the review of income and wealth

Review of Income and Wealth
Series 66, Number 4, December 2020
DOI: 10.1111/roiw.12450

ASSORTATIVE MATING AND EARNINGS INEQUALITY IN FRANCE*

BY NICOLAS FRÉMEAUX

LEMMA—Université Paris 2

AND

ARNAUD LEFRANC

THEMA—Université de Cergy-Pontoise

This paper analyzes 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. Second, we assess the contribution of assortative mating to earnings inequality between couples. We provide a new method for assessing the contribution of assortative mating to inequality in couple's potential earnings, that accounts for selection bias arising from labor force participation. Our results indicate a strong degree of assortative mating in France. The correlation in earnings is around 0.17 for annual earnings, around 0.35 for full-time equivalent earnings and up to 0.49 when using multi-year average earnings. Assortative mating tends to increase inequality among couples. For annual earnings, the effect accounts for 3 to 9 percent of measured inequality. The effect of assortative mating on household potential earnings is much larger and amounts to 10 to 20 percent for observed inequality.

JEL Codes: J12, J22, D31

Keywords: assortative mating, earnings, France, inequality, labor supply

1. INTRODUCTION

An abundant sociological literature has provided evidence of a high correlation of educational and social attributes within couples, in most developed countries.[1] 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,[2] uncertainty remains on the contribution of assortative mating to earnings inequality, as evidence is largely confined to the US case and

Note: This research was supported by the FAMINEQ project (ANR-17-CE41-0007). It 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.

*Correspondence to: Arnaud Lefranc, THEMA—Université de Cergy-Pontoise, 33, boulevard du Port - 95011 Cergy-Pontoise cedex, France (arnaud.lefranc@u-cergy.fr).

[1]See e.g. Mare (1991), Blossfeld and Timm (2003), Goux and Maurin (2003), Schwartz and Mare (2005), Kalmijn (1991), Uunk *et al*. (1996).

[2]See in particular Karoly and Burtless (1995), Cancian and Reed (1998), Burtless (1999), Schwartz (2010), Eika *et al*. (2017), Greenwood *et al*. (2014), Harmenberg (2014), Pestel (2017).

dependent on methodological choices. 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, earnings), as usually investigated in the literature. Second, we bring together in one paper several methodological issues that have been covered separately in previous papers. Specifically, in order to account for endogenous labor supply, we examine the association within couples in individual potential earnings, measured by full-time equivalent earnings (Hyslop, 2001). Moreover, we account for potential biases in the estimation of assortative mating arising from sample-selection into the labor force (Shaw, 1989). Third, we assess the contribution of assortative mating to inequality between couples, in France, in both observed annual earnings and potential earnings.

Several recent papers have examined the statistical association between male and female labor earnings within couples.[3] Available evidence for the United States points to a sizable correlation, of up to 20 percent (e.g. Burtless, 1999; Schwartz, 2010). Apart from the US case, evidence for other countries is rather sparse, although papers have examined several European societies: Sweden (Nakosteen *et al.*, 2004), Germany (Pestel, 2017; Eika *et al.*, 2017), the United Kingdom, Norway and Denmark (Eika *et al.*, 2017), and Switzerland (Ravazzini *et al.*, 2017). The present paper contributes to the growing evidence on the effect of assortative mating by looking at the French case.

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 incorporate sizable measurement errors and transitory shocks. If such errors and shocks are poorly correlated between partners, these components will lead to underestimate the association in partners' long-term earnings. 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. This is of course a relevant measure in its own right. However annual earnings reflect both individual productivity characteristics and endogenous joint labor supply decisions taken within the couple. 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. The confounding effect of labor supply decisions might then jeopardize the assessment of the degree of assortative mating in core individual attributes. In this paper, this issue is addressed by also analyzing the statistical association in potential earnings within couples, defined by the individual full-time

[3]Correlation in other economic dimensions such as individual preferences has been much less analyzed. Arrondel and Fremeaux (2016), Dohmen *et al.* (2012) and Kimball *et al.* (2009) are some of the few exceptions.

equivalent earnings an individual would command on the labor market. 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 assess inequality in welfare between households.

Potential earnings is only observed for individuals working and is latent otherwise. 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 sample-selection regression model.

One of the main economic motivations for studying assortative mating lies in its potential contribution to economic inequality between couples. This contribution has only been studied recently and is generally found to be modest. Specifically, Greenwood *et al*. (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 *et al*. (2017) conclude that the contribution of assortative mating to inequality is around 5 percent. 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 adequately deal with the endogeneity of labor supply decisions and the self-selection of individuals into couples, on the basis of their unobserved characteristics.

We review the main approaches taken in the recent literature[4] and develop an alternative method 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 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.[5] 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 is 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.

[4]See e.g. Karoly and Burtless (1995), Cancian and Reed (1998), Burtless (1999), Schwartz (2010), Hryshko *et al*. (2017), Greenwood *et al*. (2014), Harmenberg (2014), Pestel (2017), Eika *et al*. (2017).

