DISCUSSION PAPER SERIES

Assortative Mating and Earnings Inequality in France

Nicolas Frémeaux
Arnaud Lefranc

DISCUSSION PAPER SERIES

IZA DP No. 11084

Assortative Mating and Earnings Inequality in France

Nicolas Frémeaux
Université Paris 2

Arnaud Lefranc
University of Cergy-Pontoise, THEMA and IZA

OCTOBER 2017

ABSTRACT

Assortative Mating and Earnings Inequality in France[*]

This paper analyzes economic assortative mating and its contribution to inequality in France. We first provide descriptive evidence on the statistical association in several socio-economic attributes of partners among French couples (annual earnings, potential earnings, education, occupation). Second, we assess the contribution of assortative mating to earnings inequality between couples. Contrary to previous estimates, we account for possible biases in the estimation of assortative mating arising from sample-selection into the labor force. We also provide a new method for assessing the contribution of assortative mating to inequality in couple's potential earnings. Our results indicate a strong degree of assortative mating in France. The correlation coefficient for education is above 0.6. The correlation in earnings is lower but sizable: around 0.17 for annual earnings, when including zeroes; around 0.35 for full-time equivalent earnings and up to 0.49 when using multi-year average earnings. We show that assortative mating tends to increase inequality among couples, compared to random mating. For annual earnings, the effect is non-negligible and accounts for 3 to 9% of measured inequality. The effect of assortative mating on household potential earnings is much larger and amounts to 10 to 20% for observed inequality.

JEL Classification:	J12, J22, D31
Keywords:	assortative mating, inequality, earnings, labor supply, France

Corresponding author:
Arnaud Lefranc
Université de Cergy-Pontoise
33 boulevard du Port
95011 Cergy-Pontoise
France
E-mail: arnaud.lefranc@u-cergy.fr

1 Introduction

An abundant sociological literature has provided evidence of a high correlation of educational and social attributes within couples, in most developed countries (e.g. Mare 1991, Blossfeld and Timm 2003). In comparison, available evidence on the extent of assortative mating according to economic characteristics is much more limited. Investigating the degree of homogamy in modern societies is however crucial for at least three reasons. First, the propensity to mate into homogenous couples might amplify existing earnings inequality between individuals. Although several papers have recently investigated this issue[1], the extent to which assortative mating contributes to economic inequality between couples remains largely unknown. Second, as discussed in Becker (1973) and Zhang and Liu (2003), observed assortative mating patterns might shed light on the nature of intra-household production and allocation decisions. Lastly, to the extent that it shapes household resources, assortative mating will largely condition child upbringing decisions and might contribute to the intergenerational transmission of inequality (e.g. Becker and Tomes 1979, Black and Devereux 2011).

In this paper, we study economic assortative mating in France. Our contribution is threefold. We first provide comparable evidence on assortative mating among French couples for various attributes (occupation, education, annual earnings), as usually investigated in the literature. Second, in order to account for endogenous labor supply, we examine the association within couples in individual potential earnings, measured by full-time equivalent earnings. Moreover, we account for potential biaises in the estimation of assortative mating arising from sample-selection into the labor force. Third, we assess the contribution of assortative mating to inequality between couples in both observed annual earnings and potential earnings.

Available evidence on the extent of economic assortative mating appears relatively sparse. Most studies have focused on assortative mating by education (e.g. Goux and Maurin 2003, Schwartz and Mare 2005) or social origin (e.g. Kalmijn 1991, Uunk, Ganzeboom, and Róbert 1996). Assortativeness along other economic dimensions such as individual earnings or preferences has been much less analyzed[2]. This represents an important

---

[1]See in particular Karoly and Burtless (1995), Cancian and Reed (1998), Burtless (1999), Schwartz (2010), Eika, Mogstad, and Zafar (2017) Greenwood, Guner, Kocharkov, and Santos (2014), Harmenberg (2014), Pestel (2017)

[2]Arrondel and Fremeaux (2016), Dohmen, Falk, Huffman, and Sunde (2012) and Kimball, Sahm, and Shapiro (2009) are some of the few exceptions.

limitation for at least two reasons. First, it does not allow to fully capture the contribution of marital choices to economic inequality. Second, in a period of rising returns to skills, a constant degree of educational or occupational assortativeness might hinder a rising polarization of the distribution of family resources. To partially address these issues, recent research has examined the statistical association between male and female labor earnings within couples. Available evidence points to a sizable correlation, of up to 20%, in individual earnings (e.g. Burtless 1999, Nakosteen, Westerlund, and Zimmer 2004, Schwartz 2010). The analysis is however largely confined to the United States and much less is known of the situation in European societies.[3]

Existing studies suffer from several empirical limitations. First, estimates are generally based on cross sectional data in which earnings are only observed on a single year. However, annual earnings might incorporate sizable measurement errors and transitory shocks that can bias downward the estimates and lead to an underestimation of the association in partners' earnings.[4] In this paper, we exploit panel data to compute average earnings over multiple years in order to address this issue. Second, most papers have focused on the statistical association in annual earnings. However annual earnings reflect both individual productivity characteristics and endogenous joint labor supply decisions taken within the couple. The confounding effect of labor supply decisions might jeopardize the assessment of the degree of assortative mating. An important concern, in this respect, is that a sizable share of women in couples report zero earnings as they do not participate in the labor force. If labor force participation is positively associated with partner's earnings, this will lead to underestimate the degree of assortative mating in individual economic characteristics.

In this paper, this issue is addressed by analyzing the statistical association in potential earnings within couples. Potential earnings are defined by the individual full-time equivalent earnings. Potential earnings of a couple thus represent the earnings it would receive if each partner worked full time, given the individual market wage rate of its members. Compared to reported annual earnings, potential earnings provide a more extensive measure of the total economic resources commanded by the couple, which is more relevant to

---

[3]Among the few exceptions are : Nakosteen, Westerlund, and Zimmer (2004) on Sweden,Pestel (2017) on Germany, Eika, Mogstad, and Zafar (2017) on Norway, Germany and Denmark. The present paper only uses the French version of the EU-SILC database. The analysis will be extended to other European countries in future research.

[4]The incidence of measurement errors has been widely documented in the related field of intergenerational earnings mobility studies. See for instance Solon (1992) and the survey of Black and Devereux (2011).

assess inequality in welfare between households. First, individuals out of the labor force might have a positive contribution to the household's consumption of goods and services through domestic production, as emphasized in Gronau (1977). Available estimates indeed suggest that domestic production represent a sizable fraction of household consumption.[5] Measures of household production value usually combine individual market wage information with time-use surveys to value the domestic production of basic services (cleaning, gardening, shopping...). This leaves aside the value of leisure enjoyed, and, for households with children the value of human capital investment undertaken at home. Our measure of potential earnings values total available time at the prevailing individual market wage. This can be seen as an encompassing measure of the resources available that ultimately determine household welfare.

Of course, one of the difficulties in assessing the intra-household correlation in potential earnings is that the market wage rate is not observed for individuals out of the labor force. This problem does not arise when using observed earnings (including zeroes). We explicitly account for sample selection due to non-participation and provide estimates of the intra-couples correlation in (possibly latent) potential earnings by extending the usual regression with sample-selection model.

One of the main economic motivations for studying assortative mating lies in its potential contribution to economic inequality between couples. Empirical analyses of earnings inequality have mainly stressed the influence of aggregate shocks (rise in the returns to skills, skill-biased technological change, globalization, etc.), institutions and policies (labor market deregulation, decrease in marginal income tax rates, etc.) as the main drivers of the recent rise in inequality in most developed countries. The effects of demographic factors, in particular assortative mating patterns, have only been studied recently and the effect of assortative mating on inequality is generally found to be modest. Specifically, Greenwood, Guner, Kocharkov, and Santos (2014) estimate that the Gini coefficient for the United States would decrease from 0.43 to 0.42 when random matching is imposed while Eika, Mogstad, and Zafar (2017) conclude that the contribution of assortative mating to inequality is around 5% . The main route taken in the literature is to compare the observed earnings distribution to a counterfactual distribution built under alternative hypothetical mating patterns. However, the construction of this counterfactual distribution requires to

---

[5]See for instance House, Laitner, and Stolyarov (2008), Frazis and Stewart (2011), Ahmad and Koh (2011), Roy (2012).

adequately deal with the endogeneity of labor supply decisions and the self-selection of individuals into couples, on the basis of their unobserved characteristics.

Two main approaches have been taken, in the recent literature, to build these counterfactual distributions. The accounting approach, also referred to as 'addition randomization' in the literature, treats observed annual earnings as a fixed individual characteristic and simulates the distribution that would prevail if individuals kept their labor earnings unchanged and were randomly matched into couples (e.g. Karoly and Burtless 1995, Cancian and Reed 1998, Burtless 1999, Schwartz 2010, Hryshko, Juhn, and McCue 2014). Hence, this approach ignores the labor supply responses that would result from the random rematching of individuals. The so-called behavioral approach (or 'imputation randomization') characterizes individuals by some observable earnings determinants, in general education (e.g. Greenwood, Guner, Kocharkov, and Santos 2014, Harmenberg 2014, Pestel 2017, Eika, Mogstad, and Zafar 2017). Individuals are then randomly rematched into counterfactual couples. The joint earnings of the counterfactual couples are simulated on the basis of the observed distribution among actual couples with similar observable earnings determinants. Hence, this approach takes into account the endogeneity of labor supply decisions, but only to the extent that it is driven by observable characteristics of the mates. Furthermore, it ignores the self-selection of individuals into couples on the basis of their unobservable attributes.

In this paper, we develop a third approach in which we characterize the effect of assortative mating on inequality in couples' *potential* earnings. Compared to existing studies, our approach offers three main advantages. First, as previously discussed, potential earnings provide a broader and more relevant measure of household resources. Second, since potential earnings are defined as the earnings an individual would receive if he/she worked full-time, this alternative measure of resources is largely independent of joint-labor supply decisions in the couple, contrary to annual earnings.[6] Our assessment of the impact of assortative mating on inequality relies on a statistical model of the joint distribution of the potential earnings of both partners that allows for sample selection in the observed distribution and correlation across partners in their unobservable earnings determinants. The third advantage of our approach, compared to other simulation methods discussed in

[6]One limitation is the possibility that individual market wage is determined by the past labor supply decision, as discussed for instance in Eckstein and Lifshitz (2011). In this paper, we do not accounting for the dynamics of human capital and employment opportunities.

the previous paragraph, is thus the ability to account for self-selection of individuals into couples on the basis of their unobservable attributes.

