

Assortative mating and earnings inequality in France ^{*}

Nicolas FRÉMEAUX (THEMA)[†]

Arnaud LEFRANC (THEMA)[‡]

September 2015

Abstract

This paper analyzes economic assortative mating in France. We first provide descriptive evidence on the statistical association in various economic attributes of spouses (education, earnings, market wage rate) among French couples. Second, we assess the contribution of assortative mating to earnings inequality between couples. Our estimates account for possible biases in the estimation of assortative mating arising from sample-selection into the labor force. We also develop a new methodology for assessing the disequalizing impact of marital choice when labor force participation is endogenous with respect to match characteristics. Our results indicate a strong degree of assortative mating in France. The correlation coefficient for education is above .5. The correlation in earnings is lower but sizable : around .2 for annual earnings and .35 for full-time equivalent earnings. Assortative mating tends to increase inequality among couples. The effect on the distribution of annual incomes remains moderate and explains 3 to 10% of measured inequality, depending on the counterfactual we use. The effect is however much larger for inequality in earnings potential and represents between 16% and 30 % of observed inequality.

JEL codes: J12, J22, D31

Key words: assortative mating, inequality, earnings, labor supply, France.

^{*}This research was developed under the auspices of the Labex MME-DII Center of Excellence (grant ANR11-LBX-0023-01), the support of which is gratefully acknowledged.

[†]THEMA - Université de Cergy-Pontoise, 33, boulevard du Port - 95011 Cergy-Pontoise cedex (France); nicolas.fremeaux [at] u-cergy.fr

[‡]Corresponding author. THEMA - Université de Cergy-Pontoise, 33, boulevard du Port - 95011 Cergy-Pontoise cedex (France); arnaud.lefranc [at] 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 issue¹, 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 twofold. We first provide descriptive evidence on the statistical association in various economic attributes of spouses (education, earnings, market wage rate) among French couples. Second, we assess the contribution of assortative mating to earnings inequality between couples.

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 analyzed². This represents an important 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

¹See in particular Karoly and Burtless (1995), Cancian and Reed (1998), Burtless (1999), Schwartz (2010), Eika, Mogstad, and Zafar (2014) Greenwood, Guner, Kocharkov, and Santos (2014), Harmenberg (2014), Pestel (2015)

²Arrondel and Fremeaux (2015), Dohmen, Falk, Huffman, and Sunde (2012) and Kimball, Sahm, and Shapiro (2009) are some of the few exceptions.

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 between spousal earnings.⁴ In this paper, exploiting panel data allows us to compute average earnings over multiple years 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 labor force. 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. We explicitly account for sample selection due to non-participation and provide estimates of the intra-couples correlation in (possibly latent) earnings potential.

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), 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 no consensus has yet emerged on the size of these effects. The main approach taken in this literature is to compare the observed earnings distribution to a counterfactual distribution built under alternative hypothetical mating patterns. However, the construction of this counterfactual

³Among the few exceptions are : Nakosteen, Westerlund, and Zimmer (2004) on Sweden, Pestel (2015) on Germany, Eika, Mogstad, and Zafar (2014) on Norway. 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.

⁴The 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).

distribution requires to 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 treats observed annual earnings as a fixed individual characteristics and simulate the distribution that would prevail if individuals kept their labor earnings unchanged and were randomly allocated 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 behavioral approach characterizes individuals by some observable earnings determinants, in general education (e.g. Greenwood, Guner, Kocharkov, and Santos 2014, Harmenberg 2014, Pestel 2015, Eika, Mogstad, and Zafar 2014). 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. 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. Our approach allows to characterize the effect of assortative mating on inequality in couples' *potential* earnings. Potential earnings are defined as the earnings a couple would receive, given the wage rate of its members, if each spouse worked full time. This provides a more extensive measure of the total economic resources commanded by the couple. Compared to existing studies, our approach offers three main advantages. First, focusing on potential earnings allows to account for the fact that individuals out of the labor force might have a positive contribution to their household's welfare through domestic production and leisure consumption. Second, it offers a consistent measure of the disequalizing impact of assortative mating that does not require to model labor supply decisions. Third, our method relies on a statistical model of the joint distribution of the earnings potential of both partners that allows for sample selection in the observed distribution and correlation across partners in their unobservable earnings determinants.

Our empirical analysis is based on the French waves of the EU-Statistics on Income

and Living Conditions, covering the period 2004-2011. Our results indicate a strong degree of assortative mating in France. The correlation coefficient for education is above .5. Correlation in earnings is lower but sizable : around .2 for annual earnings and .35 for full-time equivalent earnings. Sample-selection leads to a moderate upward bias on 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. The effect on the distribution of annual incomes remains moderate and explains 3 to 10% of measured inequality, depending on the counterfactual we use. The effect is however much larger for inequality in earnings potential and represents between 16% and 30 % of observed inequality.