[5]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 account for the dynamics of human capital and employment opportunities.

Lastly, our estimates indicate that assortative mating tends to increase inequality among couples by up to 20 percent. But the magnitude of these effects appears to vary with the earnings measure, the imputation method and the inequality index used in the analysis. In particular, the effect of assortative mating is found to be larger for potential earnings than for annual earnings and for inequality indices more sensitive to the tails of the distribution. 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 French waves of the EU-SILC surveys. We focus on the waves 2004 to 2011. The EU-SILC is a longitudinal household survey which focuses on income, poverty, social exclusion and living conditions. Although the EU-SILC provides information for all EU member states, data harmonization is only partial.[6] In particular, it is not based on a harmonized survey questionnaire. Data is collected nationally using a mixture of administrative data and country-specific surveys. Furthermore, panel length is set at four years for most countries but it extends to eight years for France. For these reasons, the analysis of assortative mating in this paper is confined to the case of France.[7] We leave to future research the assessment of assortative in the EU at large, which will necessarily be constrained by data harmonization issues.

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 spouse 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 or out of the labor force because of retirement or studying. This results in a sample of 7,966 couples. We also exclude couples in which earnings are zero for both partners in all available years (102 couples). In the end, our analysis is based on a sample of 7,864 couples with valid information on age, years of education and earnings (for at least one year). Appendix A provides general descriptive statistics on our final sample. The

[6]See e.g. Iacovou *et al*. (2012).

[7]Quality reports of the French EU-SILC data can be found at https://ec.europa.eu/eurostat/web/income-and-living-conditions/quality/eu-and-national-quality-reports.

exact number of couples and observations in the tables of results below may vary slightly due to missing observations on some variables.

2.2. *Main Variables*

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

Earnings

Annual total 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. This variable is denoted $w0$. In some estimations, we restrict attention to individuals with strictly positive annual earnings. This variable is denoted w and is missing for individuals out of salaried employment.

Annual 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.[8] To compute FTE earnings, we rely on the history of monthly labor force participation in the preceding year, as reported in the survey. For individuals out of salaried work, FTE earnings are missing, by construction. In later tables, this variable is denoted w^{FTE}.

In the estimations based on either measure of annual earnings, we only keep one observation per couple, to avoid under-representing couples with higher attrition risk. 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).

For both earnings measures, we also compute multi-year averages of individual earnings. For each individual, the average is computed over the full set of available yearly observations. For annual earnings excluding zeros (w) and for FTE earnings (w^{FTE}) we only consider positive earnings. In other words, for an individual observed for 3 years and who reports earnings equal to zero in one wave, we only estimate the average earnings over the 2 years during which the individual's earnings are positive. However, for the multi-year average value of $w0$ we keep all available observations, including zeros. The number of years of observation in our sample varies between 1 and 8 years, with an average of 3.4 years.

[8]The survey provides two measures of work duration: an indicator of full-time vs part-time work (for each month in the preceding year) and the regular number of hours worked (per week, in the current job). While the second measure is more precise it only refers to the situation as of the survey date and is missing for around 10 percent of the sample. This corresponds to individuals who are unable to report a regular time usually worked per week. When we focus on the sample for which both variables are available, we obtain very close estimates of correlations of FTE earnings between partners for the two definitions of work duration.

Potential Earnings

One obvious limitation of using reported annual earnings for welfare analysis is that it ignores the welfare value of non-participation, which arises from two important components: domestic production on the one hand and the value of leisure on the other hand. The contribution of individuals out of the labor force to the household's consumption of goods and services through domestic production, was first emphasized in Gronau (1977). Available estimates indeed suggest that domestic production represents a sizable fraction of household consumption.[9] Besides the value of leisure enjoyed, a third welfare contribution of individuals out of the labor force, in households with children, is the value of human capital investment undertaken at home.

How this welfare contribution should be measured is a difficult issue. Measures of household production usually combine individual market wage information with time-use surveys to value the domestic production of basic services (cleaning, gardening, shopping…). However, this falls short of integrating the value of leisure. In any case, information on time use is not available in our data. In this paper, we measure potential earnings by the individual full-time equivalent earnings. This amounts to value total available time at the prevailing individual market wage and can be seen as an encompassing measure of the resources available that ultimately determine household welfare. Of course, equating the value of domestic time to the market wage rate is only true at the margin[10] and the value of infra-marginal units of time is higher, thus leading to underestimate true welfare contribution. However, we believe that this represents an improvement over the assumption that the welfare contribution of non-participation is equal to zero.

Last, it is worth emphasizing that potential earnings is a latent variable. When it is observed, it is equal to FTE earnings. But sample selection issues must be taken into account when asserting the correlation in potential earnings within couples, as we do in Section 4.

2.3. *Other Socioeconomic Variables*

Education

We consider two measures of 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). The second variable is an ordered qualitative measure 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.

[9]See for instance House *et al.* (2008), Frazis and Stewart (2011), Ahmad and Koh (2011), Roy (2012).

[10]And for an interior solution to the optimal time allocation problem.

