LEMMA – Document de travail

DT 2023-04

Assortative mating and earnings inequality in South Korea

Nicolas Frémeaux

Université Paris-Panthéon-Assas, LEMMA

SeEung Jung

Inha University

Arnaud Lefranc

CY Cergy Paris Université, CNRS, THEMA

LEMMA EA 4442 - 4 rue Blaise Desgoffe 75006 Paris

Assortative mating and earnings inequality in South Korea

Nicolas Frémeaux* SeEun Jung† Arnaud Lefranc‡§

July 2023

Abstract

We analyze economic assortative mating and its contribution to earnings inequality in South Korea from 1998 to 2018. Our analysis is based on cross-sectional and panel data and accounts for several methodological issues, including measurement error and sample selection bias. Despite a very high level of assortativeness in education, Korea exhibits a negative correlation in earnings between spouses due to low female labor force participation and its negative correlation with male earnings. However, the correlation is large and positive for hourly earnings, among dual-earner couples. Cohort analysis reveals significant changes in earnings correlations, as rising female labor force participation offsets slightly declining educational sorting among younger cohorts. As a result, assortative mating contributes to a very limited extent to inequality between households in observed monthly earnings, but accounts for a sizable fraction, around to 15%, of inequality between household in hourly earnings.

JEL codes: J12, J22, D31.

Keywords: assortative mating, inequality, earnings, education, household, labor supply, South Korea.

*LEMMA — Université Paris Panthéon-Assas, 4, rue Blaise Desgoffe — 75006 Paris (France); nicolas.fremeaux [at] u-paris2.fr

†Corresponding author. Inha University, 100 Inharo, Michuholgu, Incheon, South Korea; jse [at] inha.ac.kr

‡CY Cergy Paris Université, CNRS, THEMA, F-95000 Cergy, France; arnaud.lefranc [at] cyu.fr

§Frémeaux and Lefranc have been supported by the FAMINEQ research project (grant ANR-17-CE41-0007). This paper 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

1 Introduction

The extent of economic assortative mating, *i.e.* the propensity of individuals in couples to share similar economic characteristics, has been examined in several recent papers. They complement the rich sociological literature on the association between spouses in education and social origin by assessing the association in earnings within couples. Indeed, a better understanding of mating patterns is an important step in understanding the distribution income between households in modern societies, as economic homogamy at the household level could largely amplify the inequality that prevails between individuals. While available evidence has documented substantial correlations in spouses' earnings in the case of Western industrialized countries, the evidence for other regions of the world is much more limited. Moreover, the contribution of assortative mating to household inequality is still largely an open question.

In this paper, we study economic assortative mating in South Korea. Our main contribution is twofold. First, we provide descriptive evidence on the correlation of earnings among Korean couples and compare it with evidence from other countries. Second, we examine the contribution of assortative mating to earnings inequality across households, defined as the difference between observed inequality and the counterfactual inequality that would prevail under the assumption of random mating.

Several characteristics make South Korea an important country for the study of economic assortative mating. First, it is one of the largest Asian economies, along with China and Japan. Second, while it shares many similarities with Japan, both economically and socially[1], South Korea differs markedly from most Western societies along important socio-demographic dimensions, including access to education and female participation in the labor market.[2] In particular, despite having one of the highest rates of access to tertiary education among OECD countries (75.7% of the population aged 25-34 in Korea vs. 50.8% on average in the OECD in 2017),

[1] Japan and Korea, for example, have similar GDP per capita ($ 39808 in Japan and $ 33422 in Korea) and Gini index (32.9 in Japan and 31.2 in Korea). Historically, both countries have had relatively low female labor force participation rates (60% in Korea vs. 71% in Japan), low fertility rates (1.4 in Japan vs. 1 in Korea), and high tertiary education rates (63% in Japan vs. 85% in Korea). Both countries are known to have strong assortative mating in terms of education, but studies suggest that it is decreasing over time.

[2] See e.g. OECD (2019) for an overview.

Korea has a low rate of female labor market participation (57. 8% in Korea compared to 60.9% in the OECD in 2017) and one of the highest gender pay gaps (31.1 in Korea compared to 11.9% in the OECD in 2021).[3] Given these specificities, it is highly relevant to assess the contribution of assortative mating patterns to household income inequality in Korea.

A first step in this direction is to document the extent of economic assortative mating, as measured by the correlation between spouses in labor market earnings and other relevant economic attributes such as education. Estimating these correlations raises several methodological concerns. First, spousal earnings correlations are typically estimated using cross-sectional data in which earnings are only observed for a single year. However, earnings are subject to substantial measurement error and transitory shocks. Since these components are weakly correlated between partners, they will lead to an underestimation of the association in partners' long-term earnings. To address this problem, we use the longitudinal data of the Korean Labor and Income Panel Study (KLIPS) and use multi-year averages of individual earnings as proxies for permanent earnings components, in the spirit of Solon (1992).

A second methodological concern is that observed earnings reflect endogenous household labor supply decisions, which may complicate the assessment of within-couple assortiveness. Along the intensive margin, decisions to work part-time may call into question the relevance of monthly earnings as a measure of individual economic characteristics. Along the extensive margin, decisions to remain out of the labor force may introduce sample selection bias into the assessment of economic homogamy. Many studies often ignore this issue by using zeros as the relevant measure of earnings for individuals not in the labor force, although this ignores the fact that these individuals have latent earnings potential and could contribute to household welfare through, for example, domestic production. In this study, we explicitly address this issue. In particular, we pay special attention to the analysis of hourly earnings, which, in contrast to monthly earnings, are largely independent of the couple's joint labor supply decisions and can be seen as a better proxy for potential earnings. The use of longitudinal data also allows us to observe earnings over time and to minimize the incidence of temporary non-participation.

[3] For more information, see the OECD Gender Equality Database https://www.oecd.org/gender/data/employment/

Finally, we explicitly account for sample selection due to non-participation and provide estimates of the within-couple correlation in (possibly latent) hourly earnings that correct for sample selection using the methodology proposed in Frémeaux and Lefranc (2020).

The main difficulty in assessing the contribution of assortative mating to economic inequality between couples is to define a credible counterfactual of the income distribution that would prevail if individuals mated randomly into couples. Two main approaches have been proposed in the literature: the *addition randomization* approach (e.g. Hryshko et al., 2014)[4] and the *imputation randomization* approach (e.g. Greenwood et al., 2014)[5]. The two approaches differ in their treatment of endogeneity and unobserved heterogeneity. The addition approach assumes that female labor supply does not adjust to matching, while we find that female labor force participation is strongly associated with male earnings.The imputation approach rules out unobserved heterogeneity conditional on observed educational characteristics, which is questionable given the very high degree of educational assortative mating found in South Korea. To carefully assess the contribution of assortative mating to inequality in South Korea, we complement these two approaches with the *simulation randomization* approach developed by Frémeaux and Lefranc (2020), which allows for unobserved heterogeneity and endogenous participation decisions.[6]

Our results indicate a high degree of assortative mating in South Korea. This is particularly the case for education, where the correlation is among the highest in developed countries, despite a declining trend in recent cohorts. The correlation is also high for earnings. For dual-earner couples, the correlation of monthly earnings is around 0.25-0.30 and stable over the period, and slightly higher when we look at hourly earnings. However, due to a negative correlation between female participation and male earnings, the within-couple earnings correlation is negative when considering all couples, including those where one spouse has zero earnings. Longitudinal analysis reveals a decline in the earnings correlation over the life cycle of couples, driven by women's withdrawal from the labor market, which is indicative of intra-household specialization. The cohort analysis shows that the decreasing educational sorting among younger cohorts is compen-

[4]See also Burtless (1999) and Aslaksen et al. (2005).

[5]See also Eika et al. (2019), and to some extent Pestel (2017)

[6]See below, section 6.1, for more details.

sated by the increasing female labor force participation. Finally, measurement error and sample selection issues affect these results.

Overall, assortative mating contributes to increase earnings inequality between households, although the extent of this contribution varies depending on the earnings measure and the counterfactual scenario. Using monthly earnings, the contribution of assortative mating appears to be limited. The imputation and simulation methods indicate that random mating would lead to slightly lower inequality (between -4 and -6% on average over the period), while the addition randomization method leads to slightly higher inequality (+3%). This limited effect is mainly due to the fact that the strong sorting by education between spouses is compensated by the low participation rate of women. However, when using hourly earnings, the effect on assortative mating appears to be much stronger: under the assumption of random mating, inequality would be much lower, around -8% for the Gini index and up to 17% for the Theil index. The simulation method leads to a larger reduction in inequality. The impact of assortative mating on inequality is quite stable over time.

The rest of the paper is organized as follows. Section 2 provides a review of the literature. Section 3 presents the data. We then study assortative mating in section 4 before analyzing sample selection issues in section 5. Section 6 estimates the impact of assortative mating on inequality.

2 Related literature

In recent years, a growing number of studies have analyzed the extent of economic assortative mating and its consequences for inequality. For the US, they provide evidence of a substantial correlation, around 0.2, in spouses' earnings (in particular Burtless, 1999; Schwartz, 2010). This issue has also been studied in a number of European countries, including Sweden (Nakosteen et al., 2004), Germany (Pestel, 2017; Eika et al., 2019), the United Kingdom, Norway and Denmark (Eika et al., 2019), Switzerland (Kuhn and Ravazzini, 2017) and France (Frémeaux and Lefranc, 2020). Overall, these studies indicate a significant degree of assortiveness in economic characteristics, although lower than for education (Fernandez et al., 2005). Cross-country variation typically reflects overall differences in inequality and social mobility and stratification.

The contribution of economic assortative mating to income inequality and its trends has only recently been studied and no consensus has emerged. Comparability of results across studies is hampered by differences in the methodology[7], the concept of income, or the measure of inequality used in the analysis.

Most US studies report a limited contribution of assortative mating to income inequality. Eika et al. (2019) find that income inequality in 2013 was 5% higher than in the counterfactual where spouses were randomly matched. Harmenberg (2014) finds similar results, while Greenwood et al. (2014) reports an estimate of 2%. Focusing on changes over time rather than levels, Ciscato and Weber (2020), Dupuy and Weber (2022), and Grossbard et al. (2022) find that the level of inequality would be about 5% lower if mating patterns had not changed since the 1960-1970s. However, Schwartz (2010) reports a much larger contribution of about 25% to 30% of inequality. In Europe, the influence of assortative mating varies across countries. In Germany (Pestel, 2017) and France (Frémeaux and Lefranc, 2020), assortative mating is found to contribute 10 to 20% of the observed inequality. However, Eika et al. (2019) find limited effects in the case of Denmark, Sweden, and the United Kingdom.

Inequality and assortative mating in South Korea, and in Asia more broadly, have not been extensively studied empirically. In terms of the level of income inequality, South Korea is slightly above the OECD average, with a Gini index for disposable income of 0.331 compared to an OECD average of 0.312 in 2019. Few studies have examined trends over time. The results are mixed and vary depending on the data source. Kim and Kim (2015) construct long-term series of top income shares in South Korea using income tax statistics from 1933 to 2010. They find that top income shares declined sharply after World War II, remained low during the industrialization period, and increased since the mid-1990s. However, other studies find that inequality in Korea declined after 2008 (Kim, 2011; Teichman, 2015).

Existing evidence suggests a high degree of social assortative mating in South Korea relative to other countries. The seminal article of Johnson et al. (1976) examined the association of

[7] See discussion in section 6.1 below.

cognitive ability within Korean couples and found a correlation coefficient of 0.72 in 1973-1974, a much higher measure than that typically found in Western countries. Several authors have explained this extremely high correlation by the high prevalence of arranged marriages in the Korean society. In the custom of arranged marriages, which was quite common in the 1960s and 1970s, matchmakers often arranged marriages between individuals of similar ability, parental social status, and family reputation. The later study by Hur (2003) finds, similar to Western evidence, a large positive correlation within couples in terms of educational level ($\rho = 0.63$) and religious affiliation ($\rho = 0.67$), but reports a lower correlation in cognitive ability than Johnson et al. (1976). Hur (2003) interprets the decrease in cognitive ability association as a result of the increase in free marriages in the Korean society. Park and Smits (2005) assess the evolution of educational assortative mating in South Korea from 1930 to 1998. They find that assortativeness and stratification increased significantly over this period. In particular, they show that the boundary between the educational elite and the rest of Korean society has become more pronounced since the end of the 20th century. Several authors have suggested that this may be due to an increased preference among Korean men for a highly educated spouse who can help their children achieve better results in Korea's competitive education system. To our knowledge, there are no studies of assortative mating along economic traits in Korea, nor of the influence of assortative mating on inequality.