Our empirical analysis is based on the French waves of the EU-Statistics on Income and Living Conditions (SILC), covering the period 2004-2011. Our results indicate a strong degree of assortative mating in France. The correlation coefficient for education is above 0.6. The correlation in earnings is lower but sizable. Specifically, for dual-earner couples, the correlation is around 0.3 for annual earnings and 0.35 for full-time equivalent earnings. We then show that sample-selection leads to a moderate upward bias in the estimation of the within-couple correlation. We also investigate the extent of non-linearities in the statistical association of earnings and show that positive assortative mating is particularly high at the top of the earnings distribution. Lastly, our estimates indicate that assortative mating tends to increase inequality among couples. For annual earnings, the effect is non-negligible. The addition randomization approach indicates a contribution to inequality between 9 and 18%. The imputation randomization approach points to a smaller effect of 3 to 9% of measured inequality. The effect of assortative mating on household potential earnings is however much larger and amounts to 10 to 20% for observed inequality. The effect of assortative mating is found to be larger for inequality indices more sensitive to the tails of the distribution which is consistent with the non-linearities of assortative mating. These findings are robust to the model used for simulating the counterfactual distribution and to sample selection.

The rest of this paper is structured as follows. Section 2 presents the data. Section 3 provides summary measures of the degree of assortative mating for various individual attributes (education, socio-economic status, social origin, earnings). In section 4, we focus on the issue of sample selection. Section 5 estimates the contribution of assortative mating to earnings inequality among households.

2 Data

2.1 EU-SILC

Our analysis is based on the European Union - Statistics on Income and Living Conditions (EU-SILC) surveys. We focus on the waves 2004 to 2011 of the French sample. The EU-SILC is a longitudinal household survey, coordinated by Eurostat, which gathers data from

all EU member states. The main goal of the survey is to study income, poverty, social exclusion and living conditions in the European Union. The French waves were collected by the French national statistics institute (INSEE)[7].

Data are collected annually for a rotating panel of households. In the French sample, individuals are followed for a period of up to 8 years. The survey provides information on the composition of the household, the link between its members, as well as unique individual identifiers. The main sampling unit is the household. We define a couple as a unique pair of individuals reporting to be respectively head and married or common law partner of the head in a given household. Other pairs of individuals living in the same household are not considered as a couple. Our sample includes all couples regardless of their legal status (married or not).

We restrict the sample to couples in which both partners are between 25 and 60 years old, in which neither partner is self-employed and in which neither partner is out of the labor force because of retirement or studying. We only keep one observation per couple. For each individual in a couple, we keep the observation with non-missing information of the variables of interest which is closest to the age of 35. This choice is made in order to minimize the incidence of life-cycle earnings dynamics on our measure of economic assortative mating (Haider and Solon 2006). This results in a sample of 7,966 couples. In the main analysis, we also exclude couples in which earnings are zero for both partners.[8] In the end, our analysis is based on sample of 7,864 couples. Appendix A provides general descriptive statistics on our final sample.

2.2 Main variables

We examine two types of individual characteristics : earnings and measures of socio-economic achievement. Appendix A provides detailed information about the construction of variables.

Earnings *Annual earnings* are defined as the total wage and salaries earned in the previous year deflated by the consumer price index. For individuals out of salaried employment,

---

[7]National quality reports about the EU-SILC survey are available here: `http://ec.europa.eu/eurostat/web/income-and-living-conditions/quality/national-quality-reports`

[8]This corresponds to 102 couples.

the value of annual earnings is equal to zero[9].

Full-time equivalent (FTE) earnings are defined as annual earnings/(number of months worked full-time + 0.5 × number of months worked part-time) × 12. To compute FTE earnings, we rely on the history of labor force participation reported in the survey. For individuals out of salaried work, FTE earnings are missing, by construction.

For both earnings measures, we compute multi-year averages of individual earnings. The average is computed over the full set of available yearly observations. The number of years of observation in our sample varies between 1 and 8 years, with an average of 3.4 years.

Other socioeconomic variables   *Education*- The first measure is the number of years of education, equal to the school leaving age minus 6 years (i.e. minimum age for compulsory education)[10]. Our second variable is based on the ordered classification of the highest degree completed.

Occupation- Our measure of occupation is based on the standard 6-levels French classification. In order to come close to an ordinal measure of occupation, we gather farmers and unskilled manual workers. The SILC survey investigated individual socioeconomic origin and gathered information on education and occupation of both parents of adult respondents. Information is only available in 2005.

3   Descriptive measures of assortative mating

3.1   Education and occupation

We first analyze the extent of assortative mating in socio-economic achievement by estimating the partners correlation in occupation and education. Information is available for both partners of the couple, as well as for their parents. For ordinal variables (occupation and highest degree completed), the association is measured using two indicators : the Spearman

---

[9]Given our use of panel data, the individuals with zero earnings should have never reported any salaried activity. Some of these individuals may however report unemployment period and so potentially unemployment benefits. Taking into account these benefits (as a proxy for earnings) does not change our estimates but it increases the measurement errors. In the end, we decided not to include them.

[10]For some individuals, the number of years of education appears noisy. Furthermore, although highest degree is reported for all individuals in the sample, number of years of education is missing for 9% of the sample. For this reason we estimate the correlation in *predicted* number of years of education, where the prediction is based on a regression of number of years of education on degree dummies interacted with gender and a fourth degree polynomial function of birth cohorts.

correlation coefficient measures the statistical association in the distributional ranks of two variables; the polychoric correlation assumes that the discrete variable that measures each partner's attainment (degree, occupation) is determined by a latent variable, following a multinomial model. The polychoric correlation is defined as the linear Pearson correlation coefficient for the latent variables of the two partners and is parametrically identified.[11] For the number of years of education, we report linear (Pearson) correlations and Spearman rank correlations.

Occupation Table 1 provides our estimates of assortative mating for occupation and education. Occupational correlations are given in panel A. The correlation in partners' own occupation ranges between 0.453 and 0.531 (column 1), which appears high, though in line with estimates found for other countries.

This can be compared to estimates of the correlation in social origin, as captured by parental occupation. Columns 2 and 3 compare the correlations in own occupation with the correlation in father's occupation, on the sub-sample where father's occupation is reported. Columns 4 and 5 report the same analysis for mother's occupation. On these sub-samples, the correlation among partners in own occupation (columns 3 and 5) is very similar to the whole sample (column 1). The correlation among partners in fathers' or mothers' occupation is positive and around 0.3, which indicates positive assortative mating by social origin. Note though that the correlation in parental occupation is lower than the correlation in patners own occupation, which indicates that assortativeness depends more on individual occupational attainment than on social origin. The correlation is higher for fathers' occupation (0.29-0.377) than for mothers' (0.249-0.308). It is important to keep in mind that the absence of information for a significant share of respondents' mother (mainly because of inactivity) makes the comparison difficult. The high level of assortative mating and the difference between the partners and their parents are consistent with existing evidence on French data (Bouchet-Valat 2014).

Education Panels B and C of Table 1 report statistical associations in education. Panel B uses the highest degree completed. On the whole sample, we find positive correlations between 0.559 and 0.593. The difference between the two measures of correlations (Spearman rank correlation vs. polychoric correlation) is small. These correlations appear higher

---

[11]The idea of polychoric correlation dates back to Pearson (1900) and Ritchie-Scott (1918)

for education than for occupation. The correlation between partners is also higher for own education than for social origin, as captured by parents' education. However, compared to panel A, the differences between own and parental characteristics appear smaller for education than for social class.

Panel C provides correlation estimates for a continuous measure of education, the number of years of education. The correlations are higher, around 0.62, but consistent with those obtained for the correlation in highest degree completed.

Overall, our results indicate high levels of positive assortative mating in France. These results are consistent with existing evidence on France (Goux and Maurin 2003, Bouchet-Valat 2014). They can be compared with the results presented in Fernandez, Guner, and Knowles (2005) for the correlation in education in a large set of countries. Our estimates for France appear higher than the correlation reported for most European countries, with the exceptions of Spain, Belgium and Italy. They are similar to those reported for the US and lower than those found in most Latin American countries (around 0.8).

3.2 Earnings

Annual and FTE earnings To assess the extent of economic assortative mating, we now examine the correlation between partners in annual and full-time equivalent (FTE) earnings.

Results are presented in table 2. Column 1 reports correlations in annual earnings based on all observations, including zeroes. The correlation between partners in annual earnings is around 0.175. Column 2 focuses on dual-earner couples, in which both partners report positive earnings. The correlation in this sample is significantly higher (0.31 for the Pearson correlation). The gap in the estimated correlation between the two samples is likely to be explained by non-participation in the labor force. When earnings are zero for one the partners, it is predominantly female earnings. Assume first that labor force participation of women is independent of male earnings. In this case one would expect the correlation coefficient to fall when non-participants with zero earnings are taken into consideration.[12] Whether the assumption of random participation constitutes a reasonable approximation

---

[12]In fact, under random participation, the presence of zeroes would mechanically lead to a decrease in the covariance of earnings among partners. Furthermore, the inclusion of zeroes would likely (although not surely) increase the variance of earnings in each marginal distribution. These two effects would then converge to decrease the correlation coefficient.

is of course open to discussion and we shall return to this issue below. But note, however, that if female non-participation is more likely in couples with higher male earnings this will further reinforce the fall in earnings correlation when including observations with zero earnings.

In the last column of table 2, we examine the correlation in FTE earnings. This allows to remove the correlation in labor supply decisions within the couple that affects the correlation in annual earnings and focus on the correlation in potential earnings. As in column 2, we focus on dual-earner couples. This results in a much higher correlation, up to 0.351 for the Pearson correlation. Compared to column 2, removing heterogeneity across individuals in the number of months worked full and part-time increases the correlation in earnings by about 13%. This increase indicates that the correlation within couples in hours worked is lower than the correlation in hourly wage rate. It confirms, along the intensive margin, our discussion, in the previous paragraph, of the incidence of of labor supply decisions. We address this issue more carefully in section 4.

One may suspect that part of the correlation in earnings arises from life-cycle effects, through the correlation in birth cohort within couples. For all columns, we thus estimate the correlation in earnings after netting out cohort effects.[13] Result indicate a modest fall in the estimated correlation. The Pearson coefficient falls by 3.5 to 4.5%. The effect on the Spearman correlation is even smaller.

Two conclusions can be drawn from table 2. First, results indicates that assortativeness in earnings is high in France compared to other countries. On a similar sample from the US population, Schwartz (2010) estimates a correlation of 0.12 for all couples (including couples in which one of the partners is out of the labor force) and a correlation slightly higher than 0.2 for dual earner couples. Our estimates are 45% and 55% higher, respectively, in France. Second, the table also indicates that labor supply decisions (along both the extensive and the intensive margins) attenuate the correlations of potential earnings. In other words, marital sorting according to potential labor earnings is high but the labor supply decisions pertaining to labor force participation and part-time work tend to dampen the correlation in partners' earnings.