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 French sample and use the waves 2004 to 2011. 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. As a consequence, the content of the surveys is harmonized across the 27 European countries. The French waves were collected by the French national statistics institute (INSEE).

Data are collected annually for a rotating panel of households. In the French samples, 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 spouse 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 also exclude couples in which both spouses or partners are out of the labor force. Last, to minimize the incidence of measurement errors in earnings on our estimates of assortativeness, we trim the distribution by excluding the bottom and the top 1% of the earnings distribution when earnings are positive⁵.

In the estimations reported below, we only keep one observation per couple. For each individual in a couple, we keep the non-missing observation 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). In the end, we have a sample of 8,259 couples.

2.2 Variables of interest

The characteristics of the spouse 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. Earnings are self-declared from 2004 to 2007 and matched with fiscal and administrative data afterwards. The value of annual earnings for individuals out of salaried employment is equal to zero. We also have information about the number of hours worked per week and the number of months worked full-time and part-time in the previous year. This allows us to build a measure of full-time equivalent (FTE) earnings. *FTE earnings* are defined as $\text{annual earnings} / (\text{number of months worked full-time} + 0.5 \times \text{number of months worked part-time}) \times 12$. Individuals out of paid-work are arbitrarily assigned a value of zero for FTE earnings. We explicitly account for sample selection due to non-participation in salaried work in section 4.

For both earnings measures, we compute multi-year averages of individual earnings. This average is computed over the full set of available year-observations. The number of years

⁵As discussed below, we rely on various measures of earnings (in level or log, single-year vs. multi-year average). Each distribution is trimmed separately at the extreme 1%.

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). Our second variable is based on the highest completed degree. 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 get 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 their last socioeconomic category (in case of unemployment). If people are inactive, the information is missing.

Socioeconomic origin In the 2005 wave, 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.

3 Descriptive measures of assortative mating

3.1 Education and occupation

We begin our analysis of assortative mating by focusing on variables widely used: occupation and education. The information is available for both partners of the couple, as well as their parents. For ordinal variables (occupation and highest degree completed), the association is measured using two indicators : the Spearman correlation coefficient mea-

sures 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. For the number of years of education, we report linear (Pearson) correlations and Spearman rank correlations.

Table 1 provides our estimates of assortative mating for occupation and education. Occupational correlations are given in panel A. The correlation of spouse's own occupation ranges between 0.45 and 0.50 (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 .3, which indicates positive assortative mating by social origin. Note though that the correlation in parental occupation is lower than the correlation in spouses' own occupation, which indicates that assortativeness depends more on individual occupational attainment than on social origin. The correlation is higher for fathers' occupations (0.28-0.37) than for mothers' (0.24-0.30). 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 spouses and their parents are consistent with existing evidence on French data based on contingency tables (Bouchet-Valat (2014)).

Panels B and C of Table 1 report statistical associations in education. Panel B uses the highest completed degree. On the whole sample, we find a positive correlation of around .5. The difference between the two measures of correlations (Spearman rank correlation vs. Polychoric correlation) is small. These correlations appear higher for education than for occupation. Correlation between partners is also higher for own education than for social origin, as captured by parent's education. However, compared to panel A, the differences between own and parental characteristics appear smaller for education than for the social

class.

Panel C provides correlation estimates for a continuous measure of education, the number of years of education. Although highest degree is reported for all individuals in the sample, number of years of education is missing for 9% of the sample. For some individuals, the number of years of education appears noisy. For this reason we also 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. Results for both measures (actual and predicted years) are given in the first two rows of panel C. They are consistent with those obtained for the correlation in degree completed, around .55. The imputation of missing values increases the correlation to around .6. Since the average number of years of education is positively correlated with birth cohorts, we investigate the contribution of age homogeneity to the measured association in education. To do this, we purge individual educational attainment from cohort effects using linear regression on a polynomial in birth cohort. In the last two rows of panel C, we report correlation measures for years of education, net of these cohort effects. Residual correlations are lower by about 8 points but remain large in the absolute.

Overall, our results indicate a high levels of positive assortative mating in France. These results are consistent with existing evidence on France (Goux and Maurin 2003, Bouchet-Valat 2014). From a comparative perspective, 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, around .6, 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 gauge economic assortative mating, we now examine the correlation between partners in annual and full-time equivalent (FTE) earnings. As before, we report Pearson linear correlation and Spearman correlation coefficients. To allow for non-linearity in the association between partners' earnings, we also report estimates of the Pearson coefficient for log earnings.