A few studies have also examined assortative mating in other Asian countries, such as Japan and China. The results are somewhat mixed. Using census and household survey data between 1990 and 2009 in China, Nie and Xing (2019) finds that positive educational assortative mating has increased significantly since the early 1990s, which is in line with Han (2010). In Japan, on the other hand, Fukuda et al. (2021) shows that the strength of the association for educational homogamy remained constant between 1980 and 2000, and then declined significantly in 2010. Smits and Park (2009) examine the trend in educational assortative mating in 10 East Asian countries over five decades and assert that homogamy has declined during the process of modernization, while educational homogamy among the less educated has remained stable.

3 Data

3.1 Dataset and sample selection

Our data come from the Korean Labor & Income Panel Study (KLIPS) published by the Korean Labor Research Institute. KLIPS is a longitudinal survey of a representative sample of South Korean households and individuals living in urban areas. It has been conducted annually since 1998.[8] The 1998 sample consists of 5,000 households living in South Korea and includes all household members aged 15 and older (13,321 individuals). To maintain statistical representativeness despite sample attrition, 1,415 households were added to the panel in 2009, resulting in a total sample of 6,721 households. In 2018, 5,044 households were added again.[9]

Attrition bias in KLIPS has been examined in several articles. Examining the first six waves of KLIPS, Lee (2005) finds that non-random attrition does not cause any severe bias in the estimation of key individual economic outcomes. Lee (2005) and Lee et al. (2011) both report polarized attrition, with higher rates of attrition for both individuals with high education or earnings and unemployed households. However, attrition appears to be limited among couple households. Kim (2010) finds that attrition causes a 2% underestimation of the wage gap between regular and irregular workers, and notes that the bias is smaller than expected. Finally, the attrition bias appears to diminish over time. Overall, the literature suggests that the potential bias is likely to be limited for the purposes of our study.

The available KLIPS data cover 22 waves from 1998 to 2019. When focusing on annual information (e.g., annual earnings measure), we restrict our analysis to the three waves where sample representativeness is ensured by design: 1998, 2009, and 2018. Other waves are used when conducting longitudinal analysis for individuals observed in these three survey waves.

In each of these three waves, we restrict the sample to married couples[10] in which both

[8]The urban population, which is the focus of KLIPS, represents 92% of the total Korean population. The omission of households living in rural areas may cause limited bias in this paper since we focus on couples under or equal to 65 years of age: according to Statistics Korea, only 2.2% of the population aged 20 to 65 will live in rural areas in 2020.

[9]The 2009 sample includes 3,657 original 1998 households, 1,649 branch households that were separated from the original 1998 households, and 1,415 newly added households in 2009. The retention rate of the original 1998 sample in 2009 was 74%, and the retention rate of the newly added 2009 sample in 2017 was 84%.

[10]Unmarried couples cannot be identified in KLIPS. However, their number is very low in South Korea, about 1.4% of all couples, according to Lee (2008). Unmarried cohabitation is generally temporary before marriage.

spouses are between 25 and 65 years old at the time of the survey, and in which neither partner is out of the labor force due to retirement, education, or military service. We also drop couples in which neither spouse has positive earnings.[11] This results in samples of 2,706 households in 1998, 3,174 in 2009 and 4,690 in 2018.

3.2 Variables of interest

KLIPS collects information on household characteristics, employment, labor mobility, income, expenditures, education, and training. In this paper we rely on three main variables. Income is defined as average monthly wages and salaries for employees and as self-employment income for the self-employed. Education is defined as the highest educational attainment (no high school, high school diploma, some college, and college degree) and the number of years of education (estimated from the diploma). Weekly hours worked are defined as the regular number of hours worked per week and overtime for regular employees and the self-employed, and as the average number of hours worked for non-regular employees. Individuals are defined as participating in the labor market if they report positive earnings.[12]

We consider two measures of earnings: *monthly earnings* and *hourly earnings* (defined as the ratio of weekly earnings to the number of hours worked).[13] For monthly earnings, we consider two samples : all couples (including those reporting zero earnings) and the dual-earner subsample. For hourly earnings, we focus only on dual-earner couples. Earnings and hours worked are self-reported variables in the current survey year. All values are expressed in 2015 Korean Won (KRW) and deflated by the Consumer Price Index.

There are several advantages to using hourly wages as the variable of interest. First, unlike

According to the OECD Family Database (2018), only 2% of Korean children are born from an unmarried mother against 41% on average in OECD countries.

[11]This restriction is consistent with other papers such as Eika et al. (2019) and Frémeaux and Lefranc (2020). The main reasons why men between 25 and 65 do not report earnings are old age, unemployment, physical/mental illness, and other unspecified reasons.

[12]We do not use self-reported employment status because we found inconsistencies for couples in which both spouses are self-employed. In fact, for a small fraction of these couples, the wife was employed but did not report earnings.

[13]We first convert "current year monthly earnings" to weekly earnings and then divide them by the "current average weekly hours worked". We use hourly earnings (or wage rate) rather than full-time equivalent earnings (wage rate × the typical full-time hours worked). There is little difference between the two concepts. The variance in the number of hours worked is greater for the self-employed. For example, in 2018, about 90% of male employees work between 40 and 60 hours per week. This share is 73% for self-employed men.

monthly earnings, it is less affected by joint labor supply decisions, especially along the intensive margin. In particular, it allows us to adjust for potential differences in the number of hours worked between spouses and across households. Admittedly, one might suspect a wage penalty for part-time work, which is ignored in this approach. Moreover, labor supply decisions along the intensive margin may also shape later career prospects in ways that affect husbands and wives differently within couples. However, it is difficult to account for career involvement empirically.[14]

Second, hourly earnings are similar to full-time equivalent (FTE) labor market earnings, up to a rescaling. Hourly earnings can thus be understood as a measure of earnings *potential*. Using hourly earnings also allows the contribution of non-participation to couple welfare (and inequality therein) to be taken into account. This is one of the reasons why we explicitly account for sample selection due to non-participation (section 5).

We apply several corrections to minimize the impact of measurement error on our estimates. First, we winsorize the data at the bottom and top 1% of the monthly and hourly earnings distributions when earnings are positive.[15] Second, in computing hourly wage rates, we top-code the number of hours worked per week at 90 hours per week.[16] The share of individuals working more than 90 hours is 1.6% for men and 0.8% for women. Despite this correction, we still obtain low hourly earnings for some individuals. We drop observations where the hourly wage is less than 50% of the minimum wage. If it is in the 50-100% range, we replace the observed wage with the minimum hourly wage.[17]

Cross-sectional observations of individual earnings and wages are known to reflect a combination of permanent and transitory components. The latter include a mixture of genuine

[14]If we take full-time work as a career commitment and re-estimate the correlation in hourly earnings on the subsample where both spouses work full-time, we find slightly higher correlations (+5% in 1998, +7% in 2009, and +12% in 2018).

[15]According to Bollinger and Chandra (2005), winsorizing performs better than trimming in the presence of response errors. When reported earnings are zero, the value is left unchanged. Estimates based on winsorized data appear to be more consistent across specifications and years, as shown in table B.3 in the appendix.

[16]The legal maximum workweek was set at 68 hours per week until 2018. We allow for a higher number of hours due to the large share of self-employed workers in our sample, who are not subject to legal working hours limits. This does not significantly affect our results.

[17]This correction affects a limited number of individuals: in our main sample (1998, 2009, and 2018), we drop 130 individuals (0.7% of men and 0.5% of women) and replace the observed hourly wage of about 1200 individuals (6% of men and 5% of women). Most of the affected observations are in 2009 and 2018 due to the significant increase in the minimum wage in Korea. The self-employed are over-represented among those affected by this correction. When we do not apply these corrections, the results are barely affected (see Appendix).

transitory shocks and measurement errors, as is widely documented in labor economics.[18] As noted by e.g. Ostrovsky (2012), transitory earnings components may be partially correlated across partners for reasons related to local business cycles or industry-level shocks in the case of partners working in a similar industry, although possibly less so than permanent components. One way to deal with possible measurement error bias is to take advantage of the longitudinal dimension of the KLIPS dataset and compute multi-year average earnings. We compute average earnings over a period of two and three years from our main observation years: 1998-2000, 2009-2011, and 2018-2019.[19] For monthly earnings excluding zeros and hourly earnings, non-positive values are treated as missing. In other words, for an individual observed over three years who reports zero earnings in one wave, we only estimate the average earnings over the two years in which the individual's earnings are positive. However, for the multi-year average of monthly earnings, including zeros, we retain all available observations.

3.3 Descriptive statistics

Table 1 presents the main descriptive statistics for our sample of couples in our three main years of analysis: 1998, 2009, and 2018. We find two main changes in South Korea over the period 1998-2018. First, the average number of years of education increases by 2.3 for women and 1.6 for men between 1998 and 2018. Second, labor force participation changes significantly, especially for women. While less than 30% of married women aged 25 to 65 participated in the labor market in 1998, this share is 47% in 2017. For men, however, participation remains stable (over 95%). As a result, the share of dual-earner couples increases mechanically by 20 percentage points, reaching 44% of couples in 2018. The number of hours worked decreases significantly for working women (from 54 to 41 hours per week) and, to a lesser extent, for working men (from 54 to 48 hours).[20] The share of full-time workers increases slightly for men and decreases slightly for women. The distribution of employment status also changes, as we observe a decrease in the

[18] See e.g. Solon (1992) and the survey by Black and Devereux (2011) for a discussion in the case of intergenerational studies.

[19] We also compute average earnings using different windows (from 2 to 5 years). See Table B.1

[20] Between 1998 and 2019, there were several policy changes regarding working conditions in South Korea that led to a decrease in working hours. For example, in 2000, regular working hours were officially reduced from 44 to 40 hours per week; in 2004, regular working days were reduced from six to five days per week. The maximum working week could be extended to 68 hours before the 2000 reform and 52 hours after, subject to overtime pay.

share of self-employment, especially for men. The share of couples in which both spouses are self-employed falls by more than 6 percentage points since 1998 and represents 8% of all married couples in 2018.

4 Assortative mating in education and earnings

4.1 Cross-sectional analysis

We first assess the degree of assortativeness along two dimensions, education and earnings, based on annual cross-sectional declarations.

Education Following Eika et al. (2019), we first examine the share of homogamous couples, i.e. couples in which both spouses have similar education. We consider four categories of education, shown in Figure 1a. The share of homogamous couples is stable at around 60% between 1998 and 2018. However, the composition of homogamous couples changes over the period, partly reflecting changes in the overall distribution of education in our sample. The share of couples in which both spouses have a high school diploma or less decreases significantly, while the share of couples with a college degree increases as a result of rising educational attainment in South Korea. To account for this changing distribution, we rely on the SM assortativeness index introduced in Chiappori et al. (2020), which is defined as the percentage of homogamous couples normalized by the percentage that would be observed under random mating. Our estimates are shown in Figure 1b. The SM index increases from 1.89 in 1998 to 2.09 in 2009 and reaches 1.94 in 2018. This indicates a slightly increasing and positive assortative mating in education, since under random mating the index would be equal to 1.

We also estimate rank correlations on the number of years of education to provide a synthetic measure of educational assortative mating. The results are shown in Table 2. On average, the correlation is very high, around 0.7. This confirms the findings in Fernandez et al. (2005) and Smits and Park (2009) that South Korea is among the countries with the highest levels of educational assortative mating. These estimates are close to those observed in Latin American countries and in China, Hong Kong, and Thailand, and significantly larger than those observed in the United States and most European countries. The correlation decreases slightly over the period from 0.74 to 0.68. Overall, while the various indicators confirm the high degree of educational assortative mating in Korea, they do not show a consistent trend over time.

Earnings To assess the extent of assortative mating in earnings, we estimate Pearson correlations within couples using three different measures: monthly earnings for the entire population of couples, including those in which one spouse reports zero earnings; monthly earnings for the subsample of dual-earner couples, i.e., couples in which both spouses report positive earnings; and hourly earnings, also for dual-earner couples only. The results are presented in Table 2 and Figure 2a.[21]

First, the correlations computed on the full sample of couples and including zero earnings are negative for all three observation dates, but they are small in magnitude and in most cases not significantly different from 0. This seemingly very low level of assortative mating in earnings is contradicted by the estimates based on dual earners. For this subsample, the correlation in monthly earnings is 0.27 on average over the three years of observation. Estimating correlations in hourly earnings allows to further control for endogenous labor supply decisions, such as the decision to work part-time, and to remove the correlation in hours worked that may underlie the correlation in monthly earnings. The correlation in hourly earnings is slightly higher (around 0.30) than that based on monthly earnings, but the difference is not statistically different. The fact that part-time employment is limited for both men and women in South Korea explains the small difference between the two measures.