---

[13]This is achieved by first regressing earnings on a quartic function of birth cohort and taking residuals.

Contribution of education and social origin As noted in the introduction, most papers focus on assortativeness by education or social origin. Both variables capture dimensions along which marital sorting should obviously occur, given the interplay between socialisation processes and mating decisions. However, it is also relevant, for understanding the socio-economic determinant of mating decisions, to investigate whether sorting also occurs once individual social characteristics have been taken into account. Actually, one may object to the analysis of assortativeness by earnings that it merely reflects the correlation in partners' education and social origin. To address this issue, we examine whether earnings remain correlated, once they have been purged from the effect of education and social origin.

Table 3 presents estimates for correlations based on earnings residuals after controlling for education, social origin or both variables[14]. First, labor earnings remain positively correlated, even after controlling for individual educational attainment and social origin. Controlling for education alone (Panel B) decreases the correlation by about 35%. Controlling for social origin (Panel C) has a smaller impact on the correlation that falls by 20 to 25%. Last, comparing panels B and D indicates that once education is accounted for, further conditioning on social origin leaves the correlation in earnings almost unchanged. As a conclusion, even if assortativeness in terms of social background and of education is high, as discussed in section 3.1, there is still significant sorting along other dimensions not captured by these variables.

Multi-year average earnings A potential challenge to the measurement of earnings correlation is the incidence of measurement errors and transitory income components. Under measurement error, the correlation in annual measures of earnings might underestimate the correlation among partners in permanent earnings. The degree of underestimation will depend on the variance of measurement errors and the correlation among partners in transitory earnings components, compared to permanent components.

One way of moderating the incidence of these biases is to use average earnings, computed over multiple years of observations. This is undertaken in table 4. For each individual and each measure or earnings (annual and full-time equivalent), we compute average earnings using all available time observations. Since the number of observations over which individuals are observed varies across individuals, these averages are computed over vari-

---

[14]We restrict the sample to couples with valid information on education and social origin.

able horizons. We consider two sub-samples. In panel A, we estimate earnings correlations on the sample of couples observed during at least 3 years; in panel B, we focus on couples who are observed during at least 5 years.

Using multiple-year averages has a limited effect on our measure of the correlation in annual earnings. The linear correlation coefficient increases by 13% when averaging annual earnings over at least three-years. Using average earnings has a similar effect on the correlation in full-time equivalent earnings that increases by about 17% to reach a high value of 0.466. Again, estimating correlations on earnings residuals net of cohort effects barely changes the results. When averaging earnings over a period of at least five years, the estimated correlations reach an even higher value : 0.416 for annual earnings and 0.49 for FTE earnings.

While averaging earnings affects our measure of assortativeness in the expected direction, the size of the effect is lower than expected a priori. In a related context, intergenerational elasticity estimates indicate that using current earnings in place of permanent earnings leads to underestimate the intergenerational association in earnings by about one third. This is consistent with available evidence indicating, first, that measurement errors in annual earnings account for 10 to 15% of the variance in earnings (e.g. Duncan and Hill 1989, Hagneré and Lefranc 2006) and, second, that transitory components account for roughly one fourth of total earnings variation (Moffitt and Gottschalk 2011). However, in our case, earnings data are derived from administrative data after 2007. Additionally, as discussed in Appendix A, winsorizing the extreme one percent of the distribution should also reduce the incidence of measurement error. Furthermore, contrary to what occurs for intergenerational estimates, transitory earnings and not just permanent components are likely to be correlated within couples, to the extent that they relate to factors such as local labor market conditions or other household level shocks. Ostrovsky (2012) reports supportive evidence. In our case, given limited sample size and time-series depth, we cannot directly investigate this issue.

In the end, using average earnings reinforces the view that earnings are highly correlated within couples in France.

Non-linearities in assortative mating We now examine the extent of non-linearities in the association in earnings among couples. Figures 1 and 2 provide evidence that the sta-

tistical association in earnings vary along the earnings distribution. These figures present the contour plot of the bivariate earnings distribution among couples. The first panel gives the contour plot of earnings in level. For annual earnings, there seems to be little correlation in the lower tail of the distribution and a stronger one at the top. This is confirmed by the second panel, which represents the joint distribution of the ranks. Under the assumption of joint normality (or joint log normality) of the earnings distribution, this contour plot should be symmetric around the middle point of the box and should display two equal-sized peaks at the bottom and at the top of the distribution. In our case, the distribution of ranks is bimodal, but displays a much higher peak at high quantiles, indicating that earnings correlation is larger at the top of the earnings distribution.

4   Sample selection and assortative mating

4.1   Model

The results of the previous section indicate that the correlation in labor earnings is influenced by labor supply decisions, along both the intensive and extensive margins. Unfortunately, none of the above estimations provides a satisfactory measure of the extent of the partners correlation in both economic resources and potential earnings. On the one hand, using all observations, including those with zero earnings amounts to ignore that people out of the labor force might produce economic resources domestically or enjoy higher welfare due to increased leisure consumption. On the other hand, the simple correlation in full-time equivalent earnings computed from the sample of dual-earner couples ignores possible sample selection into participation. Since participation decisions depend on the earnings of both partners, selection is likely to be non-random. In this case, the correlation in full-time equivalent would provide a biased estimate of the correlation in potential earnings, although the direction of the bias is a priori unknown.

Unbiased estimates of the correlation in potential earnings can be derived from a wage regression model that explicitly accounts for sample selection. Let w_s denote the earnings of partner s, with $s = m$ for the male partner and $s = f$ for the female partner. We assume

that (w_m, w_f) follows a bivariate log-normal distribution :

$$\begin{pmatrix} w_m \\ w_f \end{pmatrix} \to \ln \mathcal{N}(\mu, \Sigma) \quad \text{with} \quad \mu = \begin{pmatrix} \mu_m \\ \mu_f \end{pmatrix} \text{and } \Sigma = \begin{pmatrix} \sigma_m^2 & \rho\sigma_m\sigma_f \\ \rho\sigma_m\sigma_f & \sigma_f^2 \end{pmatrix}$$

Under the assumption of bivariate log-normal distribution the relationship between male and female earnings can be written as :

$$\ln w_f = \beta_0 + \beta \ln w_m + \varepsilon \tag{1}$$

where the regression slope satisfies $\beta = \rho\sigma_f/\sigma_m$ and is thus equal to the correlation coefficient of the variables in logarithm, rescaled by the standard errors ratio of male and female.

Assume first that w_m is always observed but that w_f is only observed for women in the labor force[15]. In the likely case where participation decisions depend on both partners' potential earnings, the sample of dual earners is no longer representative of the entire population. In this case, the partners' correlation cannot be directly assessed, based on observed earnings alone. Likewise, the distribution of w_f will be censored by participation decisions and the estimation of the standard errors of female earnings from observed data will be biased. However, equation 1 can be consistently estimated using Heckman's sample selection correction model. This yields consistent estimates of both β and σ_ε.

The estimates obtained from the sample selection regression model can be combined with estimates of σ_m to obtain an estimate of the within-couple correlation in log-earnings, ρ. It is given by:

$$\rho = \beta \frac{\sigma_m}{\sqrt{\sigma_\varepsilon^2 + \beta^2\sigma_m^2}}$$

We use this approach to estimate the partners correlation in residual earnings, i.e. net of age and time effects. The participation equation includes controls for the number of children in the household, household capital income, a quadratic function of the annual labor earnings of the husband, an indicator of whether the husband holds a long-term labor contract and a quadratic form in the age of both partners.

In principle, estimates of this model could also be biased if there is non-random selection

---

[15]Table A.1 shows that the share of men reporting positive earnings equals 94% while this share equals 77% for women.

in the observability of male earnings, although this is much more rarely the case in our sample. We investigate this issue in appendix C where we estimate a double selection model. Results indicate that selection based on the observability of male earnings can be ignored in the analysis of assortativeness within couples.

4.2 Results

Estimation results are given in Tables 5 and 6. Table 5 provides estimates of the regression coefficient, correlation coefficient, both in logarithm form, and earnings standard-deviations. Given the pattern of female labor participation and the incidence of part-time work among female, the assumption of joint log-normal distribution, discussed in the previous section, does not appear relevant for annual earnings. Hence, we concentrate here on FTE earnings.

Estimates in panel A ignore sample selection issues. The results found here are very similar to those reported earlier: The estimate of the correlation in log-FTE earnings is .326, compared to .337 for the correlation in levels, once cohort effects have been removed (Table 2, panel B). Estimates in panel B are obtained using Heckman's sample selection model. Ignoring sample selection issues leads to slightly overestimate the extent of the earnings correlation. Specifically, the correlation falls from 0.326 to 0.31 and from 0.361 to 0.353 in the case of multi-year average FTE earnings. This fall in the estimated correlation arises from two effects : first, a fall in the partners earnings elasticity (β), once selection is taken into account; second, a rise in the dispersion of female earnings, once we account for the fact that the distribution of female earnings in truncated owing to the participation decision. This suggests that sample selection into employment has only a moderate impact on the estimated earnings correlation.

Table 6 gives the estimates of the Heckman sample selection model. Analyzing the results of the selection equation allows a better understanding of the process that determines whether female partners work for pay. ρ_{res} indicates the correlation coefficient of the error terms of the selection and wage equations. For all specifications, this coefficient is negative. This indicates that women with a positive earnings residual, conditional on their partner's earnings have a lower probability of working for pay. In other terms, for female partners, "undermarriage" (i.e. women with high potential earnings conditional on their partner's earnings) is associated with lower participation and "over marriage" is associated

with higher participation. This result illustrates that the idiosyncratic disutility of work, captured by labor supply unobserved determinants, are not independent of the idiosyncratic potential earnings of the mate.

Table 6 also allows assessing the relationship between the male earnings and female labor market participation. The coefficients of w_m and w_m^2 indicates a hump-shaped relationship. Table 7 provides additional evidence on female labor market characteristics conditional on the male FTE earnings. The female employment rate rises along the male earnings distribution. After a sharp increase between the first and second deciles (D1 vs. D2), the employment rate increases steadily up to the sixth decile and plateaus to about 80% until the ninth decile but significantly falls in the top decile. The lower female employment rates at the tails of the distribution of male earnings mostly reflect a low participation rate, rather than a higher risk of unemployment (columns 2 and 3). As previously discussed, under random participation to the labor market, we would expect that excluding individuals with zero earnings would increase the observed correlation in earnings. This is partly reinforced by the hump-shaped pattern in labor-force participation observed in column (1). Second, the number of months worked (conditional on being in employment) follows a similar hump-shaped pattern, although the variation across male earnings deciles is rather limited. In sum, there seems to be more variation, across male deciles, in female labor supply along the extensive margin than along the intensive margin. Third, although, overall, female earnings increase with male earnings, the relationship is relatively flat in the bottom half of the distribution (D1 to D4). This seems particularly true for FTE earnings. However, the gradient in female earnings conditional on male earnings, at the top of the distribution seems steeper for FTE earnings than for annual earnings. Hence, the increase in the observed correlation in earnings when using FTE earnings rather than annual earnings seems largely driven by a rise in the statistical association between male and female earnings at the top of the earnings distribution.