Socio-Economic Background

The SILC survey investigated individual socioeconomic origin and gathered information on education and occupation of both parents of adult respondents. For both parents, education is measure by the highest degree obtained. Occupation is recorded as above. Information is missing for a large part of our sample, as it is only collected in the 2005 wave.

Detailed classifications, for each of these variables, are given in Appendix A.

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. For occupation and highest degree completed, the association is measured using the Spearman rank correlation coefficient. For the number of years of education, we report the linear Pearson correlation coefficient.

Table 1 provides our estimates of assortative mating for occupation and education. Occupational correlations are given in panel A. Columns 1 presents the correlations for own occupation for the whole sample. Columns 2 and 3 presents the correlation for father's (resp. mother's) occupation, on the sub-sample where father's (resp. mother's) occupation is reported.[11] The correlation in partners' own occupation equals 0.453, which appears high, though in line with estimates found for other countries. This can be compared to assortativeness in social origin, as captured by parental occupation. The correlation among partners in fathers' or mothers' occupation is positive and between 0.249 and 0.291, which indicates positive assortative mating by social origin. The correlation is higher for fathers' occupation than for mothers'. The absence of information for a significant share of respondents' mother (mainly because of inactivity) makes the comparison difficult.

Panels B and C of Table 1 report statistical associations in education. For the highest degree completed (Panel B), we find positive correlations of 0.559. This correlation appears higher 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. The correlations for the number of years of education (Panel C) are higher, around 0.62, but consistent with those obtained for the 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). Moreover, our estimates for France appear

[11]We also estimated correlations for own occupation and education on the sub-sample for which the information about social background is available. We found very close estimates compared with correlations based on the whole sample, as reported in column 1.

TABLE 1
Correlation coefficients—occupation and education

	(1)	(2)	(3)
A: Occupation			
	own occ.	father's occ.	mother's occ.
ρ_S	.453	.291	.249
	[.434,.471]	[.255,.326]	[.203,.294]
Obs.	6928	2559	1635
B: Highest degree			
	own degree	father's degree	mother's degree
ρ_S	.559	.437	.401
	[.543,.574]	[.403,.47]	[.368,.433]
Obs.	7864	2202	2571
C: Years of education			
	own education		
ρ	.624		
	[.611,.638]		
Obs.	7864		

Note: 95% confidence interval in square brackets. ρ indicates the Pearson correlation coefficient, ρ_S indicates the Spearman correlation coefficient Estimates in columns 2 to 3 are restricted to the sample of couples for which information on parental occupation or degree is available.

higher than the correlation reported for most European countries but similar to those reported for the US.[12]

3.2. *Earnings*

Annual and FTE Earnings

To assess the extent of economic assortative mating, we examine the correlation between partners in annual and full-time equivalent (FTE) earnings. Results are presented in Table 2. Panel A presents gross correlation coefficients. Panel B presents residual correlations net of individual characteristics: age and education. Panel C provides residual correlations net of age, education and parents' occupation. For the computation of residual correlations, we purge earnings from the effect of the individual characteristics and estimate the correlation in earnings residuals. For all panels, we also add year fixed effects. While estimates in Panels A and B are based on the same sample, estimates in Panel C are restricted to the sub-sample for which the information about the spouses' social background is available.

Gross correlation coefficients are presented in panel A of Table 2. Column 1 reports correlations in annual earnings based on all observations, including zeros. 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 (0.308). The gap in the estimated correlation between the two samples is likely to be explained by

[12]Specifically, Fernandez *et al.* (2005) estimate correlations for the number of years of education of 0.4 for Great-Britain, around 0.4–0.5 for Scandinavian countries and around 0.5 for Germany and the Netherlands. The correlation for the United States equals 0.6. For in-depth sociological assessment of mating insitutions and processes in France, see Bozon and Rault (2012) and references therein.

TABLE 2
CORRELATIONS COEFFICIENTS—LABOR EARNINGS

Control variables		(1) w^0	(2) w	(3) w^{FTE}
A: Gross correlations				
none	ρ	.175	.308	.349
		[.153,.196]	[.286,.331]	[.327,.371]
Obs.		7864	5983	5983
B: Residual correlation, net of individual characteristics				
age	ρ	.169	.296	.337
		[.148,.191]	[.273,.319]	[.315,.360]
education	ρ	.073	.221	.271
		[.051,.095]	[.197,.245]	[.247,.294]
age, education	ρ	−.011	.115	.182
		[.−033,.011]	[.090,.140]	[.157,.206]
Obs.		7864	5983	5983
C: Residual correlation, net of individual characteristics and social origin				
age, education,	ρ	−.044	.082	.129
parents' occupation		[−.096,.007]	[.024,.139]	[.071,.186]
Obs.		1422	1145	1145

Note: w^0: annual labor earnings, including zeros; w: annual labor earnings, excluding zeros; w^{FTE}: full-time equivalent annual labor earnings, excluding zeros. 95% confidence interval in square brackets. ρ indicates the Pearson correlation coefficient.