Results are presented in table 2. Column 1 reports correlations in annual earnings based on all observations, including observations where earnings are equal to 0. Column 2 excludes observations where reported earnings are equal to zero. The correlation between spouses' annual earnings (column 1) is slightly lower than 0.2. Focusing on the couples in which both partners are employed leads to an increase in the correlation. This should come as no surprise since we exclude unequal couples in which only one spouse (generally the male partner) declares positive earnings. In column 2, estimates of the correlation in log earnings are lower than estimates of the correlation in levels. Since the log transformation compresses earnings differentials at the top of the distribution, this indicates that the association at the top of the distribution is higher than at the bottom of the earnings distribution. We shall discuss this point further in a subsequent section.

In the last two columns of table 2, we examine the correlation in FTE earnings. This allows to focus on the correlation in earnings potential and remove the correlation (or lack thereof) in labor supply decisions within the couple that affects the correlation in annual earnings. Column 3 presents correlation estimates for the entire sample, where FTE earnings are set at 0 for individuals out of salaried work. Correlation estimates on this sample are close to those found for annual earnings. Column 4 excludes individuals with zero earnings from the estimation. This results in a much higher correlation, around .3. 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 20%.

Two main conclusions can be drawn from table 1. 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 spouses is out of the labor force) and a correlation slightly higher than 0.2 when both couples have positive labor earnings. These estimates are 55% and 22% higher, respectively, in France. Second, the table also indicates that labor supply decisions (both at the extensive or intensive margins) attenuate the correlations of earnings potential. 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.

As noted in the introduction, most papers focus on assortativeness by education or social origin. Both variables capture dimensions along which marital sorting should ob-

viously occur, given the role of the socialisation process in 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. In fact, one may object to the analysis of assortativeness by earnings that it merely reflects the correlation in spouse’s 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 and then for both education and social origin⁶. The main result of this table is that labor earnings remain positively correlated, even after controlling for individual educational attainment and social origin. The incidence of controlling for education and social origin varies with the measure of earnings we consider. In column 1, we focus on annual earnings and consider all observations, including those with zero earnings. Controlling for education reduces the correlation by around 25%. Further controlling for social origin reduces the coefficient by another 25%. The same pattern can be observed on the sub-sample with strictly positive earnings. However, when focusing on FTE earnings, controlling for education and social origin has a more limited effect on the correlation of earnings between spouses: the correlation only falls by about one third and remains sizable, around .2. As a conclusion, even if the assortativeness in terms of social background and of education is high, there is still significant sorting along economic dimensions or other unobserved individual characteristics that can be proxied by these economic measures.

Multi-year average earnings A potential challenge to the measurement of earnings correlation is the incidence of measurement errors and transitory income components. As discussed in the context of intergenerational mobility estimates (e.g. Solon 1992, Zimmerman 1992), 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 of 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 individ-

⁶We restrict the sample to the couples for whom the information about both education and social origin is available.

ual and each measure of earnings (annual vs. FTE, with or without zeroes), 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 variable horizons. We consider two sub-samples. In panel A, we estimate earnings correlations on the sample of couples observed during at least 3 waves; in panel B, we focus on couples who are observed during at least 5 waves.

Using multiple-year averages has a limited effect on our measure of the correlation in annual earnings. The linear correlation coefficient increases by 8% when using averages of annual earnings including zeroes and by 11% when zeroes are excluded. Using average earnings has a larger effect on the correlation in full-time equivalent earnings that increases by about 20% to reach a high value of .4.

While averaging earnings affects our measure of assortativeness in the expected direction, the size of the effect is lower than expected a priori. 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, which should 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. Comparing the value of the correlation coefficient for earnings in levels and in logs provides a first indication that the statistical association in earnings vary along the earnings distribution. Table 2 reports a correlation

of .26 in levels and .18 in logs. Since the log transformation amounts to emphasize more strongly the bottom of the distribution and to put less weight on deviations from the mean at the top, this is indicative of a stronger association at the top of the distribution than at the bottom.

This is confirmed by figures 1 and 2. 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 spousal correlation in both economic resources and earnings potential. On the one hand, using all observations, including those with zero earnings amounts to ignore that individuals 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 spouses selection is likely to be non-random. Hence the correlation in full-time equivalent earnings should be seen as a biased estimate of the correlation in earnings potential, although the direction of the bias is a priori unknown.

Unbiased estimates of the correlation in earnings potential can be derived from a wage

regression model that explicitly accounts for sample selection. Let w_s denote the log full-time equivalent earnings of spouse 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 normal distribution :

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

Assuming that the distribution of log earnings is a bivariate normal distribution yields the following regression model :

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

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

Assume that w_m is always observed but that w_f is only observed for women in the labor force. In the likely case where participation decisions depend on both spouses' 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. 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 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 spousal 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 spouses.