5 The contribution of assortative mating to earnings inequality among households

5.1 Methods

Assessing the contribution of assortative mating to earnings inequality among households requires comparing the observed distribution of earnings to a counterfactual distribution that would prevail under alternative mating patterns. In line with several recent papers, the counterfactual mating pattern we consider corresponds to the hypothesis of random matching.[16]

As discussed in Harmenberg (2014), two main methods have been used in the literature to build a counterfactual earnings distribution, under the assumption of random mating. The first approach is followed by Hryshko, Juhn, and McCue (2014) and to some extent Burtless (1999). It amounts to take observed labor earnings of male and female as a fixed individual characteristic and to randomly match individuals into simulated couples. Household earnings are computed as the sum of the labor earnings of both partners in the simulated couples. In this case, the counterfactual distribution is simply a convolution of the marginal earnings distribution of female and male partners observed in the population. Following Harmenberg (2014), we refer to this method as *addition randomization.* The major limitation of this approach is to assume that individual labor supply decisions are exogenous with respect to match characteristics.

An alternative approach is implemented in Greenwood, Guner, Kocharkov, and Santos (2014) and Eika, Mogstad, and Zafar (2017). In this approach individuals are characterized by some observable characteristics Z, such as education. The total earnings of a household are determined by the characteristics of both partners, Z_m and Z_f. For each combination of partners' characteristics, a (conditional) household earnings distribution can be computed. Randomization amounts to create pseudo-couples in which the characteristics Z of both partners are randomly drawn from the observed distributions of Z characteristics (among male and female partners) in the population. Once the characteristics of both partners of the pseudo-couple are defined, household earnings are randomly drawn from the observed distribution of household earnings, conditional on partners' characteristics. Hence, the

---

[16]Several papers focusing on the effect of *changes* in assortative mating on the income distribution (e.g Karoly and Burtless 1995, Burtless 1999) rely on a different counterfactual, usually the mating pattern observed in a reference year.

counterfactual distribution is a mixing of observed conditional earnings distribution, where the mixing weights are defined by the random mating hypothesis. We refer to this approach as *imputation randomization*.

To illustrate the imputation approach, assume that the population of individuals is split *equally* into two groups, regardless of gender : high education individuals, denoted by H and low education denoted by L. Based on education, we distinguish four types of couples : HH, HL, LH, and LL. For each type, we observe the cumulative earnings distribution function among couples with this type : $F_{HH}(y)$, $F_{HL}(y)$, ... Let p_{HH}, p_{HL}, p_{LH}, p_{LL} denote the weight of each type in the population of couples. The actual CDF of the distribution of earnings among couples is equal to : $F(y) = p_{HH}F_{HH}(y) + p_{HL}F_{HL}(y) + p_{LH}F_{LH}(y) + p_{LL}F_{LL}(y)$. If the characteristics of partners were drawn randomly in the population, the share of each type among couples would be equal to $\frac{1}{4}$ (again assuming equal shares of H and L individuals among males and females). Hence the counterfactual distribution under imputation randomization is, in this case, given by $\tilde{F}(y) = \frac{1}{4}\{F_{HH}(y) + F_{HL}(y) + F_{LH}(y) + F_{LL}(y)\}$.

The advantage of the imputation randomization, compared to the addition randomization approach, is to allow for endogenous labor supply responses, but only as long as they depend on the conditioning variables Z.[17] In other words this amounts to rule out the possibility that household labor supply decisions and earnings be also determined by partners' unobserved characteristics whose distribution may differ across observed couples with different combinations of Z. The results in section 4 suggest that this assumption may fail to hold, as labor supply unobserved determinants seem to depend on the productivity characteristics of the match. It is also worth stressing that, according to the results in table 3, the correlation in earnings cannot be fully accounted for by the correlation in the conditioning variables (education).

Both approaches above attempt to quantify the effect of assortative mating on inequality of *realized* household annual earnings. We also implement a third approach that allows assessing the effect of assortativeness on inequality of household *potential* earnings, defined as the earnings the couple would earn if both partners worked full-time. Contrary

---

[17]The procedure developed by Pestel (2017) may be linked to the imputation approach. It amounts to randomize individuals with different wage rates into counterfactual couples and to simulate labor supply decision based on a household labor supply model. Wage rates are, however, predicted on the basis of socio-demographic characteristics such as education. The model thus fails to account for assortative mating along unobserved earnings determinants.

to realized earnings, which are partly determined by joint labor supply decisions within the household, potential earnings can largely be considered as an exogenous individual characteristic, with respect to couple composition.[18]

The contribution of assortative mating to inequality across couples in household potential earnings can be assessed using three approaches. We can first implement the addition and imputation randomization approaches to the distribution of FTE earnings, on the sample where both partners work. This raises the same concerns as previously discussed. The third approach is to use the model of section 4 in order to parametrically identify the joint distribution of partners' potential earnings among observed couples. Under the assumption of joint-log normality, this distribution is characterized by three parameters: the variance of earnings in the marginal earnings distribution of female and male and the covariance of earnings within the couple. The estimated parameters can be used to compute the degree of inequality in the distribution of household potential earnings, although potential earnings are a latent, unobserved variable for some couples where one of the partners is out of employment. Furthermore, once the parameters of this joint distribution have been estimated, it is easy to simulate the distribution of household potential earnings under the assumption that the correlation of partners' potential earnings is zero, holding constant the characteristics of the marginal distributions.

Regardless of the specific method used to construct the counterfactual earnings distribution, an additional issue arises regarding whether the randomization process should operate on the overall population or within age groups. As previously discussed, part of the correlation of economic outcomes within couples is driven by the fact that partners are homogenous in terms of birth cohort. This cohort-wise homogamy would likely survive even if partner's choice was independent of individual social and economic characteristics. For this reason, one may suggest that the randomization process used to build the counterfactual should occur conditional on the age of partners. In the rest of the analysis, we follow this assumption and only allow rematching to occur conditional on the age of both partners.

Last, one should also mention that none of the three above approaches takes into consideration the changes in the distribution of earnings and wage rates. Such changes

---

[18]This is true, at least, in the short term. In the long run, due to the accumulation of experience and seniority, potential earnings also depend on past labor supply decisions. We do not account for this source of endogeneity here.

could indeed result from general equilibrium effects driven by changes in the composition of households and in their labor supply decisions. They are however rarely taken into consideration in such counterfactual decompositions of inequality.

The three randomization algorithms are described in Appendix B.

5.2 Results

Our estimates of the effect of assortative mating on earnings inequality are given in table 8. For the observed and simulated earnings distributions we compute standard inequality indices (Gini, Theil, Atkinson and P90/P10). We also report the variation of the inequality indices between the actual distribution and the counterfactual distribution, which indicates the inequality reduction obtained by randomizing mating patterns among couples.

Annual earnings Panel A reports the results for addition randomization. Inequality in the actual distribution, for instance the Gini coefficient of 0.27, is slightly lower than the degree of inequality in the overall distribution of earnings in France. This reflects the greater homogeneity of our sample, compared to the overall population, induced by our sample selection rules.[19] The equalizing effect of randomizing individual annual earnings across couples, conditional on age, appears relatively modest. The Gini index falls by 8.5%. The effect on the other inequality measures is larger : the Theil and Atkinson indices fall by about 17-18%. Of course one of the difficulties of this approach is that it fails to take into account the labor supply responses that would occur if individuals were randomized into less homogenous couples. These labor supply responses would be likely to occur, especially in the case of female. However, the consequence of these labor supply adjustments for overall earnings inequality is a priori unclear.

Panel B provides actual and counterfactual inequality measures for the imputation randomization procedure. The effect of randomizing educational attainment across couples (conditional on age) is smaller than in Panel A. The Gini falls by 2.8% is in line with the results reported in Eika, Mogstad, and Zafar (2017), Greenwood, Guner, Kocharkov, and Santos (2014) and Harmenberg (2014) who also report a modest contribution of assortative mating to inequality between couples. However, the effect on the other inequality indices is significantly larger, especially for the Atkinson(2), which falls by about 8.6%. Though one

---

[19]Excluding single-headed households will, in particular, drive down inequality measures.

of the advantages of the imputation randomization approach is to allow for labor supply responses, one obvious limitation of this approach is to rule out selection on unobservable characteristics and to assume that heterogamous couples are a good counterfactual for the behavior of individuals observed in homogamous couples if these individuals were rematched with more heterogeneous partners. Unfortunately, it is hard to guess how selection on unobservable characteristics would bias the counterfactual experiment.

FTE earnings Panels A, B and C also provide evaluations of the effect of assortative mating on inequality in FTE earnings. First, one should stress that using FTE earnings as the variable of interest reduces inequality in the distribution, by reducing heterogeneity across individuals arising from differences in labor supply. This explains the relatively low observed value of the inequality measures.

Overall the results indicate a larger contribution of assortative mating to potential earnings inequality than for annual earnings. The simulations conducted under addition randomization (Panel A), predict a sizable fall in inequality as a result of random rematching. The Gini coefficient would fall by 10.1% and the Theil index by 21.3%. Unlike the results obtained for annual earnings, imputation randomization also indicates a sizable effect of assortative mating on FTE earnings inequality. For instance, imputation randomization predicts a fall in the Gini of 8.3% (against only 2.8% for annual earnings) and a fall in the Theil index of 17.4%. Controlling for sample selection in Panel C provides consistent results on the disequalizing effect of mating patterns. Under random matching, the Gini coefficient would fall by 8.7% and the Theil index would fall by about 16.6%.

In summary, the three approaches to randomization produce similar and consistent results in the case of FTE earnings. They all point to a sizable contribution of assortative mating to earnings inequality. The effect is also much higher than the one observed for annual earnings. Two conclusions can be drawn from these results. First, the effect of assortative mating on annual earnings inequality seems to be partially mitigated by endogenous labor supply decision. Second, the small contribution of assortative mating to annual earnings inequality may mask a greater contribution to overall inequality across households. In this respect, FTE earnings provide a broader measure of the resources available to the household and might be more relevant to assess the consequences of mating decisions on inequality.