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, since on the sub-sample of non participants, the spouse correlation in earnings is null.[13] Whether the assumption of random participation constitutes a reasonable approximation 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 (0.349).[14] 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 percent. This 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 labor supply decisions. We address this issue more carefully in Section 4.

[13]In fact, under random participation, the presence of zeros would mechanically lead to a decrease in the covariance of earnings among partners. Furthermore, the inclusion of zeros 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.

[14]This result is consistent with estimates of Shaw (1989).

It is worth stressing that using FTE earnings may not suffice to fully account for the confounding effect of labor supply decisions on the intensive margin. This would for instance be the case if involuntary part-time work resulted in a wage rate penalty, thus leading to underestimate the true FTE earnings of part-time workers. To correct for this possible source of bias, we re-estimated the earnings correlation on the sub-sample of couples where both partners reported working full-time for at least 80 percent of the preceding year. On this subsample, we can presumably rule out the existence of a sizable part-time wage penalty. This resulted in slightly higher correlations for FTE earnings, at 0.4 against 0.349 for the full sample. Of course these sensitivity checks should be interpreted with caution, as there might be other factors explaining the difference in point estimates between these samples, including differences in individual characteristics and even the degree of similarity within the couple. However, these complementary results suggest that estimated degree of correlation in FTE earnings reported in Table 2 could be a lower bound estimate of the degree of assortative mating in earnings.

Two preliminary conclusions can be drawn from Table 2. First, results indicate 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 percent and 55 percent 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.

Contribution of Age, Education and Social Origin

One may suspect that part of the correlation in earnings arises from the correlation in several individual characteristics. It may for instance be driven by life-cycle effects, through the correlation in age within couples. Moreover, as noted in the introduction, many 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. In fact, one may object to the analysis of assortativeness by earnings that it merely reflects the correlation in partners' age, education and social origin. To address this issue, we examine whether earnings remain correlated, once they have been purged from the effect of age, education and social origin.

In panel B, we first estimate the correlation in earnings after netting out age effects.[15] Results indicate a modest fall in the estimated correlation. The correlation coefficient falls by around 3.5 percent for all earnings definitions. Then, controlling

[15]This is achieved by first regressing earnings on a quartic function of age and taking residuals.

for education alone decreases the correlation by about 58 percent for the whole sample. The effect is around twice smaller for dual-earner couples when we consider annual or FTE earnings. Finally, once controlling for both age and education, the residual correlation goes to zero on the full sample, including individuals out of the labor force with zero earnings. However, the residual correlation remains significant and sizable for dual-earner couples, at .115 for annual earnings and .182 for FTE earnings. Panel C confirms that after netting out social origin effects as well as age and education the remains a sizable residual correlation in particular for FTE earnings. As a conclusion, even if assortativeness in terms of age, social background and education is high, as discussed in Section 3.1, there remains significant sorting along other dimensions that are not captured by these variables.

Multi-Year Average Earnings

A potential challenge to the estimation 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. The incidence of measurement errors and transitory shocks has been widely documented in the related field of intergenerational earnings mobility studies.[16] Here, unlike the case of intergenerational mobility, transitory earnings components may however be correlated across partners, for reason owing to local business cycle or industry level shocks in the case of partners working in a similar industry.

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 3. 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 variable 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 correlation coefficient increases by 13 percent when averaging annual earnings over at least three-years. Using average earnings has a similar effect on the correlation in FTE earnings that increases by about 17 percent to reach a high value of 0.466. 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

[16]See for instance Solon (1992) and the survey of Black and Devereux (2011).

TABLE 3

Correlation coefficients—multi-year average of labor earnings

	(1)	(2)	(3)	(4)	(5)	(6)
	$w0$		w		w^{FTE}	
	Annual	Multi-year average	Annual	Multi-year average	Annual	Multi-year average
A: Couples observed at least 3 years						
ρ	.192	.222	.336	.39	.389	.466
	[.163,.221]	[.193,.25]	[.304,.367]	[.36,.419]	[.359,.418]	[.438,.493]
Obs.	4258	4258	3106	3106	3106	3106
B: Couples observed at least 5 years						
ρ	.235	.257	.367	.416	.385	.49
	[.189,.28]	[.212,.301]	[.317,.415]	[.368,.462]	[.335,.432]	[.446,.532]
Obs.	1677	1677	1185	1185	1185	1185

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

earnings by about one third. This is consistent with available evidence indicating, first, that measurement errors in annual earnings account for 10 to 15 percent 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.[17]

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

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.

Let w_s denote the potential 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:

$$\begin{pmatrix} w_m \\ w_f \end{pmatrix} \rightarrow \ln \mathcal{N}(\boldsymbol{\mu}, \Sigma) \quad \text{with} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_m \\ \mu_f \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_m^2 & \rho\sigma_m\sigma_f \\ \rho\sigma_m\sigma_f & \sigma_f^2 \end{pmatrix} \tag{1}$$

The difficulty in deriving estimates of the parameters of the (latent) joint distribution, $(\rho, \sigma_m, \sigma_f)$, lies in the fact that potential earnings is subject to non-random sample selection. However, as we now discuss, unbiased estimates of these parameters can be derived from a wage regression model that explicitly accounts for sample selection.