4.2 Results

Estimation results are given in Tables 5 and 6. Table 5 provides estimates of the regression coefficient, correlation coefficient, and earnings standard-deviation. Estimates in panel A ignore sample selection issues. Estimates in panel B are obtained using Heckman’s sample selection model. Ignoring sample selection issues leads to overestimate the extent of the earnings correlation. This is especially true in the case of annual earnings for which the correlation coefficients fall from around .19 to about .14. For other earnings variables (mean annual earnings, FTE earnings), the fall appears milder. This fall in the estimated correlation arises mechanically from two effects : first, a fall in the spousal 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 is truncated owing to the participation decision.

Table 6 gives the estimates of the Heckman sample selection model. Analyzing the results of the selection equation allows a better understanding of the selection process implied by the participation decision. 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 spouse’s earnings have a lower probability of labor force participation. In other terms, for female, “undermarriage” (i.e. women with high earnings potential conditional on their spouse’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 productivity characteristics of the match.

It is also instructing to examine the association between female participation and their spouse’s annual earnings. Table 6 indicates a hump-shaped relationship, which is also confirmed by the descriptive statistics in table 7. The lowest employment rate is found for partners of males in the first quintile. The employment rate rises with male earnings quintile but fall in the last quintile. As a result of both effects, sample selection will most likely affect undermarried women in couples with low male earnings.

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 to compare 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 on the basis of relevant economic characteristics.⁷

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 (2014). In this approach individuals are characterized by some exogenous characteristics Z such as education and age. 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 characteris-

⁷Several 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.

tics. Hence, the 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. By construction the CDF of the distribution of earnings among couples is equal by construction 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 1/4. 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 .⁸ In other words this amounts to rule out the possibility that household labor supply decisions be also determined by couple's 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. Furthermore, results in table 3 also indicate that 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 earnings. We also implement a third approach that allows assessing the effect of assortativeness on inequality of household *potential* earnings. Contrary 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.⁹ The model of section 4 allows

⁸The procedure developed by Pestel (2015) 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.

⁹This is true, at least, in the short term. In the long run, due to the accumulation of experience and

to parametrically identify the joint distribution of partners' potential earnings in the observed population of 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, defined as the earnings the couple would earn if both partners worked full-time. It is also easy to simulate the distribution of household potential earnings under the assumption that the correlation of spouses earnings potential is zero.

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.

The details of the simulation algorithm are described in the appendix.

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, interdecile ratios - P90/P10, P90/P50, P50/P10). We also report the ratio of the inequality indices in the actual distribution and in the counterfactual distribution, which indicates the inequality reduction obtained by randomizing mating patterns among couples.

Panel A reports the results for addition randomization for annual earnings. Inequality in the actual distribution, for instance the Gini coefficient of .26, is slightly lower than the degree of inequality in the overall distribution of earnings in France. This reflects the seniority, earnings potential also depend on past labor supply decisions. We do not account for this source of endogeneity here.

greater homogeneity of our sample, compared to the overall population, induced by sample selection rules.¹⁰ The equalizing effect of randomizing individual annual earnings across couples, conditional on age, appears relatively modest. The Gini index falls by about 6%. The effect on the other inequality measures is larger : the Theil and Atkinson indices fall by about 10%. 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. One may suspect that these labor supply adjustments would lead to further decrease earnings inequality. As previously discussed, female participation follows a hump-shaped pattern as we move along the distribution of male earnings. Highly educated women with zero earnings, who are more likely married to high earnings men, would probably increase their labor force participation when rematched to a low earnings man, as a result of negative income effects. This would lead to increase earnings at the bottom of the earnings distribution. Symmetrically, women with low and positive earnings would likely decrease their labor supply when rematched to a high earnings man. Again, this would decrease overall earnings inequality. As a consequence, one may suggest that addition randomization provides a lower bound estimate of the effect of random rematching on earnings inequality.

Panel B provides actual and counterfactual inequality measures for the imputation randomization procedure. Measured inequality is slightly different from panel A, owing to small differences in the samples.¹¹ As in the case of panel A, the effect of randomizing educational attainment across couples (conditional on age) is relatively modest. Most inequality measures fall by 3 to 6%, with the exception of Atkinson(2), which falls by about 10%. This modest effect of imputation randomization is in line with the results reported in Eika, Mogstad, and Zafar (2014) and Harmenberg (2014) who also report a modest contribution of assortative mating to inequality between couples. Though one 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

¹⁰Excluding single-headed households will, in particular, drive down inequality measures.

¹¹Since imputation randomization rematches individuals on the basis of their level of education, individuals with missing information on education were excluded from the sample.

rematched with more heterogeneous partners. Unfortunately, it is hard to guess how selection on unobservable characteristics would bias the counterfactual experiment.

