I_{96}^h	I_{97}^h	I_{98}^h	I_{99}^h	I_{100}^h
I_{75}^w	0.018***	$\cdots$	0.012***	0.009***	0.013***	0.009***	0.002	0.000
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_{95}^w	0.011***	$\cdots$	0.030***	0.028***	0.036***	0.043***	0.049***	0.031***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
I_{96}^w	0.010***	$\cdots$	0.023***	0.035***	0.038***	0.045***	0.054***	0.038***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
I_{97}^w	0.009***	$\cdots$	0.027***	0.035***	0.030***	0.044***	0.057***	0.060***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
I_{98}^w	0.012***	$\cdots$	0.027***	0.033***	0.044***	0.049***	0.064***	0.083***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
I_{99}^w	0.010***	$\cdots$	0.024***	0.033***	0.035***	0.047***	0.064***	0.103***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
I_{100}^w	0.010***	$\cdots$	0.021***	0.027***	0.026***	0.038***	0.072***	0.160***
	(0.002)	$\cdots$	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

Notes: Control variables include the age of both spouses, the size of the town of residence, number of children of each spouse at the wedding year and year fixed effects. Standard errors in parentheses. * for p<.05, ** for p<.01, and *** for p<.001.

Figure 4: OLS estimated probability of a woman belong to a set of selected income groups to get married to a husband of income groups above the median, when they both are observed three years before wedding (smoothed lines)

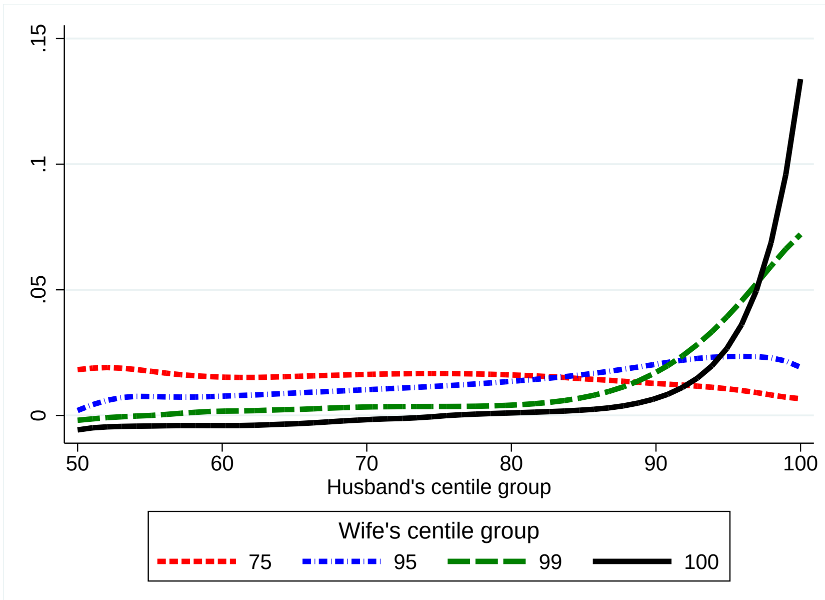


Table 5: OLS estimates of the probability of a woman belonging to percentile I_k^w to get married to a man belonging to percentile I_j^h, for some, selected k and for $j = \{50, ..., 1000\}$, when their incomes are observed three years before wedding.

	I_{75}^h		I_{95}^h	I_{96}^h	I_{97}^h	I_{98}^h	I_{99}^h	I_{100}^h
I_{75}^w	0.015***	$\cdots$	0.016***	0.011***	0.007*	-0.001	0.004	0.001
	(0.003)	$\cdots$	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_{95}^w	0.009*	$\cdots$	0.024***	0.029***	0.036***	0.024***	0.050***	0.034***
	(0.004)	$\cdots$	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
I_{96}^w	0.010**	$\cdots$	0.020***	0.026***	0.030***	0.028***	0.056***	0.031***
	(0.004)	$\cdots$	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
I_{97}^w	0.006	$\cdots$	0.025***	0.034***	0.025***	0.035***	0.039***	0.041***
	(0.004)	$\cdots$	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
I_{98}^w	0.009*	$\cdots$	0.023***	0.028***	0.041***	0.052***	0.061***	0.078***
	(0.004)	$\cdots$	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
I_{99}^w	0.007	$\cdots$	0.031***	0.026***	0.032***	0.033***	0.069***	0.096***
	(0.004)	$\cdots$	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
I_{100}^w	0.007	$\cdots$	0.013***	0.026***	0.050***	0.032***	0.072***	0.133***
	(0.004)	$\cdots$	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)

