I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16513

Marital Sorting, Household Inequality and Selection

Ivan Fernandez-Val
Aico van Vuuren
Francis Vella

OCTOBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16513

Marital Sorting, Household Inequality and Selection

Ivan Fernandez-Val
Boston University

Aico van Vuuren
University of Groningen, Gothenburg University and IZA

Francis Vella
Georgetown University and IZA

OCTOBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Marital Sorting, Household Inequality and Selection*

Using CPS data for 1976 to 2022 we explore how wage inequality has evolved for married couples with both spouses working full time full year, and its impact on household income inequality. We also investigate how marriage sorting patterns have changed over this period. To determine the factors driving income inequality we estimate a model explaining the joint distribution of wages which accounts for the spouses' employment decisions. We find that income inequality has increased for these households and increased assortative matching of wages has exacerbated the inequality resulting from individual wage growth. We find that positive sorting partially reflects the correlation across unobservables influencing both members' of the marriage wages. We decompose the changes in sorting patterns over the 47 years comprising our sample into structural, composition and selection effects and find that the increase in positive sorting primarily reflects the increased skill premia for both observed and unobserved characteristics.

JEL Classification: C30, J12, J31

Keywords: marital sorting, inequality, selection

Corresponding author:
Francis Vella
Economics Department
Georgetown University
37th and O Streets, NW
Washington, DC 20057
USA
E-mail: fgv@georgetown.edu

* We are grateful to David Card and Stella Hong for useful comments and to Richard Blundell, Bram De Rock and Costas Meghir for several important references.

1 Introduction

While most empirical studies evaluate inequality via an examination of individuals' wages or earnings (see, for example, Katz and Murphy 1992, Murphy and Welch 1992, Juhn, Murphy, and Pierce 1993, Welch 2000, Autor, Katz, and Kearney 2008, Blau and Kahn 2009, Acemoglu and Autor 2011, Autor, Manning, and Smith 2016, Murphy and Topel 2016, and Fernández-Val et al. 2023a, b), for many individuals their household's income is more determinative of their economic welfare. As households are composed in a variety of ways, it is useful for policy purposes to analyze income inequality for different household types. One type of particular interest is married couples comprising individuals both working full time full year (FTFY). Although the proportion of husbands and wives in the total population aged between 24 and 65 years decreased from 77.7% to 58.2% from 1976 to 2022, this group represents an increasingly larger share of married couples. In 1976 26.2% of married couples in which both the husband and the wife were aged between 24 and 65 years had both members working FTFY and by 2022 this number increased to 51.3%. As the percentage of married males in FTFY employment increased from 82.8 to 84.1 while that for married females went from 31.1 to 56.3, the growth of this group reflects the increasing employment rates of married females.

Examining the patterns and determinants of income inequality across FTFY married couples is interesting from a number of perspectives. First, studies on wage and income inequality generally do not distinguish between married and unmarried individuals. It is useful to examine if inequality among the married shows similar patterns to those of their unmarried counterparts and married individuals with non-working spouses. Second, as household income combines the earnings of both spouses, positive sorting on wages will exacerbate inequality while negative sorting will mitigate it. Finally, the large increase in the participation rates of females, with its implications for selection bias, has been shown to affect females' wages and income inequality (see, for example, Mulligan and Rubinstein 2008, Blau, Kahn, Boboshko, and Comey 2021 and Fernández-Val et al. 2023b). While selection bias has been largely ignored for males, the existing evidence indicates it is not economically important. However, as increasing married female participation rates increase the sample of males in FTFY married couples, female selection into FTFY employment may have implications for the wage and earnings distributions of married males with working spouses.