For all three measures, within-couple earnings correlations show a rather flat profile over the years, with a moderate increase between 1998 and 2009 and a slightly declining trend after 2009. An analysis of the pattern of female labor force participation allows us to understand the difference in earnings correlations observed between the full sample and the dual-earner subsample. Figure 3 shows the female participation rate. In 1998, the low correlation of earnings between spouses is explained by two factors. First, about 3 out of 4 women report zero earnings, imply-

[21] We present only the estimates based on the Pearson correlations in the core of the paper. If we replicate the analysis using Spearman rank correlations, the results are similar.

ing for these individuals a zero correlation with their husband's earnings. Second, while human capital is positively correlated within couples, female participation is negatively associated with the male earnings decile, contributing to the overall weakly negative correlation in earnings. Over time, two main changes occur, the joint effect of which is close to zero on the correlation in monthly earnings for the full sample: first, the average participation rate increases, leading to an increase in the correlation, all else equal; second, the gradient of female participation along the male distribution becomes more negative, leading to a decrease in the correlation.

Finally, figure 2b also shows estimates for correlations based on residual earnings after controlling for age and education. Residual correlations are always lower than gross correlations. For the full sample, the correlation becomes more negative, with a coefficient of about -0.1. For dual-earner couples, the correlation is low and not statistically different from zero at conventional levels, but becomes positive and significant, slightly below 0.15 in 2009 and 2018. Thus, while age and education capture most of the initial earnings correlations in 1998, this is much less the case in 2009 and 2018, suggesting that there remains significant sorting along other dimensions not captured by these variables.

4.2 Longitudinal analysis

To purge our estimates of the earnings correlation from the influence of transitory components and measurement error, we take advantage of the longitudinal dimension of the KLIPS dataset and compute multi-year average earnings. For each of our reference years, we average earnings over two- or three-year periods (1998-2000, 2009-2011, and 2018-2019) for those couples who remain observed in the KLIPS data and meet our sample selection rules.[22]

The estimates are shown in Table 2 and Figure 2c. For all couples, including those with zero earnings, the correlations in multi-year average earnings are little different from those estimated using annual observations of earnings, suggesting that female labor force withdrawal is a persistent phenomenon. For dual-earner couples, the correlation in multi-year averages is slightly higher than for single-year earnings, averaging 0.28 versus 0.25 for annual earnings. The

[22] When we compute single-year correlations for the restricted sample of individuals observed for two or three years, the results appear very similar to those presented in Panel A of Table 2.

gap between the two estimates is larger for hourly earnings, with estimates based on multi-year averages about 20% higher than for single-year measures.[23] In line with previous results in the literature, this confirms the higher incidence of transitory shocks and measurement errors in measures of hourly earnings.[24]

Following Nakosteen et al. (2004), we also use the longitudinal dimension of the data to assess changes in earnings correlations over the life cycle of couples. A potential limitation of our cross-sectional estimates is that couples are observed at different points in their lives. In addition, we face potential selection problems since we only observe surviving couples. Therefore, we restrict our sample to couples who are observed for at least 10 years. We estimate the earnings correlation in each year during the 10 years after the couples entered our sample. The results are shown in Figure 4. When we consider all couples (4a), we find a small and non-significant decrease in the correlation over the 10-year period. This may be due to the fact that we are mixing couples observed at different stages of their life cycle. When we focus on younger couples (4b), the downward trend is more pronounced with a significant gap (at the 90% confidence level). However, if we restrict our analysis to dual-earner couples, we do not find a similar downward trend in the earnings correlation (4c and 4d). This result is consistent with specialization after marriage at the extensive margin, as women withdraw from the labor market.

4.3 Cohort analysis

Because of the KLIPS sampling structure, the cohort composition of our sample of couples evolves over time. To assess cohort trends in assortative mating, we refine our analysis by considering four different cohorts defined on the basis of the wife's year of birth: born before 1955, between 1955 and 1964, between 1965 and 1974, and after 1974.[25] These cohorts differ in several respects. In particular, more recent cohorts have higher educational attainment, higher labor force participation, and lower fertility.

[23] In table B.1, we estimate correlations using different windows for estimating the multi-year average: 1, 2, 3, and 5 years. The correlations increase as the window increases to 3 years and then stabilize for hourly earnings and slightly decrease for monthly earnings.

[24] See, e.g., Bound et al. (1994).

[25] Using the husband's year of birth does not affect the results because of the stability of the age gap between spouses over time (about 2.5-3 years).

Table 2 (Panel C) and Figure 5 report measures of assortative mating in terms of education and earnings by cohort. Educational assortative mating shows a hump-shaped pattern across cohorts (5a). The Spearman correlation coefficient falls from around 0.7-0.75 for those born before 1964 to 0.58 for the most recent cohort. This trend is statistically significant at the 95% level. It could be explained by the decline in arranged marriages discussed in Hur (2003).

Correlations in multi-year average earnings show opposite trends, depending on the earnings measure and the sample (Figure 5b). For monthly earnings, including zeros, the correlation shows a statistically significant increase from -0.14 for the cohort born before 1955 to 0.03 for the cohort born after 1974. For dual-earner couples, the correlation is larger for the cohort born before 1955, about 0.4 (or 0.5) for monthly (or hourly) earnings, and stable in later cohorts, about 0.3 for monthly earnings and 0.36 for hourly earnings.

It may be objected that our analysis is affected by life-cycle bias, as cohorts are observed at different ages (Haider and Solon, 2006). In the appendix, we attempt to proxy for permanent earnings in two ways: by using the multi-year average of earnings between ages 30 and 50, and by retaining only the earnings observation closest to age 40.[26] While these robustness checks are not precisely estimated due to sample limitations, they support the results discussed above.

5 Accounting for sample selection

The fact that a very large proportion of women in our data are not in the labor force raises the issue of sample selection bias. For inactive women, the assumption that their contribution to household resources is zero ignores both household domestic production and the leisure value of time. However, in the likely case that participation decisions depend on the potential earnings of both partners, the sample of dual earners is no longer representative of the population as a whole, and the correlation cannot be assessed directly on the basis of observed earnings alone.

To account for sample selection, we rely on the methodology introduced in Frémeaux and Lefranc (2020). Their model provides a regression-based estimator of the earnings correlation, corrected for the endogeneity of the participation decision using the Heckman (1979) approach.

[26]See Figure B.1.

We present the details of their approach in the Appendix A and summarize its intuition here. The main equation of interest is the simple bivariate regression of female earnings on male earnings within couples:

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

As is well-known, the regression slope in this model is related to the earnings correlation by the following equation: $\rho = \beta\sigma_m/\sigma_f$, where σ denotes the standard-deviation and m, f respectively index male and female within couples. Due to sample selection, OLS provide a biased estimator of β, in equation 1 above. Furthermore, since the distribution of female earnings is truncated, as a result of selection into participation, the usual estimator will also provide a biased estimate of the standard deviation of female earnings σ_f. However, both β and σ_f can be consistently estimated using Heckman's sample selection correction model. Frémeaux and Lefranc (2020) thus suggests to rely on these estimators to provide a consistent estimator of the within-couple earnings correlation given by $\hat{\rho} = \hat{\beta}^{Heckman} \times \sigma_m/\hat{\sigma}_f^{Heckman}$.

In our estimation of equation 1, we focus on hourly earnings, net of age and time effects. As discussed earlier, we believe that compared to monthly earnings, hourly earnings are a better proxy for *potential* earnings because they are more independent of the couple's joint labor supply decisions. Our estimates are based on the Heckman selection model estimated using maximum likelihood on our full sample of couples, separately for each reference year (1998, 2009, 2018). The dependent variable of the selection equation is a dummy variable that equals one if we observe strictly positive earnings of the wife. The selection equation includes controls for characteristics of the wife (age, education, number of children, and age of oldest child) and of her husband (age, education, log hourly earnings, experience, and self-employment status). The full results are presented in Appendix A and show that selection is positively associated with the wife's education and negatively associated with the household's number of children and the husband's earnings.[27]

Since our focus is on the overall correlation in earnings, the main equation of the Heckman model is simply the bivariate model of equation 1, with no additional controls except that

[27] Table A.1.

hourly earnings are net of age and time effects. Both earnings variables are measured in log.[28] Table 3 and Figure 2d present estimates of the hourly earnings correlation with and without correcting for sample selection, along with associated moments. While derived from a log-linear OLS regression model, the estimates in Panel A are close to those reported in Table 2 (Panel A), ranging from 0.30 to 0.36 for the correlation in logarithm and from 0.27 to 0.33 for the correlation in level. Panel B reports correlations corrected for sample selection. These correlations are about 10% lower than the uncorrected correlations for the years 1998 and 2009. However, the correlations estimated for 2018 are very close in both cases and slightly higher when sample selection is taken into account.

Overall, these results suggest, first, that using the sample of dual-earner couples to estimate the correlation of earnings potential within the population of Korean couples introduces a limited sample selection bias. Second, since the bias changes sign over time, correcting for sample selection slightly reinforces the hump-shaped trend already observed in Figure 2. [29]

We confirm the relevance of our sample selection model using the longitudinal dimension of our data. Our model allows us to predict hourly earnings for individuals who do not work at all during the time they respond to the survey. Alternatively, we can also use the longitudinal dimension of the dataset to infer hourly earnings for individuals who report positive earnings at least once. In a first approach, we replace the missing values of the hourly wage rate with the average hourly wage rate over the period (estimated from the available observations). Second, we replace the missing values with the closest available observations. Indeed, the average hourly wage may not be representative for the missing year(s) if the characteristics of the individual and the macroeconomic context have changed over time.[30]

Figure B.2 in the appendix compares the hourly earnings correlations obtained using these

[28] The normality assumption of the ML estimator thus requires that the joint distribution of spouses' earnings is a bivariate log-normal distribution, which is a reasonable assumption in the case of the hourly earnings distribution.

[29] Appendix A extends these estimates to the correlation in monthly earnings and yields similar results: using dual-earner couples to assess the latent correlation in potential earnings within Korean couples leads to a slight overestimation of the true correlation at the beginning of the period (1998, 2009) and an underestimation at the end of the period (2018), but the bias appears small in absolute value, around 10%, except for 1998, where it reaches around 20%.

[30] This procedure allows to double the number of observations compared to the unadjusted estimates. In total, 1998, 2009, and 2018 allow us to use 3,543 observations for hourly earnings correlations without correction and 6,962 when using the nearest observation or the average.

different approaches. For the full sample, ignoring sample selection leads to an overestimation of the hourly earnings correlation. The correlation after correcting for sample selection is on average 12% lower (as opposed to 8% for the full sample). Compared to the uncorrected estimates, the correlation is 3% lower if we use average hourly earnings as a reference for the missing values and 9% lower if we use the closest available observation. The gap between these estimates narrows over time, and by the end of the period all three correction methods are remarkably similar. Overall, these alternative methods confirm the results found with the sample selection correction model and show that, if anything, the sample selection corrected estimate provides a lower bound on the earnings correlation.

6 Contribution to earnings inequality

We now have a clearer picture of the degree of assortative mating in South Korea. We have highlighted significant sorting by education, which leads to a positive correlation in potential earnings, despite the negative correlation between wages and labor force participation. We also pointed to some changes across cohorts and selection effects. Our goal now is to estimate the contribution of assortativeness to earnings inequality between couples.

6.1 Methodology

We rely on three different approaches discussed in Frémeaux and Lefranc (2020), which we summarize here. All three approaches have in common that they differentially assess the contribution of assortative mating by comparing the level of household income inequality observed in the actual situation with the level that would prevail under a counterfactual scenario of random mating. They differ in the construction of this counterfactual distribution.

The *addition randomization* approach consists of taking observed individual earnings as a fixed individual characteristic and randomly mating individuals into simulated couples.[31] Household income is the sum of the income of both partners in simulated couples. In this case, the counterfactual distribution is simply a convolution of the individual marginal earnings distribu-

[31] See, for example, Hryshko et al. (2014).

tion of the female and male partners as observed in the population.

In the *imputation randomization* approach, individuals are characterized by observable characteristics, such as education. The total earnings of a household are determined by the characteristics of both partners. A conditional household income distribution can be computed for each combination of partner characteristics. Randomization consists of creating simulated couples in which the partners' conditional *characteristics* are randomly drawn from the observed distributions of those characteristics in the population.[32] The counterfactual distribution is thus a mixture of the observed conditional earnings distribution, with the mixture weights defined by the random mating hypothesis.

The main limitation of the addition approach is that it assumes that female labor supply does not adjust to rematching, while we find that female labor force participation is strongly associated with male earnings. In comparison, the advantage of the imputation approach is that it allows for endogenous labor supply responses, but only to the extent that they depend on the conditioning variables. However, a major limitation of this approach is that it excludes selection into couples based on unobservable characteristics. In fact, the imputation approach assumes that heterogamous couples are a good counterfactual for the behavior of individuals observed in homogamous couples if these individuals were reassigned to heterogeneous partners.