Notes: Control variables include the age of both spouses, the size of the town of residence, number of children of each spouse at the wedding year and year fixed effects. Standard errors in parentheses. * for p<.05, ** for p<.01, and *** for p<.001.

economic approach using the theory of social welfare functions. The Atkinson index may be interpreted as 1 minus the proportion of mean income that would be needed to maintain, with an equal distribution of income, the existing level of welfare, which depends on how averse one is to inequality. Here we consider inequality aversion equal to $a = 0.5, 1, 2$ (for an introduction to these indices, see Cowell, 2011). We also considered the share of total income produced by the sample of new couples at each given year that goes to some groups of the population, namely those of couples where the husband's income at the year of wedding is above the 99^{th} percentile, between the 95^{th} and the 99^{th} percentile, between 90^{th} and 95^{th}, between 50^{th} and 90^{th}.

We simulate the effect of assortative mating on income inequality by simulating two extreme reshuffles of couples. The first assumes that the husband with top income is married with the bottom income at the wedding year, the second top earning husband with the second lowest earning wife and so on. The second reshuffle does the contrary and matches the top earning husband with the top earning wife, the second top husband with the second top wife, etc. The first simulated new population of couples will remove assortative mating completely and the remaining level of inequality will all be due to other factors determining market income inequality. The second simulated population will magnify the role of assortative mating. Using these two simulated populations of couples we then estimated the measures of inequality mentioned above by year and took their average over time.

By plotting the inequality indices of those two simulated populations of couples one can have an idea of the role of assortative mating for inequality and of how far is the actual measure from its the maximum or minimum level due to assortative mating. Figure 5 plots the actual average inequality index and the simulated maximum and minimum. It shows that actual inequality is relatively closer to the maximum level, which could be reached if assortative mating on income was at its maximum level. It also shows that assortative mating accounts for a very large share of income inequality. For instance, the actual average Gini index is 0.38 but it could reach 0.17 if assortative mating was at its minimum and it would reach 0.47 if assortative mating was at its maximum. The importance of assortative mating for inequality is even more clear using the Atkinson index with $a = 2$, which suggests that with such a relatively large measure of inequality aversion the Atkinson index would range between a value of 0.79 in case of maximum assortative mating and of 0.10 in case of minimum assortative mating.

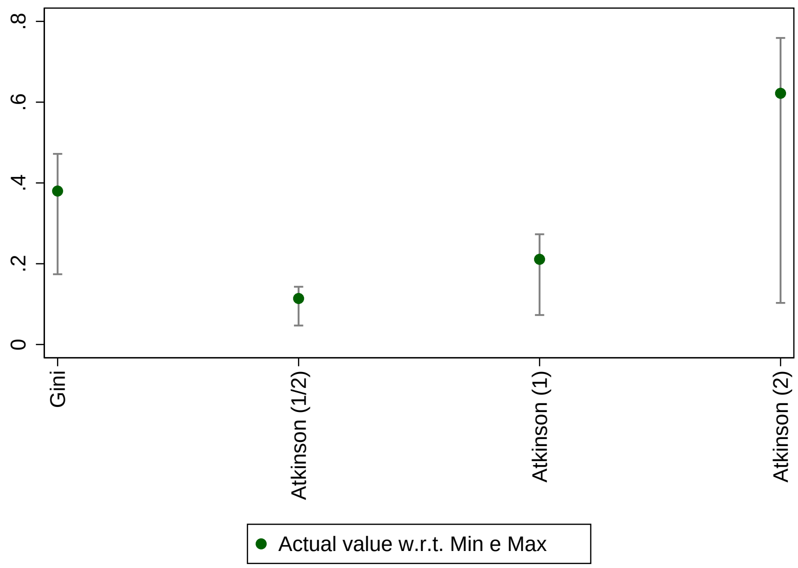
A description of the role of assortative mating for income share changes is depicted in Figure 6 where we plot the income share of each quantile group normalized by the size of the group, for quantile groups 50-90, 90-95, 95-99, 99-100. It shows that the top 1% couples earn around 7 times more that they would have under equal distribution of income, the 95-99 group earn 3 times more than under perfect equality, the 90-95 about twice more than under perfect equality an the 50-90 quantile group about the same they would get under equal distribution. Our simulated couple reshuffles would impact largely on top earning groups and progressively less on lower income groups. In particular, had couples been reshuffled so to marry the top earners to the top earners, the second top with the second top, and so on, the normalized share would have increased to more than 9; had couples been reshuffled so to marry the top earners to the bottom earner, the second top to the second bottom, and so on, the normalized share would have been about 6.