There is a vast literature, employing a variety of methodological approaches, on the relationship between marriage, sorting and inequality and it provides mixed evidence on sorting behavior. Kremer (1999) documents that marital sorting on education in the United States, measured by the correlation between spouse's education levels, had declined over the period 1940-1990 and via a calibration exercise finds that marital sorting had a larger effect on intergenerational mobility than inequality. Fernández and Rogerson (2001) develop and calibrate a dynamic model of intergenerational education acquisition, fertility and marital sorting based on the PSID and conclude that an increase in sorting is likely to increase the degree of income inequality. Greenwood, Guner, Kocharkov, and Santos (2015) examine US census data and document an increase in positive assortative matching on education and via the comparison of counterfactual income distributions conclude that the level of income inequality measured by the Gini coefficient has increased. Chiappori, Salanié and Weiss (2017) provide a model in which increased returns to education results in higher paid couples spending more time with their children and increasing assortative matching on education. Examining US marriages for individuals who are married and born between 1943 and 1972, they find that this occurred for white, but not black, individuals. Eika, Mogstad, and Zafar (2019) examine data for Denmark, Germany, Norway, the United Kingdom, and the United States and show that there is a considerable amount of educational assortative matching although it has changed little since the 1980s. Moreover, while there was an increase in sorting at the bottom of the educational distribution, there is a remarkable decrease of assortative matching at the top of the educational distribution. They conclude that the increases in the Gini coefficient cannot be explained by changes in assortative matching. However, they also note that assortative matching on education contributes to the cross-sectional inequality in household income in each country. This supports the earlier work of Breen and Salazar (2011) for the United States that finds a very small impact of educational sorting on changes in income inequality between the late 1970s and early 2000s. Gihleb and Lang (2018) employ a range of statistical measures of assortativeness to the CPS data for 1970 to 2010 and the 2010 American Community Survey to test for increased educational homogany among married couples in the US and matching conclude that there is no evidence of increased assortative matching. However, they do not explore sorting on wages due to the non trivial non participation of wives in market employment. Chiappori, Costa-Dias and Meghir (2020,2023) provide alternative measures of assortativeness to illustrate the difficulty in quantifying how it varies across economies with different marginal distributions of the variable on which it is being evaluated. They conclude that the degree to which educational homogany has changed among married couples in the US depends on the measure employed. The conclusion also varies on where in the educational distribution it is measured. Chiappori, Costa-Dias, Crossman and Meghir (2020) employed related measures in evaluating educational assortativeness in the U.K and conclude that there are no clear patterns. However, they also conclude that the changes appear to have only slightly increased income inequality. Our empirical work adds to this literature as we provide a more rigorous and detailed analysis of sorting on wages while accounting for selection although we do so by examining a more homogenous, albeit large, group of workers. However, as our focus is primarily on the role of sorting on inequality we focus on this issue rather than adopting the measures proposed by Gihleb and Lang (2018) and Chiappori and co-authors.

While earnings appear a more appropriate measure than wages for evaluating the welfare implications of inequality, examining different measures may lead to substantially different conclusions. While substantial evidence suggests female wage inequality has risen, Fernández-Val, Van Vuuren, Vella and Perrachi, hereafter FVVP (2023a), show that annual earnings inequality has decreased for females in the United States due to the shifts in their annual hours of work distribution. This finding is similar in spirit to Cancian and Reed (1998a,1998b) who find household inequality decreased due to the large shifts in the female income distribution resulting from changes in their hours of work. Analyzing changes in annual income is more challenging for married couples as it requires jointly modeling annual hours and wages of both spouses. However, one could begin by restricting attention to those with a relatively homogeneous level of hours while accounting for the accompanying selection. As FTFY individuals generally work a similar numbers of hours, we can then examine income inequality via comparisons of the sum of couples' wages.