Panel C reports estimation and simulation results for FTE earnings derived from the sample selection model. 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 value of the Gini coefficient (.2) and other inequality measures. One advantage of the using FTE earnings (observed or latent) to compute inequality measures is that it may provide a more accurate measure of inequality of living conditions across couples. In fact individuals with zero earnings might engage into domestic production, the value of which is not captured by annual earned income, but which might be proxied by individual market earnings potential. Comparing estimated and simulated inequality measures indicates a much stronger effect of assortative mating on overall inequality across couples. The Gini coefficient falls by 16% in the simulated distribution. The effect on the other inequality indices is even stronger, around 30%.

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 the partners. The correlation coefficient for years of education lies slightly below .6 and the correlation in wage rates amounts to about .35. This high degree of homogamy among French couples 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 mobility in France is relatively low compared to other developed economies. Lecavelier and A (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 reflect 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 to assess 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. We develop an alternative approach that accounts for assortativeness in unobservable earnings determinants and allows to assess the effect of assortative mating on inequality across couples in earnings potential. Our results indicate that assortative mating accounts for as much as 30% of total inequality in earnings potential. The effect on realized earnings is however much smaller, around 5%.

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 (2015). More specifically, the design of couples' income taxation strongly influences the spouses' 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

- ARRONDEL, L., AND N. FREMEAUX (2015): “For richer, for poorer: assortative mating and savings preferences,” *Economica (forthcoming)*, 120(1), 273–344.
- 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.
- BOUCHET-VALAT, M. (2014): “Les évolutions de l’homogamie de diplôme, 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.
- EIKA, L., M. MOGSTAD, AND B. ZAFAR (2014): “Educational Assortative Mating and Household Income Inequality,” NBER Working Papers 20271, National Bureau of Economic Research, Inc.
- 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.

- 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): “Marry Your Like: Assortative Mating and Income Inequality,” *American Economic Review - Papers and Proceedings*, 104(5), 348–353.
- HAGNERÉ, C., AND A. LEFRANC (2006): “Étendue 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,” *Économie 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.
- HARMENBERG, K. (2014): “A note: the effect of assortative mating on income inequality,” *Mimeo*.
- 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).
- 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 L. A (2015): “Siblings correlations in socio-economic outcomes in France,” *mimeo thema*.
- LEFRANC, A. (2011): “Educational expansion, earnings compression and changes in inter-generational 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.
- PESTEL, N. (2015): “Marital sorting, inequality and the role of female labor supply: Evidence from East and West Germany,” .
- 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 corr.	0.4382	0.2829	0.4375	0.2423	0.4555
polychoric corr.	0.5167	0.3722	0.5086	0.3026	0.5306
obs	7101	2633	2633	1661	1661
B - Highest degree					
	own degree	father's degree	own degree	mother's degree	own degree
Spearman corr.	0.5452	0.435	0.5677	0.4023	0.5588
polychoric corr.	0.5808	0.5063	0.6031	0.4793	0.5953
obs	8259	2282	2282	2667	2667
C - Years of education					
	years	years pred.			
Pearson corr.	0.5513	0.6166			
Spearman corr.	0.5486	0.6205			
<i>Net of cohort effects</i>					
Pearson corr.	0.4877	0.546			
Spearman corr.	0.4743	0.523			
obs	7581	8259			

Note: For Panels A and B, in columns 2 to 5, we restrict the sample to the couples for whom the information about both the own and the parental occupation (resp. degree) is available.

Table 2: Correlations - labor earnings

	w_0	w	w_{fte}	w_{fte}
Pearson	0.1923	0.2665	0.2077	0.3193
Pearson log	0.1893	0.1893	0.3135	0.3135
Spearman	0.2037	0.2596	0.2267	0.3026
Net of cohort effects:				
Pearson	0.1905	0.2494	0.2049	0.3012
Pearson log	0.1784	0.1784	0.2942	0.2942
Spearman resid	0.2025	0.2494	0.2236	0.299
obs	8211	6353	7636	5834

Note: in w_0 we include all couples; in w we include only couples in which both partners report positive earnings. Idem for full-time equivalent (fte) earnings.

Table 3: Correlations - labor earnings residuals

	w_0	w	w_{fte}	w_{fte}
No controls				
Pearson	0.1923	0.2665	0.2077	0.3193
Pearson log	0.1893	0.1893	0.3135	0.3135
Spearman	0.2037	0.2596	0.2267	0.3026
After controlling for education:				
Pearson	0.1454	0.1777	0.1936	0.2411
Pearson log	0.1235	0.1235	0.232	0.232
Spearman	0.1485	0.1848	0.2038	0.2381
After controlling for education + social origin:				
Pearson	0.1046	0.139	0.1579	0.2091
Pearson log	0.1008	0.1008	0.179	0.179
Spearman	0.08742	0.1522	0.1294	0.1876
obs	1832	1832	1745	1745

Note: in w_0 we include all couples; in w we include only couples in which both partners report positive earnings. Idem for full-time equivalent (fte) earnings.