6 Concluding comments

In this paper, we evaluated the extent of assortative mating in France and its contribution to inequality between couples. Our estimates reveal a large statistical association in socioeconomic characteristics among partners. The correlation coefficient for years of education is high, around 0.6. Similar results are found for occupation. For annual earnings, the correlation appears much weaker, around 0.17, when computed on all individuals, including those with zero earnings. Although this value seems low, especially when compared to the correlation in other socio-economic characteristics, one should emphasize that it is markedly higher than the one found for other developed countries, in particular the US. The correlation of full-time equivalent earnings, computed on the sample of couples in which both partners are salaried, is also markedly higher than for annual earnings: this correlation is around 0.35 for yearly measures of FTE earnings and raises up to 0.49 when using multi-year averages. All in all, this points to a fairly large degree of assortative mating among French couples.

This high degree of homogamy is consistent with the picture of a highly stratified French society. For instance, Lefranc and Trannoy (2005) and Lefranc (2011) report that the degree of intergenerational earnings persistence in France is relatively high compared to other developed economies. Lecavelier and Lefranc (2015) estimates statistical association in education and earnings among siblings. Their findings indicate a high correlation in socio-economic outcomes among siblings. Interestingly, they report values of the intra-siblings correlation in education and earnings that are very similar to the value of the within-couple correlations found here. This implies that the degree homogeneity within couples is similar to the degree of homogeneity within family among siblings. In other words, from the perspective of inequality among couples, patterns of assortative mating are equivalent to a process in which individuals would randomly select their mates... from their family of origin. Chadwick and Solon (2002) and Ermisch, Francesconi, and Siedler (2006) report consistent evidence.

Economic assortative mating might not simply result from the effect of social stratification but also arises from economic determinants. Of course, economic assortative mating is expected to occur as a result of marital sorting along non-economic dimensions such as social origin or educational choice. However, our results indicate that partners' earnings remain significantly correlated, even after controlling for educational choice or family

background. This is consistent with the view that economic considerations might be an important factor in determining partner's choice. Fremeaux (2014) provides similar evidence.

Our results also allow assessing the contribution of assortative mating to earnings inequality among couples. Several papers have recently addressed this issue but no clear picture has emerged regarding the disequalizing effect of homogamy. This lack of consensus partly reflects the use of different methodologies for assessing the counterfactual distribution of earnings that would prevail under random mating. As a matter of fact, current approaches fail to fully account for the endogeneity of labor supply decisions and for assortative mating along unobserved individual characteristics. Our results indicate that assortative mating has a sizable contribution to earnings inequality. Specifically, the Gini coefficient in earnings would fall by 2 points under random mating. This fall is of the same order of magnitude as the reduction in inequality that arises from income tax redistribution in France.[20] These results are based an alternative approach that focuses on couples' potential earnings. The potential earnings are defined as the earnings a couple would receive if both partners worked full-time, given their idiosyncratic market wage rate, and are measured by the sum of the full-time equivalent earnings of both partners. Our approach also accounts for assortativeness in unobservable earnings determinants. We show that assortative mating tends to increase inequality among couples, compared to random mating. For annual earnings, the effect is moderate and accounts for 4 to 10% of measured inequality. The effect of assortative mating is however much larger when focusing on couples' potential earnings and amounts to 10 to 20% for observed inequality. The effect of assortative mating is found to be larger for inequality indices more sensitive to the tails of the distribution.

The discrepancy between the two estimates suggests that labor supply decisions tend to dampen the effect of marital sorting on inequality in labor earnings across couples and partly masks wider inequality in household resources and welfare. Labor supply decisions and their relationship with marital sorting should be investigated further. The extent of marital sorting along preferences for work and employability should be evaluated. Future research should also examine the interplay between assortative mating and fiscal policy. This issue is seldom addressed with the exception of Pestel (2017). More specifically, the

---

[20]See Immervoll, Levy, Lietz, Mantovani, O'Donoghue, Sutherland, and Verbist (2005).

design of couples' income taxation strongly influences the partners' labor supply decisions. While individual taxation encourages labor market participation, joint taxation encourages specialisation within the household since the marginal tax rate of the secondary earner depends on that of the primary earner (Crossley and Jeon 2007). A majority of rich countries has implemented an individual income tax scheme (Care 2014). However, in France, taxation occurs at the household level. Given the observed hump-shaped female labor market participation, one could expect that the effect of individual taxation on female labor supply should increase the contribution of assortative mating to inequality. Future research should address this issue.

References

AHMAD, N., AND S.-H. KOH (2011): "Incorporating Estimates of Household Production of Non-Market Services into International Comparisons of Material Well-Being," OECD Statistics Working Papers 2011/7, OECD Publishing.

ARRONDEL, L., AND N. FREMEAUX (2016): "For richer, for poorer: assortative mating and savings preferences," *Economica*.

BECKER, G., AND N. TOMES (1979): "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87(6), 1153–1189.

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

BLACK, S., AND P. J. DEVEREUX (2011): *Recent Developments in Intergenerational Mobility*. Elsevier.

BLOSSFELD, H.-P., AND TIMM (2003): *Who Marries Whom?: Educational Systems as Marriage Markets in Modern Societies*. Springer.

BOLLINGER, C. R., AND A. CHANDRA (2005): "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labor Economics*, 23(2), 235–258.

BOUCHET-VALAT, M. (2014): "Les évolutions de l'homogamie de diplome, de classe et d'origine sociales en France (1969-2011) : ouverture d'ensemble, repli des élites," *Revue française de sociologie*, 55(3), 459–505.

BURTLESS, G. (1999): "Effects of growing wage disparities and changing family composition on the U.S. income distribution," *European Economic Review*, 43, 853–865.

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.

CARE (2014): "The taxation of families - International comparisons 2012," *Care research paper*.

CHADWICK, L., AND G. SOLON (2002): "Intergenerational income mobility among daughters," *American Economic Review*, 92(1), 335–344.

CROSSLEY, T. F., AND S.-H. JEON (2007): "Joint Taxation and the Labour Supply of Married Women: Evidence from the Canadian Tax Reform of 1988," *Fiscal Studies*, (28), 343–365.

DOHMEN, T., A. FALK, D. HUFFMAN, AND U. SUNDE (2012): "The Intergenerational Transmission of Risk and Trust Attitudes," *Review of Economic Studies*, 79(2), 645–677.

DUNCAN, G. J., AND D. H. HILL (1989): "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics," *Journal of Business and Economic Statistics*, 7(4), 441–52.

ECKSTEIN, Z., AND O. LIFSHITZ (2011): "Dynamic Female Labor Supply," *Econometrica*, 79(6), 1675–1726.

Eika, L., M. Mogstad, and B. Zafar (2017): "Educational Assortative Mating and Household Income Inequality," Federal Reserve Bank of New-York Staff Reports 692, Federal Reserve Bank of New-York.

Ermisch, J., M. Francesconi, and T. Siedler (2006): "Intergenerational Mobility and Marital Sorting," *Economic Journal*, 116(513), 659–679.

Fernandez, R., N. Guner, and J. Knowles (2005): "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality," *Quarterly Journal of Economics*, 120(1), 273–344.

Frazis, H., and J. Stewart (2011): "How does household production affect measured income inequality?," *Journal of Population Economics*, 24(1), 3–22.

Fremeaux, N. (2014): "The Role of Inheritance and Labour Income in Marital Choices," *Population-E*, 69(4), 495–530.

Goux, D., and E. Maurin (2003): "Who Marries Whom in France. An Analysis of the Cohorts born between 1934 and 1978," in *Who Marries Whom?*, ed. by H. Blossfeld, and Y. Shavit, chap. 4. Oxford University Press, Oxford.

Greenwood, J., N. Guner, G. Kocharkov, and C. Santos (2014): "Corrigendum to Marry Your Like: Assortative Mating and Income Inequality," *American Economic Review - Papers and Proceedings*, 104(5), 348–353.

Gronau, R. (1977): "Leisure, Home Production, and Work-The Theory of the Allocation of Time Revisited," *Journal of Political Economy*, 85(6), 1099–1123.

Hagneré, C., and A. Lefranc (2006): "Etendue et conséquences des erreurs de mesure dans les données individuelles d'enquête: une évaluation à partir des données appariées des enquêtes Emploi et Revenus Fiscaux," *Economie et Prévision*, 174(3), 131–154.

Haider, S., and G. Solon (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96(4), 1308–1320.

Ham, J. C. (1982): "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment," *The Review of Economic Studies*, 49(3), 335–354.

Harmenberg, K. (2014): "A note: the effect of assortative mating on income inequality," *Mimeo*.

Heckman, J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

House, C., J. Laitner, and D. Stolyarov (2008): "Valuing Lost Home Production Of Dual Earner Couples," *International Economic Review*, 49(2), 701–736.

Hryshko, D., C. Juhn, and K. McCue (2014): "Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important Is Assortative Matching?," IZA Discussion Papers 8729, Institute for the Study of Labor (IZA).

Immervoll, H., H. Levy, C. Lietz, D. Mantovani, C. O'Donoghue, H. Sutherland, and G. Verbist (2005): "Household Incomes and Redistribution in the European Union: Quantifying the Equalising Properties of Taxes and Benefits," IZA Discussion Papers 1824, Institute for the Study of Labor (IZA).

Kalmijn, M. (1991): "Status Homogamy in the United States," *American Journal of Sociology*, 97, 496–523.

Karoly, L. A., and G. Burtless (1995): "Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well-Being, 1959-1989," *Demography*, 32(3), 379–415.

Kimball, M., C. R. Sahm, and M. D. Shapiro (2009): "Risk Preferences in the PSID: Individual Imputations and Family Covariation," *American Economic Review - Papers and Proceedings*, 99(2), 363–368.

Lecavelier, C., and A. Lefranc (2015): "Siblings correlations in socio-economic outcomes in France," mimeo thema.

Lefranc, A. (2011): "Educational expansion, earnings compression and changes in intergenerational economic mobility : Evidence from French cohorts, 1931-1976," THEMA Working Papers 2011-11, THEMA (THéorie Economique, Modélisation et Applications), Université de Cergy-Pontoise.

Lefranc, A., N. Pistolesi, and A. Trannoy (2009): "Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France (1979-2000)," *Journal of Public Economics*, 93(11-12), 1189–1207.

Lefranc, A., and A. Trannoy (2005): "Intergenerational earnings mobility in France: Is France more mobile than the U.S.?," *Annales d'Economie et de Statistique*, (78), 57–77.

Lise, J., and S. Seitz (2011): "Consumption Inequality and Intra-Household Allocations," *Review of Economic Studies*, 78(1), 328–355.

Mare, R. D. (1991): "Five Decades of Educational Assortative Mating," *American Sociological Review*, 56(1), 15–32.

Moffitt, R., and P. Gottschalk (2011): "Trends in the covariance structure of earnings in the U.S.: 1969-1987," *Journal of Economic Inequality*, 9(3), 439–459.

Nakosteen, R. A., O. Westerlund, and M. A. Zimmer (2004): "Marital Matching and Earnings: Evidence from the Unmarried Population in Sweden," *Journal of Human Resources*, 39(4), 1033–1044.