We document the changes in the earnings distributions of dual FTFY households and examine if household inequality has been exacerbated by sorting behavior. We do so via an examination of how the wage distributions of husbands and wives and sorting patterns have changed over time. We report how the probability of a male in the j^{th} decile of the married male wage distribution being married to a female in the k^{th} decile of the married female wage distribution has evolved over our sample period. We investigate the source of these changes by estimating the conditional joint distribution of husbands' and wives' wages via the bivariate distribution regression methodology of Fernández-Val, Meier, Van Vuuren and Vella (2023). While this provides insight into how the observed sorting patterns can be explained by observable characteristics and unobservable factors, it does not account for the selection arising from the couple's respective employment decisions. We incorporate these employment decisions by extending the selection model of Chernozhukov, Fernández-Val, and Luo (2019) (hereafter CFL) based on the Heckman (1974, 1979) selection model to bivariate selection rules. We show point identification of this model under the same exclusion restrictions as in CFL. This analysis relies on a useful result for the multivariate standard normal distribution that might be of independent interest. Lemma 1 in the appendix characterizes the derivative of this function with respect to each element of the correlation matrix and shows it is positive. While this result is well known in the bivariate case (e.g., Sibuya, 1959) we have not found it for the general multivariate case. Using the model's estimates, we evaluate the role of the different forces generating the observed marital patterns and their implications for the couples' earnings distribution.

Our empirical investigation uncovers a number of notable findings. First, we confirm that wage inequality among couples for which both spouses are working FTFY has increased for the period 1976-2022. Second, we find increasing levels of assortative sorting on wages and that these have contributed to increasing household inequality. Third, there is mixed evidence regarding the role of selection from work decisions on observed sorting patterns. Fourth, the primary factors behind the observed positive sorting patterns are the observed characteristics of the individuals and the correlation between the unobservables driving the wages of each spouse. Finally, we find that positive sorting on wages has increased over the 47 years of our sample and this reflects the increasing market value of observed and unobserved individual characteristics. Consistent with Gihleb and Lang (2018), Eika, Mogstad, and Zafar (2019) and Chiappori, Costa-Dias and Meghir (2020,2023) our results establish that the nature of sorting on observed characteristics, such as education, has not substantially changed. However, we find that the prices of these characteristics have changed to push individuals further up, or down, their wage distributions.

We highlight three important features of our modeling approach treated as exogenous. The first is the individual's decision to marry and the second is their choice of spouse (see, for example, Chiappori, Costa-Dias and Meghir 2019). While each of these is interesting, it is beyond the scope of this paper to model these decisions. We also do not address how household income is allocated across the spouses nor the implications of this allocation for the work or marriage decisions (see, for example, Lise and Seitz 2011, Lise and Yamada 2018 and De Rock, Kovaleva and Potoms 2023). While the failure to address each of these issues represents a shortcoming of our approach, the focus on explaining the sorting pattern of spouses, and its implication for inequality, within the context of endogenous employment decisions remains important.

The following section briefly describes the Current Population Survey data examined and how the sample is selected. It also presents the time series trends in earnings and the wage distributions of FTFY households. Section 3 describes the observed patterns of marital sorting and Section 4 provides our econometric model. The measures of sorting are discussed in Section 5, and estimation is discussed in Section 6. Our empirical results are presented in Section 7. Section 8 concludes.

2 Data

2.1 Data

We employ the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), or March CPS, for the 47 survey years from 1976 to 2022 which report annual earnings and hours worked for the previous calendar year.[1] The 1976 survey is the first for which information on weeks worked and usual hours of work per week last year are available.[2] To avoid issues related to retirement and ongoing educational investment we restrict attention to those aged 24–65 years in the survey year. This produces an overall sample of 2,054,502 males and 2,228,726 females. The annual sample sizes range from a minimum of 30,767 males and 33,924 females in 1976 to a maximum of 55,039 males and 59,622 females in 2001.