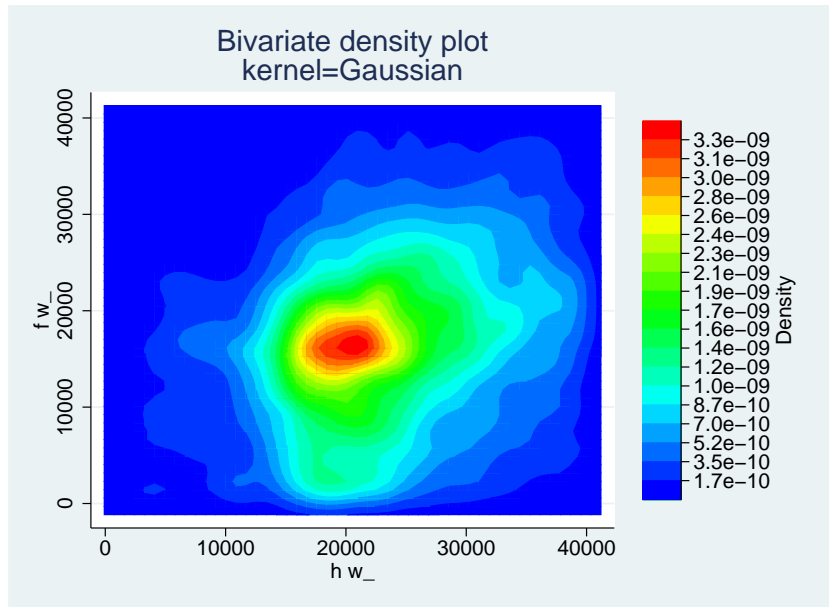
Table 4: Correlations - multi-year average of labor earnings

	w_0	mean w_0	w	mean w	$\ln w$	$\ln(\text{mean } w)$	w_{fte}	mean w_{fte}	$\ln w_{fte}$	$\ln(\text{mean } w_{fte})$
Couples observed at least 3 years										
linear corr	0.1847	0.2091	0.2682	0.2974	0.2015	0.2285	0.3415	0.4036	0.3306	0.4013
spearman	0.201	0.2175	0.2626	0.2894	0.2626	0.2894	0.3184	0.385	0.3184	0.385
linear corr residual	0.1822	0.2083	0.2566	0.2903	0.1957	0.229	0.331	0.3893	0.3204	0.3883
spearman residual	0.1979	0.2203	0.2561	0.2913	0.2554	0.2886	0.3198	0.3833	0.3116	0.3759
obs	4399	4411	3292	3460	3292	3460	3004	3180	3004	3180
Couples observed at least 5 years										
linear corr	0.1817	0.2103	0.2808	0.2879	0.2311	0.227	0.3041	0.3538	0.3038	0.3634
spearman	0.2111	0.2297	0.2826	0.285	0.2826	0.285	0.3006	0.3586	0.3006	0.3586
linear corr residual	0.1769	0.2097	0.2718	0.2845	0.2251	0.2294	0.3043	0.3463	0.3042	0.3572
spearman residual	0.2052	0.2314	0.2812	0.2913	0.2777	0.2849	0.3155	0.3626	0.3041	0.3534
obs	1713	1716	1261	1292	1261	1292	1138	1173	1138	1173

Note:: in w_0 we include all couples; in w we include only couples in which both partners report positive earnings. Idem for full-time equivalent (fte) earnings. The multi-year averages are computed all years for which the information is available.