[17]Ostrovsky (2012) reports supportive evidence.

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

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

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.[18] 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. The partners' correlation therefore cannot be directly assessed, based on observed earnings alone. Likewise, β in equation 2 cannot be estimated by linear regression. Last, the observed distribution of w_f will be censored by participation decisions and the estimation of the standard errors of female potential earnings from observed data will be biased.

However, all these parameters can be consistently estimated using Heckman's sample selection correction applied to equation 2. More specifically, this model yields consistent estimates of both β and σ_ε. Furthermore, these estimates can be combined to obtain an estimate of $\sigma_f = \sqrt{\sigma_\varepsilon^2 + \beta^2\sigma_m^2}$. Last, one can obtain an estimate of the within-couple correlation in potential log-earnings, ρ, given by:

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

We use this approach to estimate the 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 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 4 and 5. Table 4 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

[18] Table A1 shows that the share of men reporting positive earnings equals 94 percent while this share equals 77 percent for women.

TABLE 4
Correlation coefficients and sample selection — labor earnings

	(1)	(2)
	$\ln w^{FTE}$	$\ln(\text{mean } w^{FTE})$
A: Ignoring sample selection		
β_{OLS}	.329	.359
ρ	.326	.361
σ_m	.407	.396
σ_f	.411	.395
Obs.	5983	6383
B: Accounting for sample selection		
$\beta_{Heckman}$	.321	.357
ρ	.31	.353
σ_m	.421	.409
σ_f	.436	.414
ρ_{res}	−.619	−.606
Obs.	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 zeros. "mean" indicates the multi-year averages, computed over all years for which the information is available, exluding zeroes. Estimates are based on coefficients of the sample-selection model presented in Table 5.

earnings. Hence, we concentrate here on FTE earnings. We consider two variants of the selection model: using w^{FTE}, the single-year measure allows to correct for point-in-time non participation; using mean w^{FTE}, which is averaged over all available non-zero earnings observations allows to correct for persistent withdrawal from the labor force. Results are very similar across both variants.

Estimates in Table 4, panel A, ignore sample selection issues. The results 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 age effects have been removed (Table 2, panel B). The difference between the two estimates is not statistically significant. Estimates in panel B of Table 4 control for sample selection using the procedure outlined above, based on Heckman's 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, in the case of France, sample selection into employment has only a moderate impact on the estimated earnings correlation. It is also worth stressing that correlation coefficients are not statistically different between panel A and panel B.[19]

Table 5 gives the estimates of the Heckman sample selection model. Analyzing the results of the selection equation allows a better understanding of the process

[19]The confidence intervals (at 95 percent) for Panel A are [.303; .349] for ln(w^{FTE}) and [.339; .381] for ln(mean w^{FTE}).

TABLE 5
SAMPLE SELECTION MODEL—LABOR EARNINGS

	(1)	(2)
Dependent variable	$\ln w_f^{FTE}$	$\ln(\text{mean } w_f^{FTE})$
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)
Obs.	7526	7526

Note: Standard errors in parenthesis. w^{FTE}: full-time equivalent annual labor earnings, excluding zeros. "mean" indicates the multi-year averages, computed over all years for which the information is available, exluding zeroes. Indices m for the male partner and f for the female partner. *s indicate significance level:
*$p < 0.10$; **$p < 0.05$;***$p < 0.01$.

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 5 also allows assessing the relationship between male earnings and female labor market participation. The coefficients of w_m and w_m^2 indicates a hump-shaped relationship. Table 6 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 percent 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

TABLE 6
Female labor market characteristics conditional on male earnings deciles

	(1)	(2)	(3)	(4)	(5)	(6)
	Work	Unemp.	Inactivity	Months worked	w	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.

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.[20]

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 *et al*. (2017) and to some extent Burtless (1999) and Aslaksen *et al*. (2005). 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 *et al*. (2014) and Eika *et al*. (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 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 $\tilde{F}(y) = \frac{1}{4}\{F_{HH}(y) + F_{HL}(y) + F_{LH}(y) + F_{LL}(y)\}$.

[20]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.

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.[21] 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 2, 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 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.[22]

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 equation 1 in 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. To obtain this measure of inequality, we simply simulate, based on model estimates, the joint distribution of male and female potential earnings, which would be fully observed in the absence of sample selection. Furthermore, 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

[21]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.

[22]This, we believe, is a reasonable assumption, 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.

distributions), and to compute the degree of inequality in joint potential earnings that would prevail under random mating.[23]

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 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.

To summarize, we implement three randomization methods:

1. addition randomization which treats individual earnings as a fixed individual characteristic and randomly assign individual earnings into simulated couples;
2. imputation randomization which focuses on individual education, randomly assigns individual education into simulated couples and draws joint couple earnings from the observed distribution of couples with similar characteristics;
3. addition randomization with sample selection correction which treats individual potential earnings as a fixed individual characteristic, uncovers the latent joint distribution of potential earnings and randomly assign individual potential earnings into simulated couples.