Annual hours worked are defined as the product of weeks worked and usual weekly hours of work last year. Those reporting zero hours generally respond not being in the labor force (i.e., they report themselves as doing housework, unable to work, at school, or retired) in the week of the March survey. We define hourly wages as the ratio of reported annual labor earnings in the year before the survey, converted to constant 2021 prices using the consumer price index for all urban consumers, and annual hours worked. Hourly wages are unavailable for those not in the labor force. As annual earnings and hours tend to be poorly measured for the Armed Forces, self-employed, and the unpaid family workers we exclude these groups and focus on civilian dependent employees with positive hourly wages and people out of the labor force last year. This restricted sample comprises 1,783,599 males and 2,097,035 females (respectively 86.8% and 94.1% of the original sample of those aged 24–65). The subsample of civilian dependent employees with positive hourly wages contains 1,540,948 males and 1,465,165 females. Married individuals aged between 24 and 65 years make up 65% of the total sample. While this has decreased from 77% in 1976 to 57% in 2022, they still represent a substantial fraction of the total sample. The percentage of married couples with both spouses working full time increased

[1] The data are taken from the IPUMS-CPS website maintained by the Minnesota Population Center at the University of Minnesota (Flood, King, Ruggles, and Warren 2015).

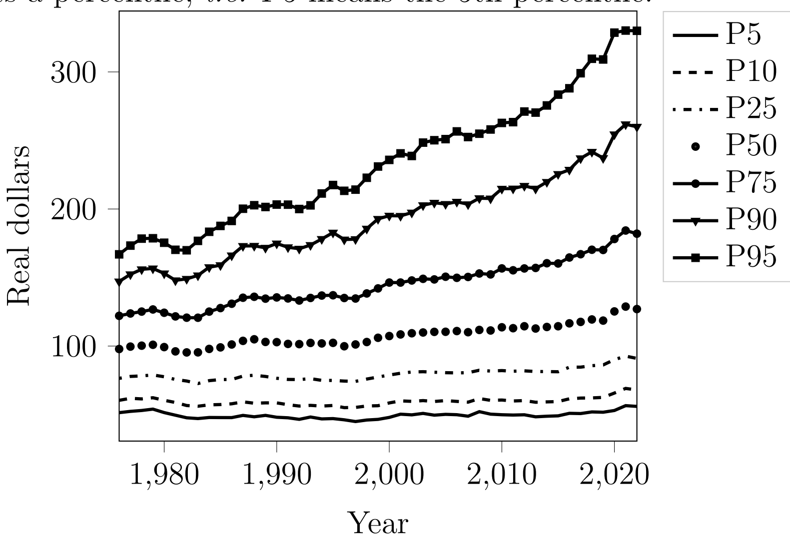
[2] We refer to the year of the survey and not the calendar year to which it refers.

drastically from 26.2 in 1976 to 51.3 in 2022.

2.2 Descriptive statistics

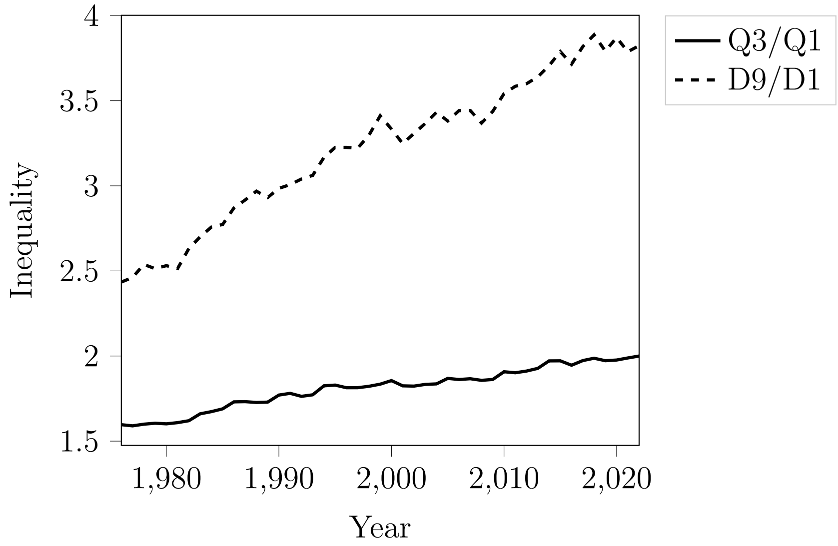
Figure 1 presents the time series of various quantiles of annual household labor earnings (in 2021 dollars) for dual FTFY households. Median (Q2) household annual income increased by 29.8% from 97.9 to 127 thousand dollars from 1976 to 2022. However, growth has been more modest at lower quantiles. For example, at the first quartile (Q1), household income increased by 19.0% from 76.5 to 91.0 thousand. Moreover, household income at this quantile was virtually constant from 1976 to 2000. There were increases of 12.5% and 8.9% at the first decile (D1) and the 5th percentile respectively. Increases at higher quantiles have been notably larger. Income at Q3 grew by 49.1% from 122.1 to 182 thousand. Increases at D9 and the 95th percentile were 76.7% and 97.8% respectively. These changes have drastically increased inequality. The Q3/Q1 and D9/D1 ratios are shown in Figure 2. The former increased from 1.60 to 2.00 and the latter from 2.43 to 3.82.