Figure 1: Bivariate density - Annual earnings

A- earnings levels



B- earnings ranks

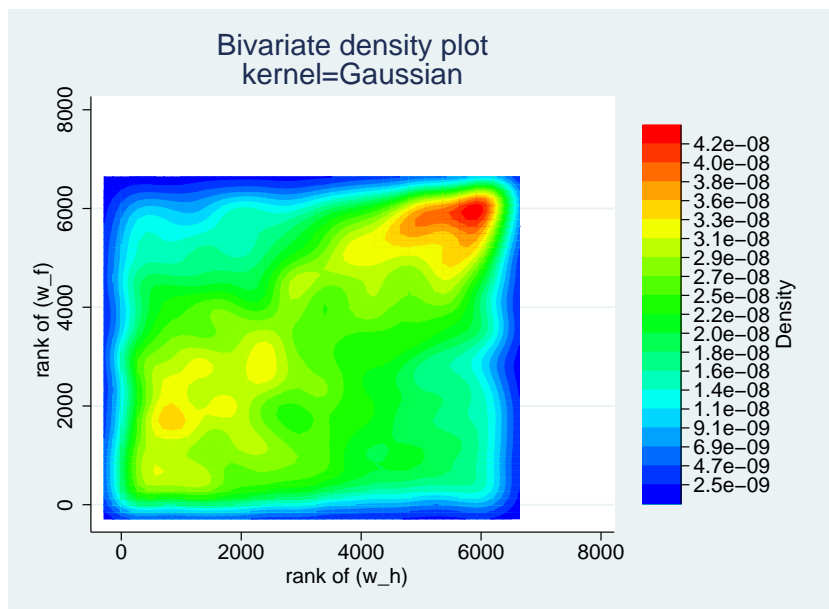
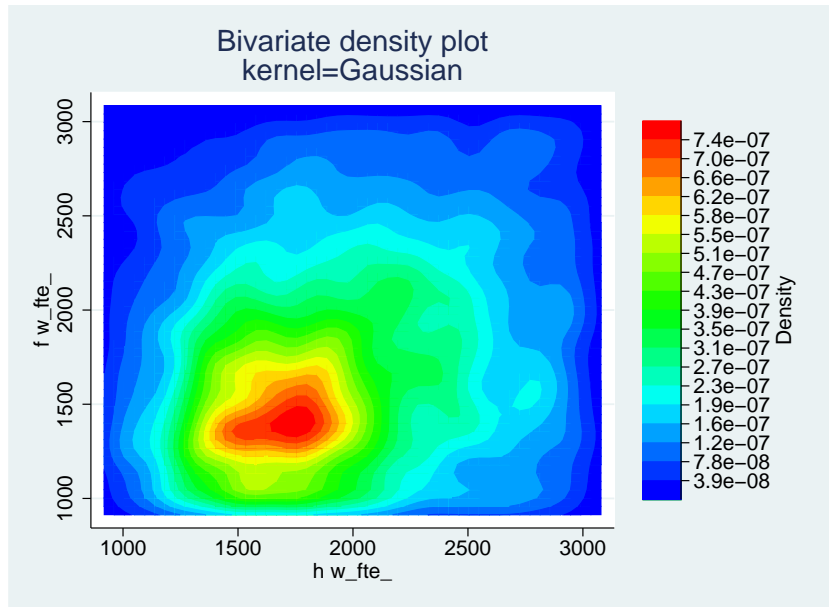


Figure 2: Bivariate density - Full-time equivalent earnings

A- earnings levels



B- earnings ranks

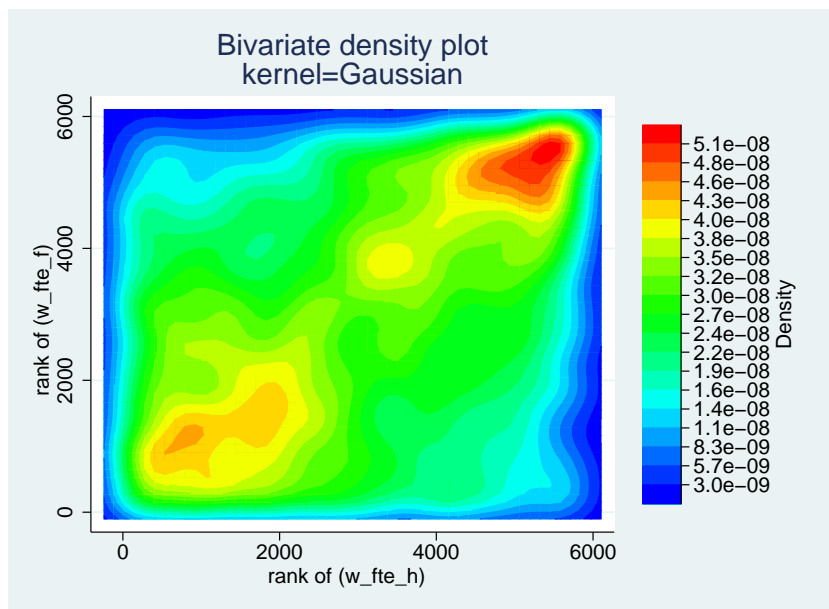


Table 5: Correlations and sample selection- labor earnings

	$\ln w$	$\ln(\text{mean } w)$	$\ln w_{fte}$	$\ln(\text{mean } w_{fte})$
Ignoring sample selection				
β_{OLS}	0.2765	0.2933	0.3072	0.3368
ρ	0.1828	0.1938	0.3074	0.3409
σ_m	0.5118	0.4925	0.397	0.3847
σ_f	0.7742	0.7451	0.3967	0.38
N	6354	6796	5834	6379
Accounting for sample selection				
$\beta_{Heckman}$	0.215	0.2371	0.2892	0.3265
ρ	0.1307	0.1446	0.2781	0.3201
σ_m	0.528	0.4975	0.408	0.3943
σ_f	0.8683	0.816	0.4244	0.4022
ρ_{res}	-0.9749	-0.971	-0.6517	-0.6475
N	7744	7744	7639	7681

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.