Table 6: Actual selected measures of inequality and income shares with their minimum and maximum simulated values, averaged over the period 2008-2011.

Gini			**Share 50-90**		
Actual	*Counterfactual*		*Actual*	*Counterfactual*	
	Minimum	Maximum		Minimum	Maximum
0.38	0.174 (-54.2%)	0.472 (+24.2%)	0.447	0.333 (-25.5%)	0.497 (+11.3%)
Atkinson (a=1/2)			**Share 50-95**		
Actual	*Counterfactual*		*Actual*	*Counterfactual*	
	Minimum	Maximum		Minimum	Maximum
0.114	0.047 (-58.7%)	0.143 (+26.2%)	0.09	0.065 (-28%)	0.106 (+18.4%)
Atkinson (a=1)			**Share 95-99**		
Actual	*Counterfactual*		*Actual*	*Counterfactual*	
	Minimum	Maximum		Minimum	Maximum
0.211	0.073 (-65.4%)	0.273 (+29.2%)	0.106	0.082 (-22.9%)	0.127 (+19.8%)
Atkinson (a=2)			**Share 99-100**		
Actual	*Counterfactual*		*Actual*	*Counterfactual*	
	Minimum	Maximum		Minimum	Maximum
0.622	0.103 (-83.4%)	0.759 (+22.1%)	0.071	0.064 (-9.9%)	0.089 (+25.5%)

Notes: In parentheses the change over actual measure.

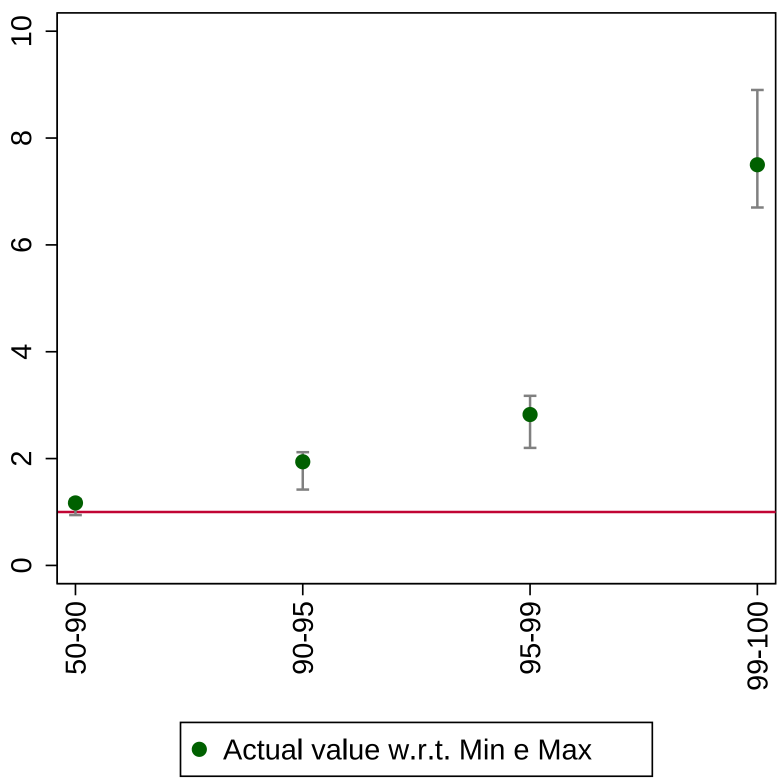
Figure 5: Actual selected measures of inequality with their minimum and maximum simulated values. Average over the period 2008-2011.