Ostrovsky, Y. (2012): "The correlation of spouses' permanent and transitory earnings and family earnings inequality in Canada," *Labour Economics*, 19(5), 756–768.

Pearson, K. (1900): "Mathematical Contributions to the Theory of Evolution. VII. On the Correlation of Characters not Quantitatively Measurable," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 195, pp. 1–47+405.

Pestel, N. (2017): "Marital sorting, inequality and the role of female labor supply: Evidence from East and West Germany," *Economica*, 84(333), 104–127.

Ritchie-Scott, A. (1918): "The correlation coefficient of a polychoric table," *Biometrika*, 12(1/2), 93–133.

Roy, D. (2012): "Le travail domestique : 60 milliards d'heures en 2010," *INSEE Premiere*, 1423.

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.

SOLON, G. (1992): "Intergenerational Income Mobility in the United States," *American Economic Review*, 82(3), 393–408.

UUNK, W. J. G., H. B. G. GANZEBOOM, AND P. RÓBERT (1996): "Bivariate and Multivariate Scaled Association Models. An Application to Homogamy of Social Origin and Education in Hungary between 1930 and 1979," *Quality & Quantity*, 30, 323–343.

ZHANG, J., AND P.-W. LIU (2003): "Testing Becker's Prediction on Assortative Mating on Spouses' Wages," *The Journal of Human Resources*, 38(1), pp. 99–110.

ZIMMERMAN, D. (1992): "Regression Toward Mediocrity in Economic Stature," *American Economic Review*, 82(1), 409–429.

Table 1: Correlations - occupation and education

	(1)	(2)	(3)	(4)	(5)
	A - Occupation				
	own occ.	father's occ.	own occ.	mother's occ.	own occ.
Spearman	.453	.29	.456	.249	.468
	[.434,.471]	[.254,.325]	[.424,.486]	[.203,.294]	[.43,.505]
Polychoric	.531	.377	.526	.308	.542
	(.01)	(.021)	(.017)	(.026)	(.021)
Obs.	6928	2559	2559	1635	1635
	B - Highest degree				
	own degree	father's degree	own degree	mother's degree	own degree
Spearman	.559	.437	.58	.401	.571
	[.543,.574]	[.403,.47]	[.551,.607]	[.368,.433]	[.545,.597]
Polychoric	.593	.506	.613	.476	.606
	(.01)	(.021)	(.017)	(.026)	(.021)
Obs.	7864	2202	2202	2571	2571
	C - Years of education				
	years				
Pearson	.62				
	[.606,.633]				
Spearman	.624				
	[.611,.638]				
Obs.	7864				

Note: 95% confidence interval in square brackets; standard-errors in parenthesis. Estimates in panels A and B, columns 2 to 5, are restricted to the sample of couples for which information on own and parental occupation (resp. degree) is available.

Table 2: Correlations - labor earnings

	(1) $w0$	(2) w	(3) w^{FTE}
A- Gross correlations			
Pearson	.175	.31	.351
	[.153,.196]	[.286,.332]	[.328,.373]
Spearman	.179	.269	.316
	[.157,.2]	[.245,.292]	[.293,.338]
B- Net of Cohort effects			
Pearson	.169	.296	.337
	[.147,.19]	[.272,.319]	[.314,.359]
Spearman	.175	.261	.318
	[.153,.196]	[.238,.285]	[.295,.34]
Obs.	7864	5983	5983

Note: $w0$: annual labor earnings, including zeroes; w : annual labor earnings, excluding zeroes; w^{FTE} full-time equivalent annual labor earnings, excluding zeroes. 95% confidence interval in square brackets.

Table 3: Correlations - labor earnings residuals

	(1) w	(2) w^{FTE}
A- Gross correlations		
Pearson	.295	.341
	[.241,.347]	[.289,.391]
Spearman	.26	.296
	[.205,.313]	[.242,.348]
B- Conditional on education		
Pearson	.178	.214
	[.122,.234]	[.158,.269]
Spearman	.163	.185
	[.106,.219]	[.128,.24]
C- Conditional on social origin		
Pearson	.203	.246
	[.147,.258]	[.191,.3]
Spearman	.159	.204
	[.102,.215]	[.148,.259]
D- Conditional on education and social origin		
Pearson	.172	.206
	[.115,.227]	[.15,.261]
Spearman	.161	.194
	[.104,.217]	[.138,.249]
Obs.	1764	1764

Note: w : annual labor earnings, excluding zeroes; w^{FTE} full-time equivalent annual labor earnings, excluding zeroes. 95% confidence interval in square brackets.

Table 4: Correlations - multi-year average of labor earnings

	(1) $w0$	(2) mean $w0$	(3) w	(4) mean w	(5) w^{FTE}	(6) mean w^{FTE}
			A- Couples observed at least 3 years			
Pearson	.192	.222	.336	.39	.389	.466
	[.163,.221]	[.193,.25]	[.304,.367]	[.36,.419]	[.359,.418]	[.438,.493]
Spearman	.197	.218	.292	.34	.342	.421
	[.168,.226]	[.189,.246]	[.26,.324]	[.309,.371]	[.311,.373]	[.392,.45]
Pearson (residuals)	.186	.215	.325	.377	.382	.454
	[.157,.215]	[.186,.244]	[.293,.356]	[.347,.407]	[.351,.411]	[.425,.481]
Spearman (residuals)	.19	.217	.284	.337	.35	.423
	[.161,.219]	[.188,.245]	[.252,.316]	[.305,.368]	[.319,.381]	[.394,.451]
Obs.	4258	4258	3106	3106	3106	3106
			B- Couples observed at least 5 years			
Pearson	.235	.257	.367	.416	.385	.49
	[.189,.28]	[.212,.301]	[.317,.415]	[.368,.462]	[.335,.432]	[.446,.532]
Spearman	.221	.248	.327	.374	.35	.435
	[.175,.266]	[.202,.292]	[.275,.377]	[.324,.422]	[.299,.399]	[.388,.48]
Pearson (residuals)	.23	.251	.361	.407	.384	.483
	[.184,.275]	[.205,.295]	[.311,.41]	[.358,.453]	[.335,.432]	[.438,.525]
Spearman (residuals)	.213	.246	.324	.374	.365	.441
	[.166,.258]	[.201,.29]	[.272,.374]	[.324,.422]	[.315,.413]	[.394,.485]
Obs.	1677	1677	1185	1185	1185	1185

Note:: $w0$: annual labor earnings, including zeroes; w : annual labor earnings, excluding zeroes; w^{FTE} full-time equivalent annual labor earnings, excluding zeroes. 'mean' indicates the multi-year averages, computed over all years for which the information is available. 95% confidence interval in square brackets.

Figure 1: Bivariate density - Annual earnings

A- earnings levels
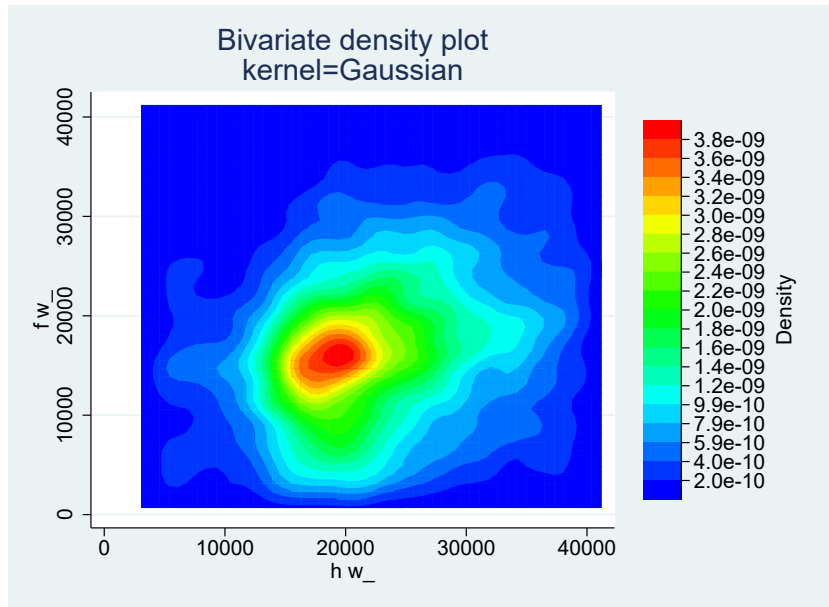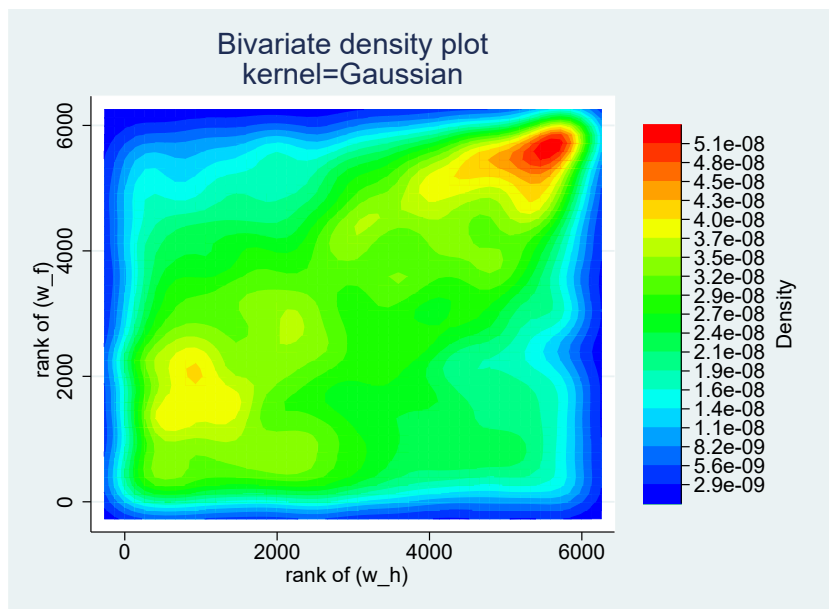
B- earnings ranks