All 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 7. For the observed and simulated earnings distributions we compute standard inequality indices (Gini, Theil, Atkinson (1 and 2) and P90/P10). We also report the variation of the inequality indices between the actual distribution and

[23]One should stress here that this randomization procedure rests on the assumption that potential earnings is a fixed individual characteristic which is unaffected by possible rematching. In particular the variance in the marginal distribution of male and female latent potential earnings, σ_m and σ_f, which we estimated in Section 4, is assumed to be unaffected by the mating pattern. Of course, as a result of random rematching, the pattern of sample selection would of course change and so would the variance of earnings in the *observed* distribution. The case where potential earnings is partly influenced by past labor market participation and work experience would deserve complementary investigations.

TABLE 7
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 90_{th} percentile over the 10_{th} percentile.

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.[24] The equalizing effect of randomizing individual annual earnings across couples, conditional on age, appears relatively modest. The Gini index falls by 8.5 percent. The effect on the other inequality measures is larger: the Theil and Atkinson indices fall by about 17–18 percent. 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.

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

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 percent is in line with the results reported in Eika *et al.* (2017), Greenwood *et al.* (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 percent. 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 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 percent and the Theil index by 21.3 percent. 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 percent (against only 2.8 percent for annual earnings) and a fall in the Theil index of 17.4 percent. The assumption that heterogamous couples are a good counterfactual for the behavior of individual observed in homogamous couples, which underlies the imputation randomization approach, leads however to larger differences for inequality measures that are more sensitive to inequality at the top or bottom of the distribution like the Atkinson index.[25] This indicates that the two methods differ in the degree of inequality predicted in the tails of the counterfactual distribution even when labor supply decisions are neutralized by the use of FTE earnings.

In panel C, we control for sample selection. Compared to panels A and B, controlling for sample selection, yields different estimates of inequality in FTE earnings. In the light of Tables 4 and 6, controlling for sample selection has two conflicting effect on the assessment of inequality. First, estimates of the dispersion in both male and female earnings increase as we allow for truncation in the observed distribution. Other things equal, this should increase the level of inequality. However, control for sample selection also reduces the correlation in earnings among partners which tends to decrease the degree of inequality in couple's total

[25]We find similar results for the Generalized Entropy measures (not included in the table).

potential earnings. The latter effect dominates for most inequality measures reported in the table. The Gini coefficient falls from .207 to .196. The Theil and Atkinson(1) fall as well.[26]

Panel C provides additional insights into the disequalizing effect of mating patterns. Under random matching, the Gini coefficient would fall by 8.7 percent and the Theil index would fall by about 16.6 percent. Compared with the addition randomisation method in panel B, the disequalizing effect of assortative mating appears lower. This is consistent with results noted in Table 4 indicating that, once sample selection is accounted for, the correlation in FTE earnings is lower than on the dual-earners couple. The disequalizing effect is close to that obtained under the imputation method, except for the Atkinson indices and the P90/P10 ratios, once again suggesting that counterfactual prediction of the distribution that would prevail under random mating is sensitive to the method used in the case of the tails of the distribution. Hence tail sensitive inequality indices are more dependent on the choice of the imputation method.

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. Three 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. Third, the effect of assortative mating on inequality varies across inequality indices. This multi-indices analysis helps to consider the non-linear female labor participation.

Compared with Eika *et al.* (2017)'s estimates, we find that the effect of assortative mating on inequality is slightly lower in France than in the US. However, our estimates are rather similar to those found for Germany[27] and the UK and higher than in Denmark and Sweden.

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 0.175, 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

[26]The P90/P10 ratio and Atkinson(2) indices remain almost unchanged, indicating that for indices more sensitive to the tails of the distribution, the two conflicting effects cancel out.

[27]See also Pestel (2017).

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 (2018) 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 *et al*. (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 using different methods 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. In this paper, we consider assortative mating regarding potential earnings, 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 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.[28] For annual earnings, the effect is moderate and accounts for 4 to 10 percent of measured inequality. The effect of assortative mating is however much larger when focusing on couples' potential earnings and amounts to 10 to 20 percent for observed inequality. The effect of assortative mating is found to be

[28]See Immervoll *et al*. (2005).

larger for inequality indices more sensitive to the tails of the distribution. Correcting for sample-selection has a moderate effect on the results.

The discrepancy between the 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.

Last, two limitations and possible extensions of the present work should be highlighted. First, our analysis has focused on labor earnings, leaving aside other sources of income. Assessing total income inequality would in particular require accounting for capital income. This is difficult in practice since in most surveys capital income is only available at the household level. And assuming that capital income is shared equally between partners would be inadequate, given recent evidence of a rising inequality within couples in capital endowments (Fremeaux and Leturcq, 2019). Second, we have exclusively focused on pre-tax inequality. 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 design of couples' income taxation strongly influences 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 developed countries has implemented an individual income tax scheme (Pearson and Binder, 2014), although France in particular implements taxation at the household level. Future research should evaluate the consequences of these differences in tax systems for inequality driven by assortative mating.