6 Concluding comments

Recent economic literature provides a large evidence that assortative mating on educational attainment is strong and increasing over time (e.g. Greenwood, Guner, Kocharkov, and Santos, 2014). Although income is on average highly correlated with past educational choices, the direct evidence of assortative mating on income is scant. By exploiting an administrative dataset of tax records for the population of residents a large region of Italy (Lombardy) over five years, and the Italian tax legislation, which allows us to trace spouses individual tax records before and after the marriage year, we provide a first measure of assortative mating on labor income levels, net of the endogeneity bias due to simultaneity that usually affects income assortative mating estimates. The administrative nature of our data and the modeling choice based on percentile groups reduce also measurement error to the minimum.

Figure 6: Normalized shares of couple incomes over the total of couples income. Average over the period 2008-2011.



Results provide evidence that top income women are much more likely to get married to top income men. Even after dealing with endogeneity due to simultaneity, assortative mating remains much higher at top income groups, at 13.3% for top 1% brides and top 1% grooms. We show that the simultaneity bias accounts for an upward bias by 2.7 percentage points, which is about 20% of the unbiased estimate.

In addition, we address the question whether this peculiar assortative mating pattern has an effect on standard inequality measures. Differently from Frémeaux and Lefranc (2015) who consider one possible counterfactual assortative mating pattern originated by a structural model of income generation and a random draw of the error from a parametric distribution, we compute the maximum and the minimum degree of assortative mating on income and found that with perfectly negative assortative mating the Gini index would halve, from 0.38 to 0.17 and it would reach a level of 0.47 in case of perfectly positive assortative mating. The range of variability of the Atkinson index would be even larger, especially if a large level of inequality aversion was assumed. For inequality aversion equal to 2, the index would range between a value of 0.79 in case of perfectly positive assortative mating and of 0.10 in case of perfectly negative assortative mating. Perfectly negative assortative mating would also reduce by 10% the share of top 1% incomes of couples.

These results are relevant also for the debate about the long term consequences of income inequality. A large literature, mainly focusing on the USA, suggests that high individual income concentration is likely to affect wealth distribution (e.g. see Saez, 2017, among others). This phenomenon will be magnified by the positive assortative mating of high income earners increasing concerns on equality of opportunities among people with different family backgrounds.

References

Arrondel, L., and N. Frémeaux (2016): "'For Richer, For Poorer': Assortative Mating and Savings Preferences," *Economica*, 83(331), 518–543.

Barban, N., E. De Cao, S. Oreffice, and C. Quintana-Domeque (2016): "Assortative Mating on Education: A Genetic Assessment," Discussion paper, University of Oxford, Department of Economics Economics Series Working Papers.

Becker, G. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–46.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012): "Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data," *The Review of Economics and Statistics*, 94(2), 371–388.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012): "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market," *Journal of Political Economy*, 120(4), 659–695.

Chiappori, P.-A., and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54(3), 832–61.

Cowell, F. (2011): *Measuring Inequality.* Oxford University Press.

Frémeaux, N., and A. Lefranc (2015): "Assortative mating and earnings inequality in France," .

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2014): "Marry Your Like: Assortative Mating and Income Inequality," *American Economic Review*, 104(5), 348–53.

Saez, E. (2017): "INCOME AND WEALTH INEQUALITY: EVIDENCE AND POLICY IMPLICATIONS," *Contemporary Economic Policy*, 35(1), 7–25.

Schwartz, C. R., and R. D. Mare (2005): "Trends in educational assortative marriage from 1940 to 2003," .