Figure 1: Household income of households working FTFY at various quantiles. *Note:* Px represents a percentile, *i.e.* P5 means the 5th percentile.



FVVP (2023a) show that there is little variation in hours for either males or females working FTFY and the income variation reflects changes in wages. Figure 3 presents wage growth for married males and females at D1, Q1, Q2, Q3 and D9. For 1976-1996 the real wage for married males at D1 decreased by 17%. It recovers

Figure 2: Inequality measures of households working FTFY at various quantiles.



somewhat but in 2022 it remains 11% below its 1976 value. There are large decreases during the financial crisis and a large decrease in the 2010's. The real wage at Q1 for married males shows a similar pattern to that at D1. There is a large decrease in the 1976-1996 period but a recovery during the second half of the sample period despite the two large dips. The 2022 value is 9% lower than the 1976 value. The married male median shows a large decrease at 1996 but virtually no change over the sample period. The overall picture at Q3 is more positive although there are several sharp decreases. However these are offset by large gains during periods of increasing real wages. The 2022 value is 21% higher than the 1976 value. At D9 there is a similar pattern to that at Q3 but with smaller dips and faster increases. This results in a dramatic gain of 42% over the whole sample period.

The time series pattern of the married female wage distribution at the lowest decile in the earlier part of the sample period is similar to that of married males although the decreases are less dramatic. Over the period 1979-1996 it decreases by 8%. For the remaining 26 years it increases by 27% and for the whole period it increases by 19%. The pattern of real wages for married females at Q1 is dissimilar to that at D1. The trend of wage growth appears to be affected by cyclical factors but generally is steadily increasing for the time period examined. This results in an increase of 29%. Growth at Q2 is even more drastic with an increase of 41%. Moreover, with the exception of a dip in the early 1980's the median real wage for this group follows a strong upward trend. At Q3 there are periods of substantial gains and these offset the small decreases which are incurred. An overall gain of 60% is achieved. A similar story is observed at D9 but the increases amount to a very substantial 84%.

Given the existence of a marriage premium in mean wages it is interesting to contrast this evidence on married individuals' wage distributions to that of all individuals. Fernández-Val et al. (2018) provide the corresponding rates of wage changes for all males and all females for the period 1976-2016. Median male wages decline by 13.6% for all working males and the decrease at Q1 is 18.2%. At Q3 there is very modest growth. This clearly suggests that there is a substantial difference between the experience of married and unmarried males. For all females, wage growth at Q1 and Q2 are 17% and 25% respectively and there are strong increases at Q3. Married females have experienced favorable wage growth compared to their unmarried counterparts.