The third approach, which we call the *simulation randomization* approach, is based on a parametric model of the bivariate distribution of spouses' earnings under sample selection, as developed by Frémeaux and Lefranc (2020) and discussed above. It focuses on the inequality of potential joint earnings across couples and attempts to overcome the limitations of the addition and imputation approaches. Under the assumption of joint log-normality, the joint distribution of partners' earnings among observed couples depends on three parameters: the variance of earnings in the marginal distribution of both men and women, and the covariance of earnings within the couple. The estimated parameters allow us to assess the uncensored distribution of potential household earnings, even though hourly earnings are unobserved for couples where one partner is not employed. It is also possible to simulate the joint distribution of male and female earnings under the assumption that the correlation of partners' earnings is zero, holding

[32] See, e.g., Greenwood et al. (2014).

constant the characteristics of the marginal distributions. We can then compute the degree of inequality in joint earnings associated with both the actual parametric distribution estimated on observed data and the simulated distribution under random mating.

The procedure of the simulation approach is based on the assumption that individual earnings (potential) are unaffected by possible re-mating. In particular, the variance of the marginal distribution of male and female potential earnings is assumed to be unaffected by mating patterns. Of course, as a result of random rematching, the pattern of sample selection would change, as would the variance of earnings in the observed distribution.

Part of the correlation of economic outcomes within couples is driven by the fact that partners are homogeneous with respect to their birth cohort. This cohort-wise homogamy would likely persist even if social and economic characteristics did not determine partner choice. Consequently, in all three approaches, we only allow mating to occur conditional on the age of both partners.

6.2 Results

Our estimates of the effect of assortative mating on earnings inequality are reported in Table 4.

To compare the observed and counterfactual distributions, we use two inequality indices: the Gini index, which is widely used in the literature, and the Theil index, which is more sensitive to inequality in the tails of the distribution. The analysis is carried out for both hourly and monthly earnings.

Before commenting on the contribution of assortative mating to earnings inequality, it is useful to discuss recent trends in earnings inequality in Korea and its contribution to overall inequality. In our sample, earnings inequality among married couples increased between 1998 and 2018. The Gini index of monthly earnings was 0.37 in 1998 and reaches 0.27 in 2018. The decline in inequality is more pronounced for the Theil index, from 0.24 to 0.12, and for hourly earnings.[33] Our analysis focuses on pre-tax income inequality, which is typically an important driver of overall income inequality. In Korea, the share of capital income is relatively low (5% in

[33]See table 4. Part of the decline in inequality can be attributed to the rapid growth of the minimum wage in Korea (e.g., +16% in 2018) and changes in working time regulations that require overtime compensation.

2007 and 7% in 2018), while labor income is the main source of income (80% in 2007 and 79% in 2018).[34] According to Kwack and Lee (2007), variations in the Gini coefficient of gross income are indeed closely related to those of labor income. Although capital income inequality fluctuates widely, its contribution to changes in the Gini coefficient of gross income distribution is small. Moreover, taxes and transfers seem to play a modest role in Korea, as the difference between pre-tax and post-tax inequality is among the lowest in OECD countries.[35] Therefore, our analyses can be extended to discuss the role of assortative mating for overall income inequality without loss of generality.

Hourly earnings Panel A of Table 4 presents inequality measures for hourly earnings in the observed data and in the counterfactual scenarios. Line (1) reports the value of the Gini and Theil indices computed from the observed data on the sample of dual-earner couples. Lines (4) and (6) report inequality under the addition and imputation randomization approaches, respectively. Estimates are given for each reference year, and the last two columns show the unweighted average of each index across years. The addition and imputation approaches both predict that random mating would lead to significantly lower inequality between couples. In both cases, the value of the Gini index would be about 2 points, or 8%, lower than observed. The decrease in inequality would be more pronounced according to the Theil index, with a decrease of about 11 to 15%.

These results are consistent with the high correlation in hourly earnings reported in the previous section. An obvious limitation is that they may lack representativeness since they are estimated for the sole sample of dual earners. One of the advantages of the simulation randomization approach is that it allows one to assess the degree of inequality in the full joint distribution of (partially latent) hourly earnings and to assess the effect of random mating on the entire sample of couples, not just dual-earners. Line (2) uses estimates of the parametric distribution of hourly earnings to compute the level of inequality in the latent joint distribution (i.e., what would prevail in the absence of selection into participation) based on observed mating

[34] OECD Income Distribution Database

[35] While the Gini coefficient falls by 27% on average in the OECD as a result of redistributive fiscal policies, this reduction is only 11% in Korea in 2018. See OECD Income Distribution Database.

patterns. On average, the inequality is very similar to the direct measure based on dual earners. If we further simulate random mating, inequality would fall by about 8% for the Gini index and 17% for the Theil index.

Overall, the results are very consistent across the three approaches, suggesting that assortative mating contributes a significant amount to the overall inequality between couples in hourly earnings. In other words, homogamy generates substantial differences between households in total earnings potential.

Monthly earnings Panel B of table 4 shows similar results for monthly earnings. An important difference is that the analysis is extended to couples where one spouse reports zero earnings. Under the addition randomization approach (lines 11-12), random rematching would slightly reduce inequality according to the Theil index (-3.8% on average), but not according to the Gini index (+2.9% on average). Under the imputation randomization approach (rows 13-14), random matching reduces income inequality between households. On average, the Gini is reduced by 3.4% and the Theil index is lower by 5.1%, with slightly larger effects at the end of the period.

Overall, these results suggest a rather limited contribution of assortativeness to inequality. However, one can question the relevance of both approaches in the case of South Korea. In a situation where female labor force participation is strongly associated with male earnings, the assumption of exogenous labor supply underlying the addition randomization approach may not be relevant. As for the imputation approach, the assumption that the observed "heterogeneous couples" are representative of what would be the earnings of "artificial couples" with heterogeneous characteristics seems unrealistic given the very strong educational assortative mating observed in Korea.[36]

Lines (15-16) present estimates based on the simulation randomization approach using a model of the joint distribution of monthly earnings. In this model, observations with zero earnings are treated as censored. The model thus corrects for participation decisions, but only

[36] For example, the share of dual-earner couples is consistently about 4 percentage points larger among couples with similar educational attainment, but their household income is slightly lower.

along the extensive margins. Estimates suggest that, on average, inequality would be about 6 to 13% lower under random mating than under the actual mating pattern.

Comparison of these estimates with the results reported in the literature is hampered by methodological differences and measurement issues.[37] Overall, when we consider monthly earnings for all couples, the effect of assortative mating on inequality appears to be lower in Korea than in France (Frémeaux and Lefranc, 2020) and Germany (Pestel, 2017) and similar to the US (Eika et al., 2019), Denmark, Sweden, and Switzerland (Kuhn and Ravazzini, 2017). The only benchmark for hourly earnings is France[38], and the impact of sorting is found to be slightly lower in Korea. However, given the increasing participation of women in the labor market for recent cohorts, as well as strong sorting by education, we can expect a growing influence of assortative mating on inequality in Korea.

Supplementary results We replicate all counterfactual analyses by cohort. The results are presented in the appendix (table B.4). For hourly earnings, the addition and simulation approaches indicate a large contribution of assortative mating to earnings inequality, especially for recent cohorts (between 10 and 15% of total inequality). The results obtained with the imputation approach are rather inconclusive and quite different from the other counterfactuals. Using the Theil index, the contribution to inequality is found to be larger for the most recent cohorts. For monthly earnings, the addition and imputation approaches suggest a more modest contribution to inequality. Using the simulation method, we estimate a contribution to inequality of 9%.

We also estimate the impact of assortative mating on permanent earnings using multi-year average earnings (table B.5 in the appendix). For the addition and simulation randomization methods, the results are unchanged. However, for the imputation method, the compensating effect of random mating is 50% greater for multi-year average earnings than for monthly measures. The results are consistent with section 4.2, where we find stronger assortativeness for average

[37] For example, the fact that the recent literature relies mostly on the Gini index, while we show that significant differences appear as soon as we use alternative indices from the class of generalized entropy, limits this comparability.

[38] Frémeaux and Lefranc (2020)

earnings. For hourly earnings, the estimates are also consistent with section 4.2. In 1998-2000 and 2018-2019, the difference between annual correlations and multi-year average earnings is limited, so the impact on inequality is roughly the same. However, the difference is larger in 2009, when the same gap is more pronounced.

7 Conclusion

In this paper, we provide original estimates of the extent of economic assortative mating in South Korea based on representative longitudinal data from the Labor and Income Panel Study over the period 1998 to 2018. Our estimates of the association of economic attributes within couples suggest a high degree of assortative mating in Korea. Our results first confirm the very high correlation in education within couples, as measured by the correlation coefficient or the degree of educational homogamy. Despite this very high correlation in human capital, the correlation in earnings across couples appears to be very low and even negative. However, this masks two opposing forces: on the one hand, among couples where both spouses work and report positive earnings, the correlation in both monthly and hourly earnings is substantial, around 0.25-0.3, and high by international standards; on the other hand, female labor force participation is strongly negatively associated with husband's earnings, so that women with high earnings potential have a strong tendency to stay out of the labor force and report zero earnings.

This paper also addresses several important methodological challenges in estimating the assortativeness of earnings within couples. First, we document that measurement error biases estimates of earnings correlations toward zero and show that using multi-year average earnings significantly reduces this attenuation bias. Estimates based on permanent earnings are found to be 10-20% larger than those based on annual observations. Second, we also show that selection problems significantly affect the measurement of assortative mating. Endogenous labor market participation is also shown to lead to an overestimation of the correlation in spouses' earnings.

We also assess the contribution of assortative mating to inequality and consider different counterfactual approaches. Standard approaches typically ignore endogenous participation behavior as well as selection into couples based on unobserved characteristics. Applying these approaches to inequality in monthly earnings across couples, we find little evidence of the contribution of assortative mating to inequality in Korea. However, when we examine inequality in the couple's joint earning potential and adequately model sample selection issues, we find a substantial dis-equalizing effect of assortative mating (between 10 and 15% of total inequality).

Although one might object that the concept of potential earnings is at odds with the typical monetary earnings measure used in inequality studies, it is important to emphasize that the partially latent contribution of assortative mating to inequality among Korean couples may become more manifest as female labor force participation increases in recent cohorts, especially at the upper end of the human capital distribution. The wage rate analysis allows us to provide an alternative benchmark because we simulate the influence of assortative mating in the absence of household specialization. In other words, this counterfactual could be seen as an upper bound on the effect of assortative mating among currently observed couples.

Competing Interests

The authors declare that they have no relevant or material financial interests related to the research described in this article.

Data availability statements

The data sets generated and analyzed in the current study are publicly available online at https://www.kli.re.kr/klips_eng/index.do. The instructions for using the datasets that support the findings of this study are available from the corresponding author upon request.

References

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, *World Inequality Report 2018*, Belknap Press, 2018.

Aslaksen, Iulie, Tom Wennemo, and Rolf Aaberge, "'Birds of a Feather Flock Together': The Impact of Choice of Spouse on Family Labor Income Inequality," *Labour*, 2005, *19* (3), 491–515.

Black, Sandra E. and Paul J. Devereux, "Recent Developments in Intergenerational Mobility," in David Card and Orley Ashenfelter, eds., *David Card and Orley Ashenfelter, eds.*, Vol. 4 of *Handbook of Labor Economics*, Elsevier, 2011, chapter 16, pp. 1487–1541.

Bollinger, Christopher R. and Amitabh Chandra, "Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data," *Journal of Labor Economics*, 2005, *23* (2), 235–258.

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers, "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 1994, *12* (3), 345–368.

Burtless, Gary, "Effects of growing wage disparities and changing family composition on the U.S. income distribution," *European Economic Review*, 1999, *43*, 853–865.

Cancian, Maria and Deborah Reed, "Assessing the Effects of Wives' Earnings on Family Income Inequality," *Review of Economics and Statistics*, 1998, *80* (1), 73–79.

Chiappori, Pierre-André, Monica Costa-Dias, Sam Crossman, and Costas Meghir, "Changes in Assortative Matching and Inequality in Income: Evidence for the UK," *Fiscal Studies*, 2020, *41* (1), 39–63.

Ciscato, Eduardo and Simon Weber, "The role of evolving marital preferences in growing income inequality," *Journal of Population Economics*, 2020, (33), 307–347.

Dupuy, Arnaud and Simon Weber, "Marriage Market Counterfactuals Using Matching Models," *Economica*, 2022, *89* (353), 29–43.

Eika, Lasse, Magne Mogstad, and Basit Zafar, "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 2019, *127* (6), 2795–2835.