Table 6: Sample selection model - labor earnings

	Female wage			
	$\ln w$	$\ln(\text{mean } w)$	$\ln w_{fte}$	$\ln(\text{mean } w_{fte})$
main equation				
male wage	0.215 (.0184)	0.237 (.0179)	0.289 (.0129)	0.326 (.012)
selection equation				
w_m	-0.0561 (.0293)	-0.0262 (.0313)	0.149 (.0399)	0.123 (.0413)
w_m^2	-5.20E-04 (.0033)	-0.0039 (.0036)	-0.0287 (.0047)	-0.0283 (.0048)
age_m	9.10E-04 (.0032)	-0.0026 (.0037)	-0.0051 (.0044)	-0.0108 (.0047)
age_m^2	-2.60E-04 (2.0e-04)	-1.90E-04 (2.2e-04)	-1.90E-04 (2.9e-04)	-1.90E-04 (3.1e-04)
age_f	0.0222 (.003)	0.0238 (.0033)	0.0305 (.0041)	0.0309 (.0044)
age_f^2	-1.00E-03 (2.1e-04)	-0.0013 (2.2e-04)	-0.002 (3.0e-04)	-0.0023 (3.1e-04)
yrs of edu $_f$	0.122 (.0322)	0.124 (.0354)	0.25 (.0483)	0.232 (.0508)
yrs of edu $_f^2$	1.70E-04 (.0012)	3.90E-04 (.0014)	-0.0018 (.0018)	-0.0011 (.0019)
kids	-0.145 (.0117)	-0.143 (.0129)	-0.224 (.0165)	-0.195 (.0171)
long-term contract	-0.0129 (.0324)	-0.0556 (.0362)	0.172 (.0466)	0.172 (.0488)
capital income	4.60E-06 (2.8e-06)	7.30E-06 (3.3e-06)	-4.90E-07 (2.9e-06)	2.20E-06 (3.5e-06)
cons	-1.38 -0.233	-1.18 -0.252	-2.99 -0.353	-2.4 -0.369
ρ_{res}	-0.9749	-0.971	-0.6517	-0.6475
lnsigma				
cons	-0.15 -0.0095	-0.214 -0.0092	-0.897 -0.0118	-0.965 -0.0109

Note:: Standard errors in parenthesis. w : annual earnings, w_{fte} : full-time equivalent earnings. Indices m for the male partner and f for the female partner.

Table 7: Female labor market participation by male earnings quintiles

	$\ln w$	$\ln w_{fte}$
<u>A - Share of female with positive earnings</u>		
Quintile 1	0.769	0.7082
Quintile 2	0.8243	0.7675
Quintile 3	0.8451	0.7892
Quintile 4	0.8554	0.8076
Quintile 5	0.8088	0.7708
<u>B - Female number of months worked</u>		
Quintile 1	9.276	9.377
Quintile 2	9.465	9.561
Quintile 3	9.664	9.575
Quintile 4	9.813	9.758
Quintile 5	9.471	9.444

Table 8: Earnings inequality - Observed and simulated matching

	Gini	Theil	A(1)	A(2)	p75p25	p10p50	p90p50
<u>A- Annual earnings, addition randomization</u>							
observed	0.265	0.116	0.127	0.336	1.880	0.479	1.733
simulated	0.249	0.102	0.112	0.306	1.803	0.497	1.663
ratio	0.940	0.884	0.888	0.911	0.959	1.037	0.960
<u>B- Annual earnings, imputation randomization</u>							
observed	0.261	0.113	0.123	0.331	1.863	0.492	1.739
simulated	0.254	0.107	0.116	0.296	1.836	0.506	1.706
ratio	0.972	0.947	0.941	0.894	0.985	1.029	0.981
<u>C- FTE earnings, addition randomization with sample selection correction</u>							
observed	0.201	0.065	0.063	0.122	1.624	0.626	1.564
simulated	0.169	0.046	0.045	0.087	1.492	0.674	1.460
ratio	0.838	0.699	0.706	0.712	0.918	1.078	0.934

Note: A(1) and A(2) denote the Atkinson inequality indices with coefficient 1 and 2 respectively; ppxy denotes the ratio of the ratio of the x^{th} percentile over the y^{th} percentile.

A Simulation algorithms

A.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.

A.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 (2014) 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.

A.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) earnings potential 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) earnings potential in the population under the assumption of random mating.