The primary focus of our empirical work below is to model the joint distribution of married couple's real wage rates. To motivate what follows we report how the distribution of the sum of the husband's and wife's wages has evolved over our sample period. This is reported in Figure 4. We define this as the household hourly wage as it captures the combined market value of an hour of work of each spouse. The household hourly wage at D1 decreases by 12% over the period 1976-1996 but increases over the remainder of the sample period. By 2022 it is 9% higher than in 1976. The household wage at Q1 shows some of the dramatic dips featured in the married males profile but over the sample period there is an increase of 15%. At Q2 it initially resembles the pattern of the male wage with multiple large dips during our sample. However, towards the end of the sample the large increases in wives' wages result in an increase of about 26% percent over the sample period. At Q3 the large shifts in the wive's wage distribution become more important. The household wage shows a dip in the 1980's and a prolonged decrease in the 1990's but over the whole sample period it increases by 43%. The trend at D9 is consistent with the male pattern at this decile combined with those of females at any of the higher quantiles. This produces a steadily rising profile and an increase of almost

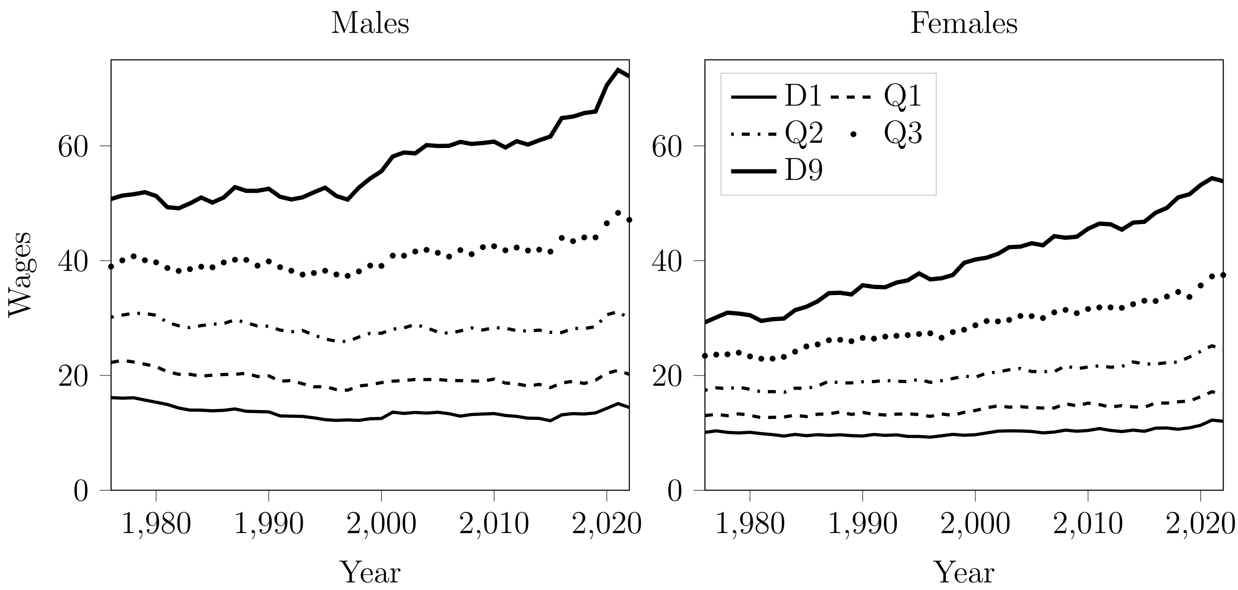


Figure 3: Time series of wages of married males and females at different quantiles.

70%.