Figure 2: Bivariate density - Full-time equivalent earnings

A- earnings levels
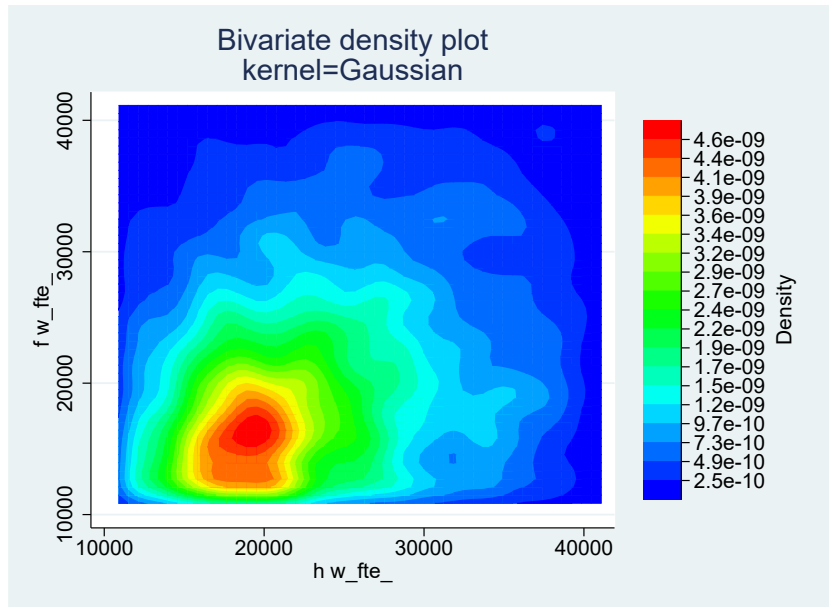
B- earnings ranks
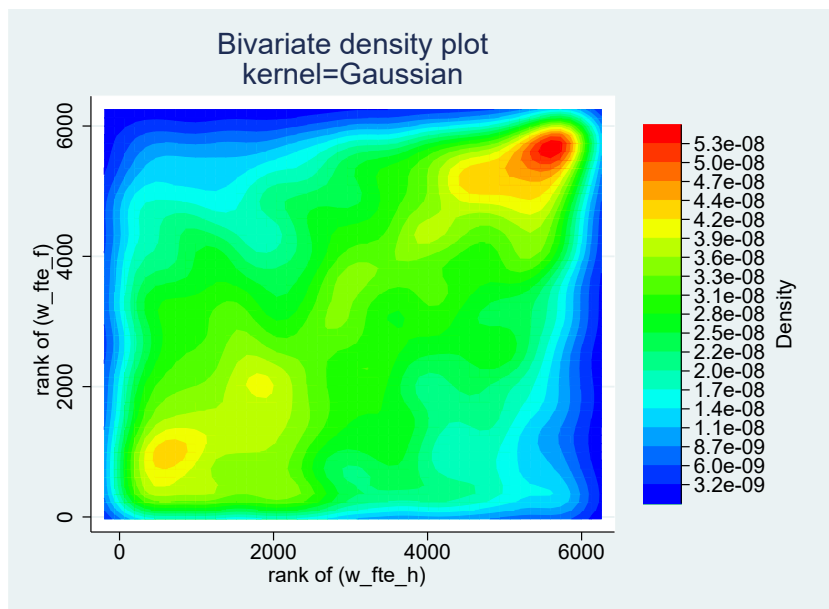
Table 5: Correlations and sample selection- labor earnings

	(1) $\ln w^{FTE}$	(2) $\ln(\text{mean } w^{FTE})$
Panel A - Ignoring sample selection		
β_{OLS}	.329	.359
ρ	.326	.361
σ_m	.407	.396
σ_f	.411	.395
N	5983	6383
Panel B - Accounting for sample selection		
$\beta_{Heckman}$	.321	.357
ρ	.31	.353
σ_m	.421	.409
σ_f	.436	.414
ρ_{res}	-.619	-.606
N	7526	7526

Note: β: regression coefficient; σ: standard deviation (for the male partner m and the female partner f); ρ: correlation coefficient; ρ_{res}: correlation coefficient of the error terms of the selection and wage equations. w^{FTE} full-time equivalent annual labor earnings, excluding zeroes. 'mean' indicates the multi-year averages, computed over all years for which the information is available.

Table 6: Sample selection model - labor earnings

Female wage	(1) $\ln w_f^{FTE}$	(2) $\ln(\text{mean } w_f^{FTE})$
Main equation		
$\ln w_m^{FTE}$	.321	.357
	(.0127)	(.012)
cons	.0802	.0606
	(.0063)	(.0055)
Selection equation		
w_m	7.0e-06	6.1e-06
	(4.0e-06)	(4.2e-06)
w_m^2	-.0219	-.0234
	(.0045)	(.0047)
age_m	.0027	-.0037
	(.0047)	(.005)
age_m^2	-6.0e-04	-6.1e-04
	(3.1e-04)	(3.3e-04)
age_f	.0209	.0215
	(.0043)	(.0046)
age_f^2	-.0019	-.0022
	(3.2e-04)	(3.3e-04)
years of education$_f$	.188	.167
	(.0494)	(.052)
years of education$_f^2$	2.6e-04	.0013
	(.0019)	(.002)
number of children	-.253	-.225
	(.017)	(.0177)
long-term contract$_m$	.157	.169
	(.0499)	(.0527)
capital income	4.3e-06	4.8e-06
	(2.3e-06)	(2.4e-06)
ρ_{res}	-.619	-.606
	(.038)	(.038)
σ_{ε}	.414	.387
	(.00467)	(.00409)

Note: Standard errors in parenthesis. w^{FTE} full-time equivalent annual labor earnings, excluding zeroes. Indices m for the male partner and f for the female partner.

Table 7: Female labor market characteristics conditional on male earnings deciles

	(1) Work	(2) Unemp.	(3) Inactivity	(3) Months worked	(4) w	(5) w^{FTE}
Male FTE earnings :						
D1	0.65	0.093	0.26	9.3	14,740	20,118
D2	0.74	0.07	0.19	9.6	15,405	20,008
D3	0.75	0.069	0.18	9.6	15,694	20,142
D4	0.76	0.074	0.16	9.8	16,062	20,425
D5	0.78	0.051	0.17	9.6	16,813	21,718
D6	0.81	0.046	0.14	9.6	17,968	23,442
D7	0.81	0.048	0.14	9.8	19,178	24,460
D8	0.8	0.049	0.15	9.7	19,975	25,674
D9	0.8	0.058	0.14	9.6	21,296	27,782
D10	0.72	0.069	0.21	9.3	24,868	33,458

Note: D1 (resp. D10) refers to the bottom (resp. top) decile of the male FTE distributions. w and w^{FTE} are expressed in 2011 Euros.

Table 8: Earnings inequality - Observed and simulated matching

	(1) Gini	(2) Theil	(3) A(1)	(4) A(2)	(5) p90/p10
A - Addition randomization					
Annual earnings					
Observed	0.270	0.121	0.124	0.268	3.722
Simulated	0.247	0.099	0.103	0.220	3.332
Δ inequality	-8.5%	-17.8%	-17.5%	-17.7%	-10.5%
FTE earnings					
Observed	0.207	0.072	0.065	0.117	2.453
Simulated	0.186	0.057	0.053	0.098	2.298
Δ inequality	-10.1%	-21.3%	-18.7%	-16.2%	-6.3%
B - Imputation randomization					
Annual earnings					
Observed	0.270	0.121	0.124	0.268	3.722
Simulated	0.263	0.114	0.116	0.245	3.515
Δ inequality	-2.8%	-5.7%	-7.0%	-8.6%	-5.6%
FTE earnings					
Observed	0.207	0.072	0.065	0.117	2.453
Simulated	0.190	0.060	0.056	0.106	2.325
Δ inequality	-8.3%	-17.4%	-13.6%	-9.3%	-5.2%
C - Addition randomization with sample selection correction					
FTE earnings					
Observed	0.196	0.062	0.060	0.116	2.474
Simulated	0.179	0.051	0.050	0.097	2.283
Δ inequality	-8.7%	-16.6%	-16.6%	-16.5%	-7.7%

Note: A(1) and A(2) denote the Atkinson inequality indices with coefficient 1 and 2 respectively; p90/p10 denotes the ratio of the ratio of the 90th percentile over the 10th percentile.

Appendix A   Main variables and descriptive statistics

The individual characteristics examined in this paper are the following.

Earnings   *Annual earnings* are defined as the total wage and salaries earned in the previous year deflated by the consumer price index. For individuals out of salaried employment, the value of annual earnings is equal to zero[21]. Earnings are self-reported from 2004 to 2007 and matched with fiscal and administrative data afterwards. Preliminary analysis suggests that self-reported earnings incorporate significant measurement error, with important consequences on the estimation of earnings correlations. Without corrections, the correlation of partners' earnings is about 25% lower for self-reported earnings than for administrative earnings. To minimize the incidence of measurement errors in earnings on our estimates of assortativeness, we winsorize the data by recoding the bottom and the top 1% of the earnings distribution when earnings are positive. When reported earnings are equal to zero, this value is kept unchanged. Bollinger and Chandra (2005) show that winsorizing performs better than trimming in the presence of response errors. After winsorizing, estimates based on self-reported earnings appear similar to those derived from administrative data.

Full-time equivalent (FTE) earnings are defined as annual earnings/(number of months worked full-time + 0.5 × number of months worked part-time) × 12. To compute FTE earnings, we rely on the history of labor force participation reported in the survey. For each month in the preceding year, individuals are asked to report their labor force status, which distinguishes between full-time and part-time salaried work. Unfortunately, for individuals working part-time we do not observe the share of working time. We thus assume that part-time work corresponds to 50% of full working time.[22] For individuals out of salaried work, FTE earnings are missing, by construction. We apply the same winsorizing procedure to FTE earnings, as described above.

For both earnings measures, we compute multi-year averages of individual earnings. This average is computed over the full set of available yearly observations. The number of years of observation in our sample varies between 1 and 8 years, with an average of 3.4 years.

Educational attainment   We use two measures of educational attainment. The first one is the number of years of education, equal to the reported school leaving age minus 6 years (i.e. minimum age for compulsory education)[23]. Our second variable is based on the highest degree completed. We consider a classification with 8 ordered levels : 1) no degree; 2) general lower secondary degree ; 3) vocational lower degree; 4) vocational upper secondary degree ; 5) general upper secondary degree; 6) college (bachelor or technical degree); 7) master's degree 8) PhD or elite schools degree (*Grandes Ecoles*).

Occupation   Our measure of occupation is based on the standard 6-levels French classification. In order to come close to an ordinal measure of occupation, we gather farmers and unskilled manual workers. This leads to the following classification: 1) Higher-grade professionals; 2) Lower-grade professionals; 3) Artisans and small proprietors; 4) Non-manual employees; 5) Farmers and manual workers. Respondents report their current or last occupation (in case of unemployment). The information is missing for individuals out of the labor force.

---

[21]Given our use of panel data, the individuals with zero earnings should have never reported any salaried activity. Some of these individuals may however report unemployment period and so potentially unemployment benefits. Taking into account these benefits (as a proxy for earnings) does not change our estimates but it increases the measurement errors. In the end, we decided not to include them.

[22]Information on hours of work is only available at the time of the interview. In our sample, 65% of part-time salaried individuals report working between 15 and 30 hours per week.