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

Frémeaux, Nicolas and Arnaud Lefranc, "Assortative mating and earnings inequality in France," *Review of Income and Wealth*, 2020, *66* (4), 757–783.

Fujihara, Sho and Fumiya Uchikoshi, "Declining association with persistent gender asymmetric structure: patterns and trends in educational assortative marriage in Japan, 1950-1979.," *Research in Social Stratification and Mobility*, 2019, (60), 66–77.

Fukuda, Setsuya, Shohei Yoda, and Ryohei Mogi, "Educational Assortative Mating in Japan: Evidence from the 1980-2010 Census," *Journal of Population Studies*, 2021, pp. 1–30.

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

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

Grossbard, Shoshana, Lucia Mangiavacchi, Williman Nilsson, and Luca Piccoli, "Spouses' earnings association and inequality: A non-linear perspective," *Journal of Economic Inequality*, 2022, (20), 611–638.

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

Han, Hongyun, "Trends in educational assortative marriage in China from 1970 to 2000," *Demographic Research*, 2010, *22*, 733–770.

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

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

Hryshko, Dmytro, Chinhui Juhn, and Kristin McCue, "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) December 2014.

Hur, Yoon-Mi, "Assortative Mating for Personality Traits, Educational Level, Religious Affiliation, Height, Weight, and Body Mass Index in Parents of a Korean Twin Sample," *Twin Research*, 2003, *6* (6), 467–470.

Johnson, R. C., J. Park, J. C. DeFries, G. E. McClearn, M. P. Mi, and M. N. Rashad, "Assortative marriage for specific cognitive abilities in Korea," *Social Biology*, 1976, *23*, 311–316.

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

Kim, Daehyun., "The Analysis of the Impact of Panel Attrition on Estimation of Regular-Irregular Worker Wage Gap in the KLIPS," *European Research Studies Journal*, 2010, *0* (4), 45–56.

Kim, J, "The dynamics of income inequality in Korea.," *Proceedings of the KDI Journal of Economic Policy Conference*, 2011, pp. 175–200.

Kim, Nak Nyeon and Jongil Kim, "Top incomes in Korea, 1933-2010: Evidence from income tax statistics," *Hitotsubashi Journal of Economics*, 2015, *56* (1), 1–19.

Kuhn, Ursina and Laura Ravazzini, "The impact of assortative mating on income inequality in Switzerland," Technical Report 1, FORS Working Paper Series 2017.

Kwack, Sung Yeung and Young Sun Lee, "Income distribution of Korea in historical and international prospects: Economic Growth and the Distribution of Income in Korea Facing Globalization in a High-Tech Skill World," *KDI report, No.10341*, 2007.

Kye, Bongoh and Robert D. Mare, "Intergenerational effects of shifts in women's educational distribution in South Korea: Transmission, differential fertility, and assortative mating," *Social Science Research*, 2012, *41*, 1495–1514.

Lee, S. H., C. Y. Park, S. S. Chung, and H. M. Choi, "Panel attrition factors in Korean labor and income panel study," *Journal of the Korean Data & Information Science Society*, 2011, (22), 18.

Lee, S.H., "The analysis of sample deviation for KLIPS," *Monthly Labor Review*, 2005, (11), 6679.

Lee, Soohyung, "Effect of Online Dating on Assortative Mating: Evidence from South Korea," *Journal of Applied Econometrics*, 2016, *31* (6), 1120–1139.

Lee, Yean-Ju, "Implications of Cohabitation for the Korean Family: Cohabiter Characteristics Based on National Survey Data," *Korea Journal of Population Studies*, 2008, *31* (2), 77–100.

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

Nie, Haifeng and Chunbing Xing, "Education expansion, assortative marriage, and income inequality in China," *China Economic Review*, 2019, *55* (C), 37–51.

OECD, *Rejuvenating Korea: Policies for a Changing Society*, Paris: OECD Publishing, 2019.

Ostrovsky, Yuri, "The correlation of spouses' permanent and transitory earnings and family earnings inequality in Canada," *Labour Economics*, 2012, *19* (5), 756–768. Special Section on: Price, Wage and Employment Adjustments in 2007-2008 and Some Inferences for the Current European Crisis.

Park, Hyunjoon and Jeroen Smits, "Educational Assortative Mating in South Korea: Trends 19301998," *Research in Social Stratification and Mobility*, 2005, *23*, 103–127.

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

Sasaki, Shoichi, "Labor Market Inequality and Marital Segregation in East Asia," Discussion Papers 1822, Graduate School of Economics, Kobe University 2018.

Schwartz, Christine R., "Earnings Inequality and the Changing Association between Spouses' Earnings," *American Journal of Sociology*, 2010, *115* (5), 1524–1557.

Smits, Jeroen and Hyunjoon Park, "Five Decades of Educational Assortative Mating in 10 East Asian Societies," *Social Forces*, 2009, *88* (1), 227–258.

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

Teichman, J., "The Role of the Middle Class in Distributional Outcomes: Chile and South Korea," *Studies in Comparative International Development*, 2015, *50*, 1–21.

Table 1: Descriptive statistics

		1998		2009		2018	
		mean	s.d.	mean	s.d.	mean	s.d.
Age	Male	42.00	8.67	45.40	9.50	48.20	9.30
	Female	38.70	8.30	42.50	9.05	45.70	9.13
Education	Male	12.20	3.33	13.00	3.10	13.80	2.72
	Female	11.00	3.03	12.20	2.95	13.30	2.59
Labor market participation	Male	0.951	0.215	0.971	0.167	0.977	0.149
	Female	0.288	0.453	0.399	0.490	0.467	0.499
	Both spouses	0.239	0.426	0.370	0.483	0.444	0.497
Self-employment	Male	0.377	0.485	0.321	0.467	0.258	0.437
	Female	0.212	0.409	0.175	0.380	0.141	0.348
	Both spouses	0.147	0.355	0.114	0.318	0.081	0.274
Weekly hours worked	Male	53.50	16.90	50.90	14.00	48.30	12.20
	Female	53.80	20.70	46.90	15.90	40.70	12.80
Full-time work	Male	0.899	0.302	0.931	0.254	0.944	0.230
	Female	0.799	0.401	0.817	0.387	0.756	0.429
Monthly labor earnings	Male	347.00	266.00	315.00	199.00	317.00	164.00
	Female	134.00	86.60	176.00	117.00	201.00	109.00
Hourly labor earnings	Male	1.86	1.65	1.68	1.19	1.76	0.95
	Female	0.79	0.70	1.03	0.68	1.32	0.67

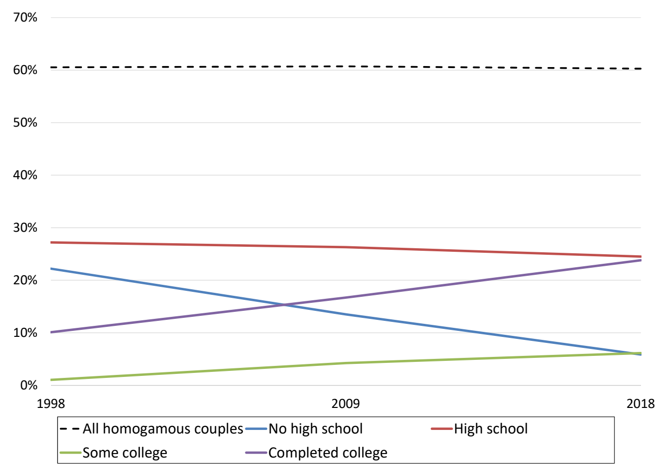
Notes – Source: KLIPS data; Education is measured as the number of years of education; Earnings are expressed in 10,000 KRW.

Table 2: Within-couple correlations - Main estimates

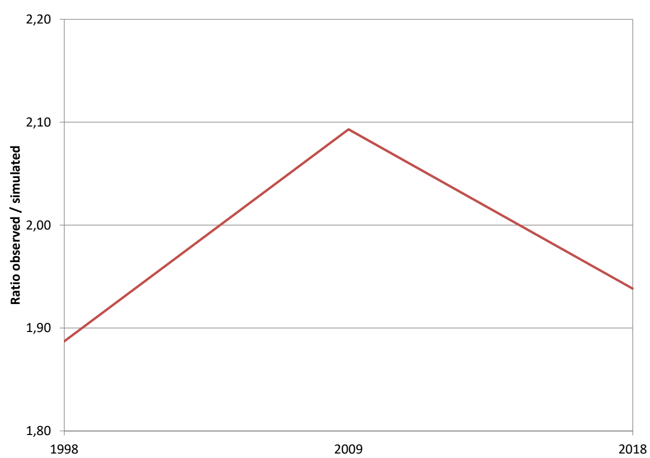
	Education (years)			Monthly earnings (incl. zeros)			Monthly earnings (excl. zeros)			Hourly earnings (excl. zeros)		
	ρ	$[\bar{\rho}$	$\underline{\rho}]$	ρ	$[\bar{\rho}$	$\underline{\rho}]$	ρ	$[\bar{\rho}$	$\underline{\rho}]$	ρ	$[\bar{\rho}$	$\underline{\rho}]$
Panel A - Cross-section observations												
1998	0.738	[0.720	0.755]	-0.0324	[-0.0699	0.00534]	0.234	[0.16	0.306]	0.275	[0.200	0.347]
2009	0.746	[0.730	0.761]	-0.00453	[-0.0393	0.0303]	0.30	[0.247	0.351]	0.333	[0.278	0.386]
2018	0.68	[0.668	0.698]	-0.0497	[-0.0782	-0.0211]	0.263	[0.223	0.303]	0.285	[0.243	0.326]
Panel B - Multi-year average earnings												
1998				-0.0197	[-0.0689	0.0297]	0.287	[0.217	0.354]	0.333	[0.265	0.398]
2009				-0.0428	[-0.0834	-0.002]	0.296	[0.244	0.346]	0.41	[0.362	0.456]
2018				-0.0529	[-0.0835	-0.0223]	0.259	[0.219	0.299]	0.329	[0.290	0.368]
Panel C - By cohort, multi-year average earnings												
1955<birth	0.69	[0.659	0.719]	-0.137	[-0.192	-0.0822]	0.43	[0.355	0.499]	0.466	[0.394	0.533]
1955-1964	0.74	[0.720	0.759]	-0.109	[-0.151	-0.0665]	0.324	[0.273	0.373]	0.375	[0.327	0.422]
1965-1974	0.686	[0.664	0.707]	0.0119	[-0.028	0.0518]	0.318	[0.272	0.362]	0.423	[0.380	0.464]
>1974	0.576	[0.549	0.601]	0.0324	[-0.006	0.0711]	0.288	[0.243	0.333]	0.350	[0.306	0.393]

Notes – Source: KLIPS data; ρ indicates the Pearson correlation coefficient for earnings and the Spearman correlation for education; $\underline{\rho}$ (resp. $\bar{\rho}$) denotes the lower (resp. upper) bound of the 95% confidence interval estimate for ρ.

Figure 1: Assortative mating by education



(a) Share of couples with similar education

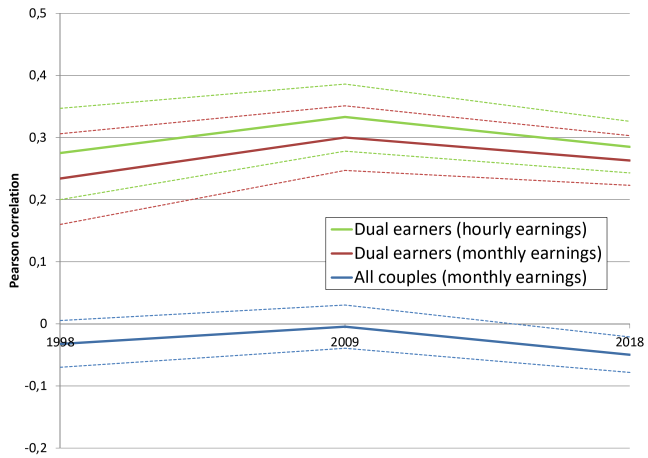


(b) Chiappori et al. (2020) assortativeness index

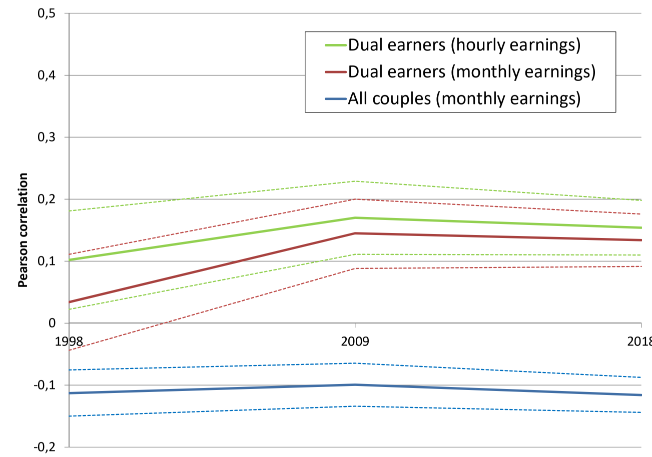
Notes – Source: KLIPS data.