References

Ahmad, N. and S.-H. Koh, "Incorporating Estimates of Household Production of Non-Market Services into International Comparisons of Material Well-Being," OECD Statistics Working Papers 2011/7, OECD Publishing, Paris, 2011.

Arrondel, L., and N. Fremeaux, "For Richer, for Poorer: Assortative Mating and Savings Preferences," *Economica*, 83, 518–43, 2016.

Aslaksen, I., T. Wennemo, and R. Aaberge, "Birds of a Feather Flock Together: The Impact of Choice of Spouse on Family Labor Income Inequality," *Labour*, 19, 491–515, 2005.

Becker, G., and N. Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87, 1153–89, 1979.

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

Black, S. and P. J. Devereux, "Recent Developments in Intergenerational Mobility," in O. Ashenfelter, and D. Card (eds), Handbook of Labor Economics, North Holland Press, Elsevier, 2011.

Blossfeld, H.-P. and A. Timm, *Who Marries Whom?: Educational Systems as Marriage Markets in Modern Societies*, Springer, 2003.

Bollinger, C. R., and A. Chandra, "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labor Economics*, 23, 235–58, 2005.

Bouchet-Valat, M., "Les Evolutions de L'homogamie de Diplome, de Classe et D'origine Sociales en France (1969–2011): Ouverture d'ensemble, Repli des Elites," *Revue française de Sociologie*, 55, 459–505, 2014.

Bozon, M., and W. Rault, "De la Sexualitïau Couple-L'espace des Rencontres Amoureuses Pendant la Jeunesse," *Population*, 67, 453–90, 2012.

Burtless, G., "Effects of Growing Wage Disparities and Changing Family Composition on the U.S. Income Distribution," *European Economic Review*, 43, 853–65, 1999.

Cancian, M., and D. Reed, “The Effects of Wives Earnings on Family Income Inequality,” *Review of Economics and Statistics*, 80, 73–9, 1998.
Chadwick, L., and G. Solon, “Intergenerational Income Mobility among Daughters,” *American Economic Review*, 92, 335–44, 2002.
Crossley, T. F., and S.-H. Jeon, “Joint Taxation and the Labour Supply of Married Women: Evidence from the Canadian Tax Reform of 1988,” *Fiscal Studies*, 28, 343–65, 2007.
Dohmen, T., A. Falk, D. Huffman, and U. Sunde, “The Intergenerational Transmission of Risk and Trust Attitudes,” *Review of Economic Studies*, 79, 645–77, 2012.
Duncan, G. J., and D. H. Hill, “Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics,” *Journal of Business and Economic Statistics*, 7, 441–52, 1989.
Eckstein, Z., and O. Lifshitz, “Dynamic Female Labor Supply,” *Econometrica*, 79, 1675–726, 2011.
Eika, L., M. Mogstad, and B. Zafar, “Educational Assortative Mating and Household Income Inequality,” Federal Reserve Bank of New-York State Reports 692, Federal Reserve Bank of New-York, 2017.
Ermisch, J., M. Francesconi, and T. Siedler, “Intergenerational Mobility and Marital Sorting,” *Economic Journal*, 116, 659–79, 2006.
Fernandez, R., N. Guner, and J. Knowles, “Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality,” *Quarterly Journal of Economics*, 120, 273–344, 2005.
Frazis, H., and J. Stewart, “How Does Household Production Affect Measured Income Inequality?,” *Journal of Population Economics*, 24, 3–22, 2011.
Fremeaux, N., “The Role of Inheritance and Labour Income in Marital Choices,” *Population-E*, 69, 495–30, 2014.
Fremeaux, N., and M. Leturcq, “Inequalities and the Individualization of Wealth,” *mimeo*, 2019.
Goux, D. and E. Maurin, “Who Marries Whom in France. An Analysis of The Cohorts Born Between 1934 and 1978,” in H. Blossfeld and Y. Shavit (eds), *Who Marries Whom?* chap. 4, Oxford University Press, Oxford, 2003.
Greenwood, J., N. Guner, G. Kocharkov, and C. Santos, “Corrigendum to Marry your Like: Assortative Mating and Income Inequality,” *American Economic Review-Papers and Proceedings*, 104, 348–53, 2014.
Gronau, R., “Leisure, Home Production, and Work-The Theory of the Allocation of Time Revisited,” *Journal of Political Economy*, 85, 1099–1123, 1977.
Hagneré, C., and A. Lefranc, “Etendue et Conséquences des Erreurs De Mesure dans les Données Individuelles d’enquête: Une Evaluation a Partir des Données Appariées des Enquêtes Emploi et Revenus Fiscaux,” *Economie et Prévision*, 174, 131–54, 2006.
Haider, S., and G. Solon, “Life-Cycle Variation in the Association Between Current and Lifetime Earnings,” *American Economic Review*, 96, 1308–20, 2006.
Ham, J. C., “Estimation of a labour supply model with censoring due to unemployment and underemployment,” *Review of Economic Studies*, 49, 335–54, 1982.
Harmenberg, K., “A Note: The Effect of Assortative Mating on Income Inequality,” *Mimeo*, 2014.
Heckman, J., “Sample Selection Bias as a Specification Error,” *Econometrica*, 47, 153–61, 1979.
House, C., J. Laitner, and D. Stolyarov, “Valuing Lost Home Production of Dual Earner Couples,” *International Economic Review*, 49, 701–36, 2008.
Hryshko, D., C. Juhn, and K. McCue, “Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important Is Assortative Matching?" *Labour Economics*, 48, 168–182, 2017.
Hyslop, D., “Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings,” *American Economic Review*, 91, 755–77, 2001.
Iacovou, M., O. Kaminska, and H. Levy, “*Using EU-Silc Data for Cross-National Analysis: Strengths, Problems and Recommendations*,” ISER Working Paper Series 2012–03, Institute for Social and Economic Research, Colchester, 2012.
Immervoll, H., H. Levy, C. Lietz, D. Mantovani, C. O’Donoghue, H. Sutherland, and G. Verbist, “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), 2005.
Kalmijn, M., “Status Homogamy in the United States,” *American Journal of Sociology*, 97, 496–523, 1991.
Karoly, L. A., and G. Burtless, “Demographic Change, Rising Earnings Inequality, and the Distribution of Personal Well-Being, 1959–1989,” *Demography*, 32, 379–415, 1995.
Kimball, M., C. R. Sahm, and M. D. Shapiro, “Risk Preferences in the PSID: Individual Imputations and Family Covariation,” *American Economic Review-Papers and Proceedings*, 99, 363–68, 2009.
Lecavelier, C., and A. Lefranc, “Siblings Correlations in Socio-Economic Outcomes in France,” *mimeo thema*, 2015.
Lefranc, A., “Educational Expansion, Earnings Compression and Changes in Intergenerational Economic Mobility: Evidence from French Cohorts, 1931–1976,” *Economica*, 85, 808–845, 2018.