[23]For some individuals, the number of years of education appears noisy. Furthermore, although highest degree is reported for all individuals in the sample, number of years of education is missing for 9% of the sample. For this reason we estimate the correlation in *predicted* number of years of education, where the prediction is based on a regression of number of years of education on degree dummies interacted with gender and a fourth degree polynomial function of birth cohorts.

Socioeconomic origin The SILC survey investigated individual socioeconomic origin and gathered information on education and occupation of both parents of adult respondents. Information is only available for a sub-sample of our data, since the questionnaire only investigated this topic in the 2005 wave. Our measure of parental occupation uses the same classification as individual occupation (see above). Occupation is missing when the parent was continuously out of the labor force during the respondent's youth. Our measure of education is based on the highest degree completed by the parents. The classification is the same as described above.

Table A.1: General descriptive statistics

	Men	Women
Age	42	40
Education		
No degree	0.12	0.12
General lower secondary degree	0.094	0.12
Vocational lower degree	0.34	0.26
Vocational upper degree	0.056	0.054
General upper degree	0.093	0.12
College - bachelor degree	0.11	0.13
Master's degree	0.082	0.12
PhD or elite schools degree	0.099	0.079
Labor market status		
Employment	0.94	0.77
Unemployment	0.047	0.062
Inactivity	0.015	0.17
Number of months worked	11	7.7
Earnings		
Share of individuals with $w > 0$	0.96	0.8
w_0 (mean)	25,079	14,581
w_0 (std error)	14,815	11,199
w (mean)	26,206	18,141
w (std error)	14,135	9,563
w^{FTE} (mean)	27,400	23,648
w^{FTE} (std error)	14,333	11,625
N	7,864	7,864

Appendix B   Simulation algorithms

B.1   Addition randomization

The addition randomization algorithm randomizes *individual* earnings within couples. Randomization is only allowed to occur given the age of both partners in the couple. Randomization relies on a parametric model of labor force participation and a semi-parametric earnings regression model. For all couples observed in the sample, the main steps of the earnings addition randomization are the following :

1. Estimate a probit model of male labor market status (0 for no earnings in the previous year; 1 for strictly positive earnings) where the probability of positive earnings is a function of a second order polynomial function of male age, female age and their interaction.

2. Estimate a linear regression model for joint earnings of the couple, on the sample of couples, where log-earnings are regressed on the number of years of education of male and female (second order polynomial), an interaction term in male and female education, a fourth order polynomial of male and female age and a second order polynomial interaction of male and female age. Store the distribution of predicted residuals.

3. Keep observations of female and male age and female labor earnings, including zeroes.

4. Randomize male labor market status by drawing from a Bernoulli distribution where the probability of positive earnings is predicted on the basis of the probit model of step 1.

5. When labor market status is 1, randomize earnings using the earnings model of step 2 : compute predicted log earnings conditional on age; randomly draw a value of the residual on the basis of the empirical distribution of predicted residuals; take the exponential of the sum of the previous two components.

B.2   Imputation randomization

The imputation randomization algorithm first randomizes education (number of years) among couples, conditional on the age of both partners. Second, it randomizes the couple's joint earnings, by randomly drawing from the observed earnings distribution of couples with similar age and education characteristics. Randomization is only allowed to occur given the age of both partners in the couple. Greenwood, Guner, Kocharkov, and Santos (2014) and Eika, Mogstad, and Zafar (2017) implement a non-parametric version of this randomization procedure. In our case, given limited sample size, randomization relies on a semi-parametric regression model of education and earnings. The steps of the imputation randomization are the following :

1. Estimate a linear regression model for years of education, on the sample of males, where years of education (in log) is regressed on a function of a second order polynomial function of male age, female age and their interaction.

2. Estimate a linear regression model for log earnings, on the sample of males with positive earnings, where log-earnings are regressed on a fourth order polynomial of male age. Store the distribution of predicted residuals.

3. Keep observations of female and male age and female years of education.

4. Randomize male number of years of education, conditional on the age of both partners, on the basis of the regression of step 1. The average number of years is predicted based on model's estimated coefficients; the residual is randomized by drawing from the distribution of predicted residuals.

5. Randomize couple's joint earnings using the earnings model of step 2: compute predicted log earnings conditional on age and education of both partners; randomly draw a value of the residual on the basis of the empirical distribution of predicted residuals; take the exponential of the sum of the previous two components.

B.3 Addition randomization with sample selection correction

Addition randomization with correction for sample selection is based on the model of section 4. Instead of estimating the model of section 4 on observed individual earnings, the model is estimated on earnings residuals computed from a preliminary regression in which earnings of both male and female are regressed on a fourth order polynomial in age. Conditional on the age of both partners, the algorithm randomizes the earnings residual based on the parametric joint log-normal model with sample selection. The steps of the addition randomization algorithm with correction for sample selection are the following :

1. Estimate a linear regression model for log FTE earnings of both male and female (separately), on the sample of individuals with positive earnings, where log-earnings are regressed on a fourth order polynomial of individual age. Store the distribution of predicted residuals and predicted values.

2. Estimate a sample selection model of female earnings residual following the model of section 4 to recover the correlation in residual earnings and the variance of female earnings without selection.

3. Keep observations of female and male age.

4. Compute predicted FTE earnings conditional on age for both male and female, using step 1.

5. Randomize male and female FTE earnings residuals by drawing residuals from a joint normal distribution with parameters estimated in step 2. This first simulation allows to derive the uncensored distribution of (latent) potential earnings in the population that corresponds to the observed degree of assortative mating.

6. Randomize male and female FTE earnings residuals by drawing residuals from a joint normal distribution with variances estimated in step 2 and covariance in residuals set equal to zero. This second simulation allows to derive the uncensored distribution of (latent) potential earnings in the population under the assumption of random mating.

Appendix C   Double selection

The selection model of section 4 assumes that endogenous sample selection only arises from the female labor participation and employment processes. Consistent with this assumption, the model is thus estimated on the subsample where male earnings is not missing. In our case, censoring is four times more prevalent for women than for men : in our sample of 7,966 couples, earnings are zero for 1,645 female partners against 440 male partners.[24] Selection issues may however also arise from the selection process that determines whether male earnings are observed and affect the estimations of the intra-household earnings correlation. We address this issue in this appendix.

We estimate a double-selection process where sample selection is allowed to be non-random due to both the observability of female earnings and that of male earnings. As in the model of section 4, the main equation of the model is given by :

$$w_f = \beta_0 + \beta w_m + \varepsilon \tag{2}$$

In the estimation of this equation, we account for the fact that both w_f and w_m may be zero. Define O_f (respectively O_m) a dummy variable indicating that w_f (resp. w_m) is non-zero. We assume that the process that determines the pair (O_f, O_m) is given by :

$$\begin{cases} O_f = \mathbb{1}\,(Z_f \gamma_f + \nu_f > 0) \\ O_m = \mathbb{1}\,(Z_m \gamma_m + \nu_m > 0) \end{cases} \tag{3}$$

where Z_f and Z_m are observable determinants of sample selection for, respectively, female and male wages and (ν_f, ν_m) is assumed to be a bivariate random normal vector.

Following Ham (1982), the model in equations 2 and 3 can be estimated by extending the two-stage procedure of Heckman (1979) to the two selection rule problem. This amounts to include two inverse Mills ratios in the estimation of equation 2, corresponding to the two selection processes of equation 3. As in the original Heckman two-stage procedure, the predicted inverse Mills ratios, $\hat{\lambda}_f$ and $\hat{\lambda}_m$, are derived from first-stage estimates of the selection rule, which is in the present case takes the form of a bivariate probit process.

The model is estimated on the full sample of couples that consists in 7,966 observations. 5,983 couples have non-zero earnings information for both partners. Variables included in the selection rule for female earnings are quadratic functions in female age and years of education, female self-assessed health, employment characteristics of the husband (indicators for non-zero earnings, unemployment and permanent job contract) and household characteristics (number of children, capital income, indicators for married couples, for living in rural areas, for the presence of a disabled household member). Variables included in the selection rule for male earnings are quadratic functions in male age and years of education, male self-assessed health, female characteristics (age and years of education) and household characteristics (number of children, capital income, indicators for married couples, for living in rural areas, for the presence of a disabled household member).[25]

Estimation results are reported in table C.1. Most variables in the bivariate selection probit model are highly significant. The correlation in the bivariate probit residuals is positive, around .3 and significant, indicating positive assortative mating in the unobserved determinants of reporting non-zero earnings. The wage regression model, accounting for sample selection, indicates a negative selection due to female earnings observability. However, the selection term for male earnings observability is very close to zero and not statistically significant.

Altogether, these results indicate that censorship due to female zero-earnings is not random. On the contrary, they support the assumption that censorship due to male zero-earnings can be ignored, as assumed in the model of section 4.

---

[24]102 couples have zero earnings for both partners and have been excluded from the estimations of sections 3 to 5.

[25]Variables that were not significant in the selection equations were omitted from the set of regressors.

	(1)	(2)
	Coef.	Std. Err.
wage equation		
dependent variable : w_f^{FTE}		
w_h^{FTE}	0.2925	0.0123
$\hat{\lambda}_f$	-0.4248	0.0253
$\hat{\lambda}_m$	0.0463	0.0604
bivariate probit selection model		
dependent variable : O_f		
age_f	0.0157	0.0025
age_f^2	-0.0028	0.0003
years of education$_f$	0.2530	0.0514
years of education$_f^2$	-0.0049	0.0019
number of children	-0.3023	0.0169
long-term contract$_m$	0.1097	0.0508
unemployed$_m$	-0.5622	0.0979
O_m	-1.2761	0.2143
capital income	-2.26e-06	-1.90e-06
married	-0.1896	0.0434
rural	0.1018	0.0382
health$_f$		
very good	REF	
good	0.0173	0.0414
fair	-0.1742	0.0532
bad	-0.4903	0.0845
very bad	-0.9913	0.2023
disabled	-0.2920	0.0656
dependent variable : O_m		
age_m	-0.0170	0.0068
age_m^2	-0.0010	0.0004
years of education$_m$	0.1273	0.0565
years of education$_m^2$	-0.0036	0.0021
age_f	0.0166	0.0062
age_f^2	-0.0006	0.0004
years of education$_f$	0.1895	0.0745
years of education$_f^2$	-0.0071	0.0028
married	0.1782	0.0596
rural	0.0995	0.0586
health$_m$		
very good	REF	
good	-0.0788	0.0655
fair	-0.1891	0.0817
bad	-1.1294	0.1001
very bad	-1.3293	0.2027
disabled	-0.5824	0.0708
ρ	0.3100	0.1106
Observations	total : 7966	O_f=1 : 6321	O_m=1 : 7526

Note: the f (resp. m) index denotes the female (resp. male) partner. ρ denotes the correlation of the error terms of the two probit processes.