Figure 2: Within-couple correlations in earnings

(a) Cross-section observations



(b) Residual earnings, net of age and education

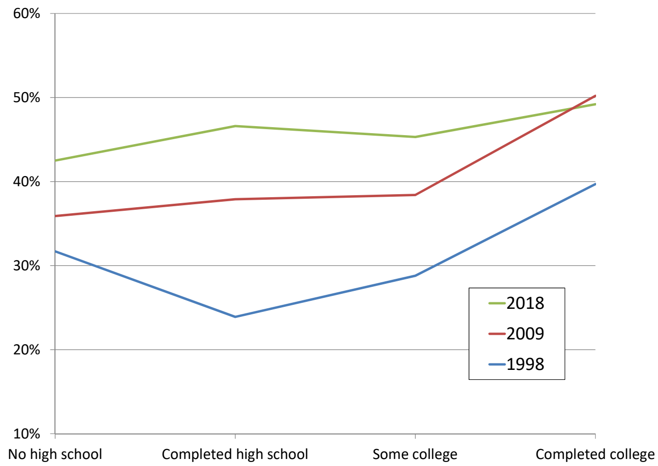


(c) Multi-year average earnings

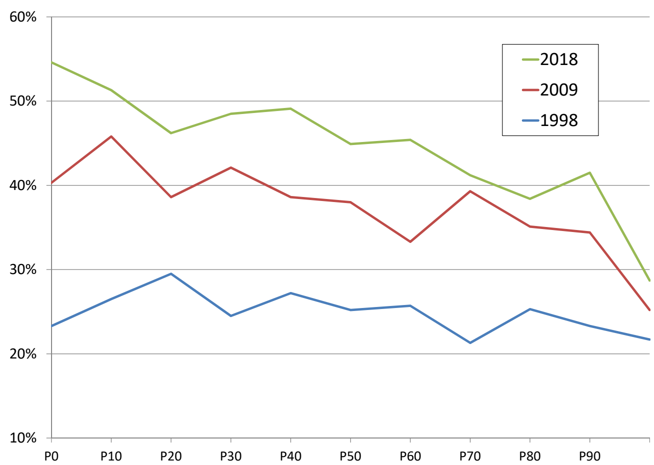
(d) Hourly earnings, corrected for sample selection

Notes – Source: KLIPS data; Estimates for all couples include individuals reporting zero earnings; Dual-earner couples include only couples in which both spouses report positive earnings. The dashed lines in 2a and 2b represent the 95% confidence intervals; In Figure 2c, "pooled" refers to the average of the yearly correlations over the period. "average" refers to the estimates based on multi-year average earnings computed each period; Figure 2d is based on Table 3.

Figure 3: Female labor market participation rate (1998–2018)



(a) By female education

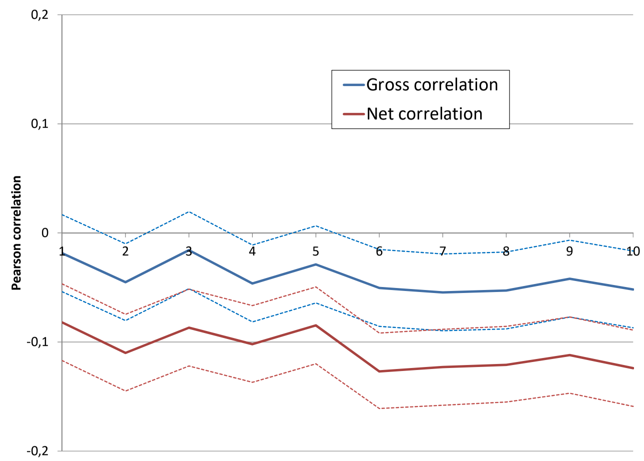


(b) By male earnings decile

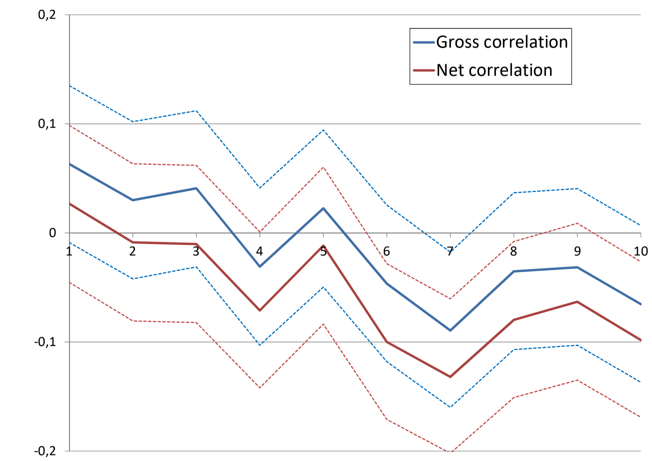
Notes – Source: KLIPS data; Participation in the labor market is a binary variable equal to 1 if the individual reports positive earnings.

Figure 4: Earnings correlation over the life-cycle of couples

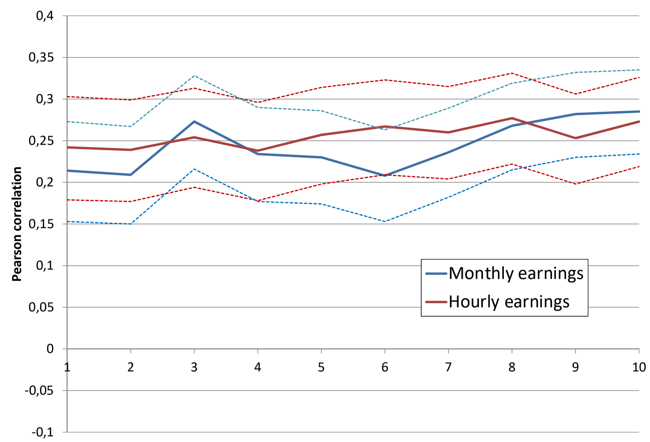
(a) All couples



(b) Couples with wives < 29 y.o.



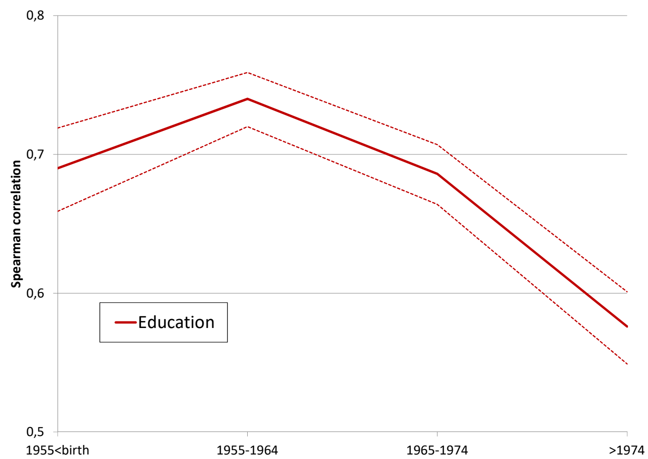
(c) Dual-earner couples



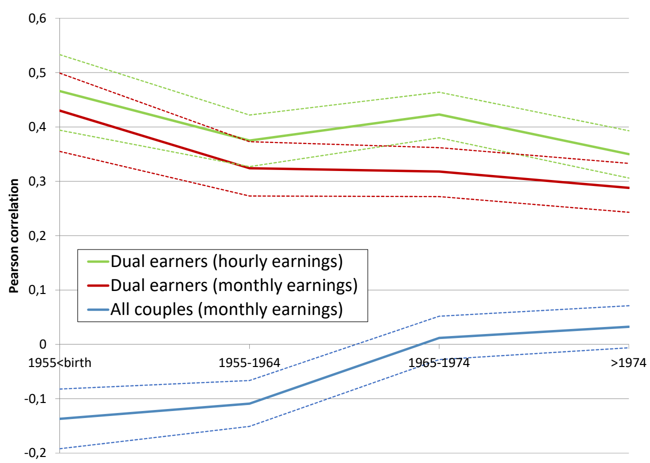
(d) Dual-earner couples with wives < 29 y.o.

Notes – Source: KLIPS data, restricted to couples observed for 10 years or more (2,556 observations); Panels (a) and (b) include couples in which one spouse reports zero earnings, while panels (c) and (d) are restricted to dual-earner couples; Correlations are estimated each year during the 10 years after they first entered the sample at least; The net correlation is the correlation of the residual earnings, net of year, age, and education; The sub-sample of couples in which the wife is under the age of 29 at the time the couple enters the panel (648 observations) proxies for recently formed couples; Dashed lines represent the 95% confidence intervals.

Figure 5: Within-couple correlation in education and earnings - by cohort



(a) Years of education



(b) Multi-year average earnings

Notes – Source: KLIPS data; Cohorts, on the x-axis, are based on wife's year of birth; Education is expressed as the number of years of education completed; Earnings refer to multi-year average earnings using all available observations (including zero earnings); Dashed lines represent the 95% confidence intervals.

Table 3: Correlation coefficients and sample selection - log hourly earnings

	1998	2009	2018
		Panel A - Ignoring sample selection	
ρ	0.309	0.360	0.304
β_{OLS}	0.265	0.330	0.273
σ_m	0.759	0.614	0.473
σ_f	0.604	0.564	0.425
Obs	604	1,040	1,899
		Panel B - Accounting for sample selection	
ρ	0.277	0.326	0.315
$\beta_{Heckman}$	0.284	0.356	0.320
σ_m	0.763	0.635	0.489
σ_f	0.784	0.692	0.496
ρ_{res}	-0.695	-0.762	-0.657
Obs	2,375	2,753	4,212

Notes – Source: KLIPS data; 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 earnings equations. Estimates are based on coefficients of the sample-selection model presented in Table A.1.

Table 4: Inequality in household earnings - observed and predicted under different random mating scenarios

			1998		2009		2018		Average	
			Gini	Theil	Gini	Theil	Gini	Theil	Gini	Theil
Panel A - Hourly earnings										
Observed	(1)	index value	0.36	0.22	0.29	0.13	0.22	0.08	0.29	0.14
	(2)	index value	0.34	0.19	0.30	0.15	0.22	0.08	0.29	0.14
Addition	(3)	index value	0.33	0.18	0.26	0.11	0.21	0.07	0.27	0.12
randomization	(4)	ratio (3)/(1)	0.912	0.835	0.919	0.852	0.930	0.850	0.919	0.843
Imputation	(5)	index value	0.33	0.20	0.27	0.12	0.20	0.06	0.27	0.13
randomization	(6)	ratio (5)/(1)	0.922	0.898	0.943	0.899	0.898	0.820	0.923	0.884
Simulation	(7)	index value	0.32	0.17	0.28	0.13	0.20	0.06	0.26	0.12
randomization	(8)	ratio (8)/(2)	0.938	0.855	0.911	0.825	0.878	0.770	0.913	0.828
Panel B - Monthly earnings										
Observed	(9)	index value	0.37	0.24	0.33	0.17	0.27	0.12	0.32	0.18
	(10)	index value	0.35	0.22	0.32	0.18	0.30	0.16	0.32	0.18
Addition	(11)	index value	0.38	0.22	0.33	0.17	0.28	0.12	0.33	0.17
randomization	(12)	ratio (12)/(9)	1.026	0.933	1.017	0.964	1.045	1.017	1.029	0.962
Imputation	(13)	index value	0.37	0.23	0.31	0.16	0.26	0.11	0.31	0.17
randomization	(14)	ratio (14)/(9)	0.983	0.985	0.958	0.928	0.955	0.912	0.966	0.949
Simulation	(15)	index value	0.34	0.20	0.30	0.15	0.28	0.13	0.31	0.16
randomization	(16)	ratio (15)/(10)	0.974	0.896	0.928	0.852	0.921	0.865	0.942	0.873

Notes – Source: KLIPS data; Lines (1) and (9) report the value of the Gini and Theil indices measured using observed data; Lines (2) and (10) use estimates of the parametric distribution of hourly earnings to estimate the level of inequality that would prevail in the absence of selection, based on observed mating patterns. The last two columns provide the unweighted average of each index over the three reference years. Lines (4), (6), (8), (12), (14) and (16) present the ratio between the counterfactual scenario and the observed level of inequality.

Appendix - for online publication only

A Accounting for sample selection

In this appendix, we provide details on the implementation of the selection model of Frémeaux and Lefranc (2020). The correlation in labor earnings is influenced by labor supply decisions along both intensive and extensive margins. 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 partners, selection is likely to be non-random. In this case, the correlation in full-time equivalents would provide a biased estimate of the correlation in hourly earnings, although the direction of the bias is a priori unknown.