Lefranc, A., and A. Trannoy, “Intergenerational Earnings Mobility in France: Is France More Mobile than the U.S.?,” *Annales d’Economie et de Statistique*, 78, 57–77, 2005.
Mare, R. D., “Five Decades of Educational Assortative Mating,” *American Sociological Review*, 56, 15–32, 1991.
Moffitt, R., and P. Gottschalk, “Trends in the covariance structure of earnings in the U.S.: 1969–1987,” *Journal of Economic Inequality*, 9, 439–59, 2011.
Nakosteen, R. A., O. Westerlund, and M. A. Zimmer, “Marital Matching and Earnings: Evidence from the Unmarried Population in Sweden,” *Journal of Human Resources*, 39, 1033–44, 2004.
Ostrovsky, Y., “The Correlation of Spouses,” *Permanent and Transitory Earnings and Family Earnings Inequality in Canada”, Labour Economics*, 19, 756–68, 2012.
Pearson, A., and D. Binder, “The Taxation of Families-International Comparisons 2012,” *Care Research Paper*, 2014.
Pestel, N., “Marital Sorting, Inequality and the Role of Female Labor Supply: Evidence From East and West Germany,” *Economica*, 84, 104–127, 2017.
Ravazzini, L., U. Kuhn, and C. Suter, “Do Opposites Attract? Educational Assortative Mating and Dynamics of Wage Homogamy in Switzerland, 1992–2014,” *Swiss Journal of Sociology*, 43, 567–86, 2017.
Roy, D., “Le Travail Domestique : 60 Milliards D’heures en 2010,” *INSEE Premiere*, 1423, 2012.
Schwartz, C. R., “Earnings Inequality and the Changing Association Between Spouses,” *Earnings”, American Journal of Sociology*, 115, 1524–57, 2010.
Schwartz, C. R., and R. D. Mare, “Trends in Educational Assortative Marriage from 1940 to 2003,” *Demography*, 42, 621–46, 2005.
Shaw, K., “Intertemporal Labor Supply and the Distribution of Family Income,” *Review of Economics and Statistics*, 71, 196–205, 1989.
Solon, G., “Intergenerational Income Mobility in the United States,” *American Economic Review*, 82, 393–408, 1992.
Uunk, W. J. G., H. B. G. Ganzeboom, and P. Robert, “Bivariate and Multivariate Scaled Association Models. An Application to Homogamy of Social Origin and Education in Hungary Between 1930 and 1979,” *Quality & Quantity*, 30, 323–43, 1996.
Zhang, J., and P.-W. Liu, “Testing Becker’s Prediction on Assortative Mating on Spouses Wages,” *Journal of Human Resources*, 38, 99–110, 2003.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web site:

Appendix A Main Variables and Descriptive Statistics
Table A.1: General descriptive statistics
Appendix B Simulation Algorithms
B.1 Addition Randomization
B.2 Imputation Randomization
B.3 Addition Randomization with Sample Selection Correction
Appendix C Double Selection
Table C.1: Double selection model