A.1 Model

Let w_s denote the hourly 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:

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

The difficulty in deriving estimates of the parameters of the (latent) joint distribution, $(\rho, \sigma_m, \sigma_f)$, lies in the fact that hourly earnings are 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.

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

$$\ln w_f = \beta_0 + \beta \ln w_m + \varepsilon \tag{3}$$

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 ratio of the standard errors of male and female earnings.

We assume that w_m is always observed and that w_f is only observed for women in the labor force. In the likely case where participation decisions depend on both partners' hourly earnings, the sample of dual earners is no longer representative of the entire population. Therefore, the partners' correlation cannot be directly assessed based on observed earnings alone. Likewise, β in equation 3 cannot be estimated using linear regression. Finally, the observed 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, all these parameters can be consistently estimated using Heckman's sample selection correction applied to Equation 3. More specifically, this model yields consistent estimates for both β and σ_ε. Furthermore, these estimates can be combined to obtain an estimate of $\sigma_f = \sqrt{\sigma_\varepsilon^2 + \beta^2\sigma_m^2}$. Finally, we can obtain an estimate of the within-couple correlation in hourly log-earnings, ρ, given by

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

A.2 Results

We use this approach to estimate the correlation in residual earnings, i.e., net of age and time effects. Results reported in the paper pertain to hourly earnings. Estimates for monthly earnings are reported below. Our estimates are based on the Heckman selection model, estimated using maximum likelihood on our full sample of couples, separately for each reference year (1998, 2009, 2018). The dependent variable of the selection equation is a dummy variable equal to one if we observe strictly positive wife earnings. The selection equation includes control for characteristics of the wife (age, education, number of children and age of the eldest child[39]) and of her husband (age, education, log hourly earnings, experience, and self-employment status), as well as regional dummies.

The selection equation comes close to a participation equation. Results of the selection model are presented in Table A.1 for the three reference years, 1998, 2009 and 2018. Results in columns 1, 3 and 5 do not control for the characteristics of the husband and are estimated by a probit model. Columns 2, 4 and 6, add information about the husband and are estimated by ML using Heckman's model.

Results indicate that having a college degree increases the likelihood of participation but the coefficient decreases over time. Fertility and labor market decisions are closely related, as female labor market participation decreases with the number of children for all years. The regression also confirms that male income is negatively correlated with female participation (columns 2, 4 and 6).

Estimation results for the main equation of the Heckman model have already presented in Table 3 and discussed in section 5. Table A.2 below presents sample-selection corrected estimates of the correlation in monthly labor earnings.[40] Results indicate that sing dual-earner couples and ignoring sample selection leads to overestimate the correlation in monthly earnings by about 20% in 1998 and 10% in 2009. Results for 2018 are underestimated by about 10%. Results in Table A.2 should however be taken with greater caution than the results found for hourly earnings as the bi-variate log-normality under which the Heckman ML model is estimated may not hold empirically.

The reason behind the fall in the estimated correlation observed in 1998 and 2009, when sample selection is taken into consideration, deserves to be examined closely. The sample selection model introduces two corrections. First, it corrects estimates of the elasticity of female earnings w.r.t to their husband's earnings, β. Second, it corrects for the truncation in the observed distribution of female earnings in the estimation of σ_f. In the case of Korea, in 1998 and 2009, the two effects work in opposite direction and the second effect dominates.

Last, it is worth noting that for all specifications, the correlation between the error terms of the selection and the earnings equations (ρ_{res}) is negative and fairly stable. This indicates that women with a positive earnings residual, conditional on their partner's earnings, have a lower probability of entering the labor market. In other words, for female partners, "undermarriage"

[39] Information about the age of other children (if any) is not available in the KLIPS survey.

[40] Results of the sample selection equation for the monthly earnings specification are very similar to those discussed in the previous paragraph and are available from the authors.

Table A.1: Determinants of female labor market participation

	1998				2009				2018			
	(1)		(2)		(3)		(4)		(5)		(6)	
	coef.	s.e.	coef.	s.d.	coef.	s.d.	coef.	s.d.	coef.	s.d.	coef.	s.d.
Wife's characteristics:												
Educ. 6y	Ref.		Ref.		Ref.		Ref.		Ref.		Ref.	
Educ. 9y	0.00477	(0.158)	0.0153	(0.191)	0.0818	(0.182)	-0.103	(0.214)	-0.00238	(0.217)	-0.0948	(0.273)
Educ. 12y	-0.371**	(0.154)	-0.126	(0.209)	-0.124	(0.172)	-0.190	(0.229)	0.0196	(0.198)	-0.0812	(0.268)
Educ. 14y	-0.0122	(0.238)	0.412	(0.308)	0.0611	(0.199)	0.140	(0.264)	-0.0453	(0.211)	-0.0271	(0.283)
Educ. 16y	0.432**	(0.190)	0.912***	(0.283)	0.487**	(0.192)	0.518*	(0.269)	0.0747	(0.207)	0.180	(0.284)
Educ. 18y	1.066**	(0.518)	1.613***	(0.588)	0.990***	(0.294)	1.043***	(0.367)	0.649**	(0.252)	0.810**	(0.330)
No children	Ref.		Ref.		Ref.		Ref.		Ref.		Ref.	
1 child	-0.496*	(0.261)	-0.485*	(0.278)	-0.569***	(0.167)	-0.674***	(0.186)	-0.511***	(0.122)	-0.565***	(0.129)
2 children	-0.567**	(0.257)	-0.493*	(0.276)	-0.365**	(0.163)	-0.426**	(0.184)	-0.324***	(0.117)	-0.315**	(0.124)
3 children	-0.843***	(0.289)	-0.716**	(0.308)	-0.472**	(0.194)	-0.581***	(0.216)	-0.631***	(0.146)	-0.597***	(0.157)
> 3 children	-0.435	(0.346)	-0.349	(0.370)	-0.354	(0.333)	-0.419	(0.364)	-0.557*	(0.335)	-0.667*	(0.354)
Husband's characteristics:												
Hourly earnings			-0.322***	(0.098)			-0.347***	(0.122)			-0.364***	(0.129)
Hourly earnings sq.			0.0250**	(0.011)			0.0183	(0.019)			0.0250	(0.024)
Educ. 6y			Ref.				Ref.				Ref.	
Educ. 9y			-0.0101	(0.217)			0.246	(0.250)			-0.308	(0.301)
Educ. 12y			0.0190	(0.227)			0.390	(0.261)			-0.338	(0.301)
Educ. 14y			0.0144	(0.302)			0.00466	(0.294)			-0.536*	(0.315)
Educ. 16y			-0.0256	(0.285)			0.482	(0.294)			-0.484	(0.315)
Educ. 18y			0.395	(0.355)			0.460	(0.344)			-0.590*	(0.342)
Experience			0.0771***	(0.027)			0.0170	(0.022)			0.0147	(0.021)
Experience sq.			-0.0006	(0.000)			0.000005	(0.000)			-0.0007*	(0.000)
Self-employed			-0.544***	(0.107)			-0.477***	(0.094)			-0.618***	(0.078)
cons.	13.98	(19.859)	20.63	(24.588)	69.60***	(17.482)	82.18***	(22.093)	13.36	(12.018)	14.22	(16.162)
N	2,529		2,529		3,058		3,058		4,571		4,571	

Notes – Source: KLIPS data; Dependent variable: indicator variable of whether the wife reports positive earnings; Models 1, 3 and 5 are estimated by Probit; Models 2, 4 and 6 report estimates of the selection equation of the Heckman model; Standard errors in parentheses; Significance levels: * = 10%, ** = 5% and *** = 1%; regressions also include controls for the age of the spouses and regional indicators.

(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. These results suggests that the idiosyncratic disutility of work, captured by labor supply unobserved determinants, is not independent of the idiosyncratic potential earnings of the mate.

Table A.2: Correlation coefficients and sample selection - log monthly earnings

	1998	2009	2018
		Panel A - Ignoring sample selection	
ρ	0.218	0.287	0.208
β_{OLS}	0.194	0.285	0.228
σ_m	0.702	0.610	0.502
σ_f	0.624	0.605	0.550
Obs	646	1,174	2,084
		Panel B - Accounting for sample selection	
ρ	0.174	0.269	0.229
$\beta_{Heckman}$	0.222	0.337	0.359
σ_m	0.700	0.622	0.503
σ_f	0.891	0.779	0.781
ρ_{res}	-0.875	-0.812	-0.909
Obs	2,529	3,058	4,571

Notes – Source: KLIPS data; 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 earnings equations. Estimates are based on coefficients of the sample-selection model presented in Table A.1.

B Supplementary results

Table B.1: Additional correlations - Estimates

	Monthly earnings (incl. zeros)			Monthly earnings (excl. zeros)			Hourly earnings (excl. zeros)		
	ρ	$\underline{\rho}$	$\bar{\rho}$	ρ	$\underline{\rho}$	$\bar{\rho}$	ρ	$\underline{\rho}$	$\bar{\rho}$
Rank correlations									
1998	-0.134	-0.171	-0.0972	0.236	0.162	0.308	0.308	0.234	0.378
2009	-0.075	-0.11	-0.0403	0.29	0.237	0.342	0.35	0.295	0.402
2018	-0.087	-0.116	-0.0589	0.25	0.209	0.289	0.302	0.261	0.343
Residual after controlling for age and education									
1998	-0.113	-0.15	-0.0754	0.0339	-0.0438	0.111	0.102	0.022	0.181
2009	-0.0993	-0.134	-0.0646	0.145	0.088	0.200	0.17	0.111	0.229
2018	-0.116	-0.144	-0.0875	0.134	0.0915	0.176	0.154	0.11	0.198
Multi-year average earnings correlations - Rank correlations									
1998	-0.155	-0.203	-0.107	0.217	0.145	0.287	0.268	0.197	0.336
2009	-0.136	-0.176	-0.0955	0.251	0.198	0.302	0.373	0.323	0.421
2018	-0.107	-0.138	-0.0771	0.235	0.193	0.275	0.321	0.281	0.36
Multi-year average earnings correlations - Varying windows									
1 year	0.0019	-0.0257	0.0294	0.285	0.245	0.325	0.294	0.251	0.335
2 years	-0.0104	-0.0399	0.0191	0.299	0.260	0.337	0.326	0.287	0.364
3 years	-0.0181	-0.0496	0.0133	0.321	0.282	0.359	0.364	0.326	0.401
5 years	-0.0470	-0.0817	-0.0121	0.273	0.231	0.314	0.359	0.319	0.398
Cohort - Multi-year average earnings (observations between 30 and 50 y.o. only)									
1955<birth	-0.0449	-0.16	0.0713	0.402	0.241	0.542	0.501	0.353	0.624
1955-1964	-0.0942	-0.146	-0.0423	0.309	0.249	0.367	0.336	0.276	0.393
1965-1974	-0.0216	-0.0659	0.0227	0.292	0.24	0.341	0.386	0.338	0.433
>1974	0.0196	-0.0214	0.0605	0.301	0.251	0.349	0.341	0.292	0.387
Cohort - Multi-year average earnings (closest observation to the age of 40)									
1955<birth	-0.118	-0.233	0.0002	0.395	0.236	0.534	0.251	0.0778	0.41
1955-1964	0.0962	0.0389	0.153	0.189	0.125	0.251	0.161	0.0971	0.225
1965-1974	0.138	0.0917	0.185	0.195	0.141	0.247	0.25	0.198	0.301
>1974	0.0915	0.0491	0.134	0.254	0.204	0.303	0.257	0.207	0.306

Notes – Source: KLIPS data; ρ indicates the Pearson correlation coefficient; $\underline{\rho}$ (resp. $\bar{\rho}$) denotes the lower (resp. upper) bound of the 95% confidence interval estimate for ρ.

Table B.2: Additional correlations - Estimates

	Annual earnings (incl. zeros)			Annual earnings (excl. zeros)			Hourly earnings		
	ρ	$\underline{\rho}$	$\bar{\rho}$	ρ	$\underline{\rho}$	$\bar{\rho}$	ρ	$\underline{\rho}$	$\bar{\rho}$
Alternative definition: splitted income for couples of self-employed									
1998	-0.0142	-0.0518	0.0235	0.34	0.283	0.395			
2009	0.009	-0.0256	0.044	0.381	0.336	0.424			
2018	-0.0384	-0.0669	-0.009	0.327	0.291	0.363			
Correlation without winzorizing									
1998	-0.0186	-0.0562	0.0192	0.226	0.152	0.298	0.188	0.11	0.264
2009	-0.0159	-0.0508	0.019	0.234	0.179	0.287	0.281	0.223	0.336
2018	-0.0196	-0.0482	0.009	0.22	0.178	0.26	0.179	0.135	0.222
No restriction about low hourly earnings									
1998							0.278	0.203	0.349
2009							0.338	0.284	0.391
2018							0.298	0.257	0.338
Maximum limit of number of hours worked set at 68 hrs / week									
1998							0.272	0.196	0.344
2009							0.334	0.279	0.387
2018							0.287	0.245	0.328
FTE earnings (hourly earnings $\times$ typical full-time hours of work (i.e. 68hrs))									
1998							0.275	0.200	0.347
2009							0.336	0.281	0.389
2018							0.289	0.247	0.330

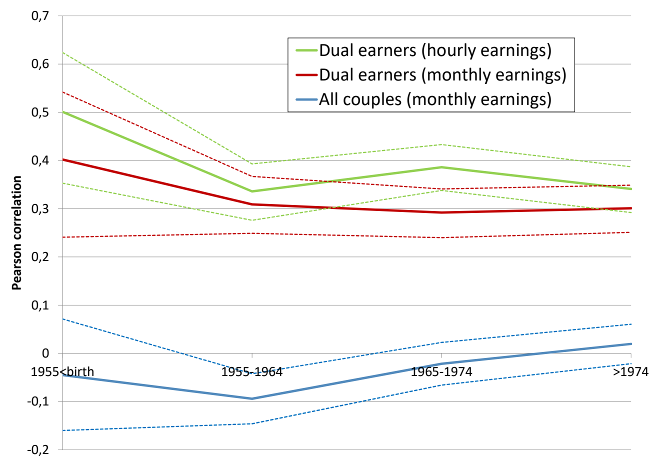
Notes – Source: KLIPS data; ρ indicates the Pearson correlation coefficient; $\underline{\rho}$ (resp. $\bar{\rho}$) denotes the lower (resp. upper) bound of the 95% confidence interval estimate for ρ.

Table B.3: Longitudinal analysis - Estimates

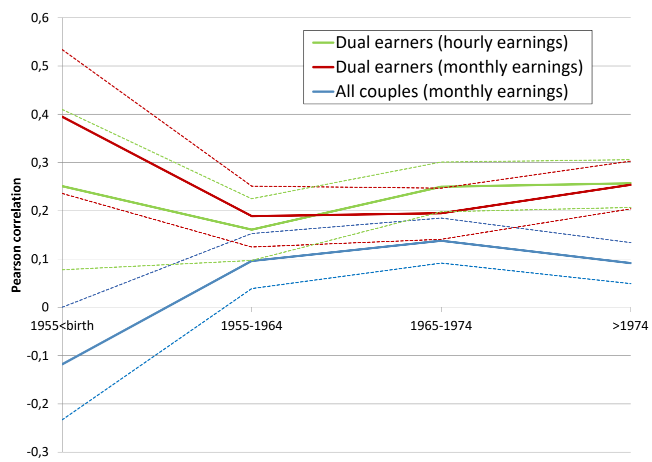
	All couples			Female partner < 29 y.o.		
	ρ	$\underline{\rho}$	$\bar{\rho}$	ρ	$\underline{\rho}$	$\bar{\rho}$
Panel A - Gross correlation - Monthly earnings (incl. zeros)						
year 1	-0.0186	-0.0539	0.0167	0.0632	-0.009	0.135
year 2	-0.0452	-0.0804	-0.0099	0.03	-0.0421	0.102
year 3	-0.0158	-0.0511	0.0195	0.0409	-0.0312	0.112
year 4	-0.0464	-0.0816	-0.0111	-0.0309	-0.103	0.0412
year 5	-0.0289	-0.0642	0.0064	0.0225	-0.0496	0.0943
year 6	-0.0505	-0.0857	-0.0152	-0.0465	-0.118	0.0256
year 7	-0.0546	-0.0897	-0.0193	-0.0895	-0.16	-0.0176
year 8	-0.0528	-0.088	-0.0175	-0.0353	-0.107	0.0368
year 9	-0.042	-0.0772	-0.0067	-0.0315	-0.103	0.0406
year 10	-0.0519	-0.087	-0.0166	-0.0653	-0.137	0.0067
Panel B - Net of year, age and education - Monthly earnings (incl. zeros)						
year 1	-0.082	-0.117	-0.0467	0.0268	-0.0453	0.0986
year 2	-0.11	-0.145	-0.0746	-0.0087	-0.0807	0.0634
year 3	-0.0869	-0.122	-0.0516	-0.0102	-0.0822	0.0619
year 4	-0.102	-0.137	-0.0666	-0.0711	-0.142	0.0009
year 5	-0.0848	-0.12	-0.0495	-0.0117	-0.0836	0.0603
year 6	-0.127	-0.161	-0.0918	-0.0999	-0.171	-0.0282
year 7	-0.123	-0.158	-0.0884	-0.132	-0.202	-0.0604
year 8	-0.121	-0.155	-0.0857	-0.0798	-0.151	-0.0079
year 9	-0.112	-0.147	-0.0771	-0.0632	-0.135	0.0088
year 10	-0.124	-0.159	-0.0891	-0.0983	-0.169	-0.0265

Notes – Source: KLIPS data; ρ indicates the Pearson correlation coefficient; $\underline{\rho}$ (resp. $\bar{\rho}$) denotes the lower (resp. upper) bound of the 95% confidence interval estimate for ρ.

Figure B.1: Assortative mating across cohorts — Alternative definitions



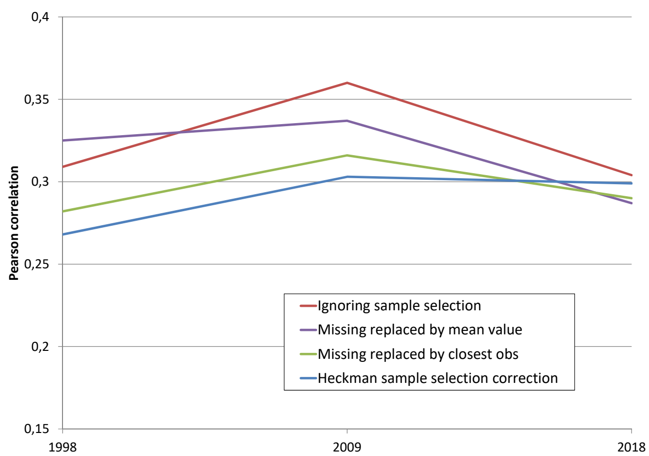
(a) Multi-year average earnings (observations between ages 30 and 50)



(b) Earnings (closest observation to the age of 40)

Notes – Source: KLIPS data; Cohorts are based on wife's year of birth; Education is expressed as the number of years of education completed; Earnings in panel (a) refer to multi-year average earnings using all using all observations between 30 and 50 years old; Dashed lines represent the 95% confidence intervals.

Figure B.2: Sample selection corrections — Alternative methods (1998–2018)



Notes – Source: KLIPS data on the sub-sample of couples with at least one observation of hourly earnings, for each spouse. The red line represents the correlation of log hourly earnings without any correction. The other lines present different types of corrections. In purple, we replace the missing value(s) with the individual's mean hourly earnings. In green, we replace the missing value(s) with the closest available observation. In blue, we apply the Heckman sample selection correction, as discussed in the text.

Table B.4: Inequality in household earnings (cohort analysis) - observed and predicted under different random mating scenarios

			Birth<1955		1955-1964		1965-1974		> 1975		Average	
			Gini	Theil	Gini	Theil	Gini	Theil	Gini	Theil	Gini	Theil
			Panel A - Hourly earnings									
Observed	(1)	index value	0.30	0.15	0.24	0.10	0.23	0.09	0.20	0.06	0.24	0.10
	(2)	index value	0.30	0.14	0.24	0.09	0.24	0.09	0.20	0.06	0.24	0.10
Addition	(3)	index value	0.29	0.14	0.24	0.09	0.21	0.07	0.18	0.05	0.23	0.09
randomization	(4)	ratio (3)/(1)	0.972	0.892	0.971	0.909	0.917	0.823	0.900	0.813	0.940	0.859
Imputation	(5)	index value	0.32	0.17	0.22	0.08	0.21	0.07	0.19	0.06	0.24	0.10
randomization	(6)	ratio (5)/(1)	1.097	1.125	0.894	0.790	0.902	0.809	0.989	1.010	0.971	0.934
Simulation	(7)	index value	0.26	0.11	0.22	0.08	0.21	0.07	0.17	0.05	0.21	0.07
randomization	(8)	ratio (8)/(2)	0.871	0.746	0.928	0.872	0.860	0.734	0.860	0.735	0.880	0.772
			Panel B - Monthly earnings									
Observed	(9)	index value	0.35	0.20	0.28	0.13	0.26	0.11	0.23	0.09	0.28	0.13
	(10)	index value	0.29	0.14	0.23	0.09	0.24	0.10	0.22	0.08	0.25	0.10
Addition	(11)	index value	0.38	0.21	0.30	0.14	0.26	0.11	0.24	0.09	0.29	0.14
randomization	(12)	ratio (12)/(9)	1.082	1.041	1.086	1.131	1.028	1.034	1.020	1.007	1.054	1.053
Imputation	(13)	index value	0.33	0.18	0.29	0.14	0.26	0.11	0.24	0.09	0.28	0.13
randomization	(14)	ratio (14)/(9)	0.939	0.894	1.054	1.125	1.004	1.014	1.031	1.065	1.007	1.024
Simulation	(15)	index value	0.26	0.11	0.21	0.07	0.22	0.08	0.20	0.07	0.22	0.08
randomization	(16)	ratio (15)/(10)	0.900	0.801	0.918	0.857	0.893	0.795	0.919	0.829	0.908	0.821

Notes – Source: KLIPS data; We present 2 "observed" levels of inequality (lines (1), (2), (9) and (10)). Line (1) and (9) report the observed value of the Gini and Theil indices. Lines (2) and (10) use estimates of the parametric distribution of hourly earnings to simulate the level of inequality that would prevail in the absence of selection, based on observed mating patterns. The last two columns provide the unweighted average of each index in for the four cohorts. Lines (4), (6), (8), (12), (14) and (16) present the ratio between the counterfactual scenario and the observed level of inequality.

Table B.5: Inequality in household earnings (multi-year averages) - observed and predicted under different random mating scenarios

			1998-2000		2009-2011		2018-2019		Average	
			Gini	Theil	Gini	Theil	Gini	Theil	Gini	Theil
Panel A - Hourly earnings										
Observed	(1)	index value	0.30	0.15	0.25	0.10	0.21	0.08	0.26	0.11
	(2)	index value	0.29	0.13	0.25	0.10	0.20	0.06	0.24	0.10
Addition	(3)	index value	0.27	0.12	0.23	0.09	0.19	0.06	0.23	0.09
randomization	(4)	ratio (3)/(1)	0.911	0.815	0.918	0.832	0.900	0.795	0.910	0.816
Imputation	(5)	index value	0.27	0.13	0.22	0.08	0.20	0.07	0.23	0.09
randomization	(6)	ratio (5)/(1)	0.903	0.841	0.864	0.745	0.920	0.882	0.895	0.820
Simulation	(7)	index value	0.27	0.12	0.22	0.08	0.18	0.05	0.22	0.08
randomization	(8)	ratio (8)/(2)	0.934	0.872	0.903	0.818	0.893	0.800	0.912	0.839
Panel B - Monthly earnings										
Observed	(9)	index value	0.30	0.15	0.27	0.11	0.26	0.11	0.28	0.13
	(10)	index value	0.29	0.14	0.26	0.11	0.26	0.11	0.27	0.12
Addition	(11)	index value	0.31	0.16	0.28	0.12	0.27	0.11	0.29	0.13
randomization	(12)	ratio (12)/(9)	1.024	1.028	1.037	1.056	1.034	1.026	1.031	1.036
Imputation	(13)	index value	0.28	0.13	0.24	0.10	0.26	0.11	0.26	0.11
randomization	(14)	ratio (14)/(9)	0.926	0.857	0.917	0.855	0.991	0.985	0.944	0.894
Simulation	(15)	index value	0.28	0.13	0.24	0.09	0.24	0.10	0.25	0.11
randomization	(16)	ratio (15)/(10)	0.945	0.929	0.937	0.893	0.942	0.867	0.941	0.889

Notes – Source: KLIPS data; We present 2 “observed” levels of inequality (lines (1), (2), (9) and (10)). Line (1) and (9) report the observed value of the Gini and Theil indices. Lines (2) and (10) use estimates of the parametric distribution of hourly earnings to simulate the level of inequality that would prevail in the absence of selection, based on observed mating patterns. The last two columns provide the unweighted average of each index in the three periods. Lines (4), (6), (8), (12), (14) and (16) present the ratio between the counterfactual scenario and the observed level of inequality.