Consider the implications of these trends for inequality. For males, females and households the Q3/Q1 ratios increased from 1.75, 1.80 and 1.58 in 1976 to 2.33, 2.23 and 1.96 in 2022. The D9/D1 ratios increase from 3.14. 2.90 and 2.39 in 1976 to 5.00, 4.48 and 3.72 in 2022. While we cannot directly infer anything definitive about sorting via the relative magnitudes of these measures we report them to reflect the extent of inequality. However, the rate of growth over the time period is similar for each which appears to suggest positive assortative matching. Fernández-Val et al. (2018) find that for the period 1976-2016 the D9/D1 ratio for all males increases from 3.6 to 5.4 and that of females increases from 3.7 to 5. This suggests that there is less wage inequality for married individuals. This appears to primarily reflect the relatively higher wages of married individuals at low quantiles.

3 Marital Sorting

We capture marital sorting patterns by reporting the propensity of the male in the j^{th} decile of the male wage distribution to be married to a female in the k^{th} decile of the female wage distribution. We consider $j,k = 1 \ldots 10$ producing 100 cells.

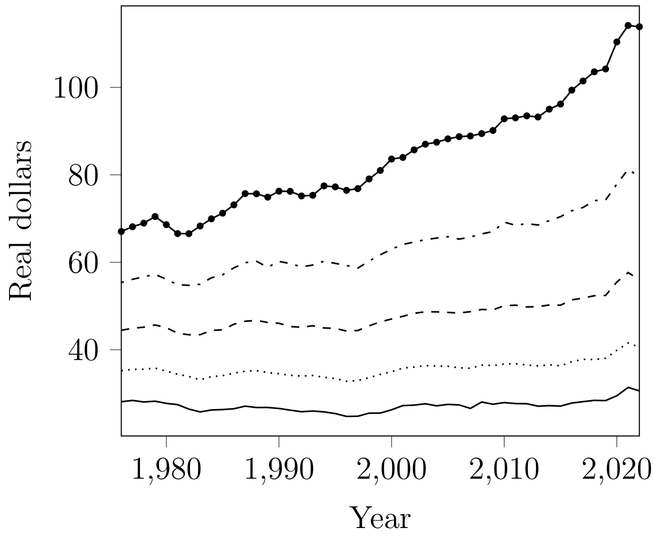


Figure 4: Time series of the households' hourly wage at different quantiles.

To overcome small sample sizes we aggregate the data into five year intervals. To contrast the changes in cell sizes over our entire sample period, we compare the beginning and end periods corresponding to 1976-80 and 2018-22. Tables 1 and 2 report the contingency tables for these two periods. As random sorting is consistent with each cell containing 0.01 of the data, we divide the sample frequencies by this number. Deviations from 1 are suggestive of sorting behavior. Perfect positive assortative matching implies all couples should appear on the diagonal going from the top left to the bottom right with each element having a value of 10. Perfect negative assortative matching implies the elements on the diagonal going from the bottom left to the top right have value 10. Note that dj denotes that an observation is located between the D(j-1) to D(j) decile.

We acknowledge that a number of factors may generate departures from random sorting. For example, individuals are likely to sort on observed factors such as age, education, race and the individual's region of residence. As these factors are determinants of wages this may spuriously support positive sorting on wages. There may also be wage variations reflecting regional specific cost of living influences. While cost of living components of wages may also partially capture unobserved factors, we do not take a position on whether this necessarily reflects sorting.

For the 1976-80 period there is evidence of non-random sorting. The most striking features of Table 1 are the cells associated with the extreme values of the joint wage distribution. The two largest observed frequencies correspond to both spouses in the bottom decile (2.19) and both in the top decile (2.39). The next two highest values correspond to the cells immediately adjacent these extremes (1.72 and 1.59). This is consistent with positive sorting. If we define the bottom as (d1+d2) and the top as (d9+d10) we obtain 6.56% and 6.60%. This contrasts with the 4% implied by random sorting. Another interesting feature of the table is revealed by the off-diagonal values. There is an almost monotonically decreasing relationship between distance from the diagonal and the probability of marriage. Finally, the sum of the elements on the diagonal is 13.78, which appears to support positive assortative matching. However, as this ignores behavior away from the diagonal and it is not invariant to the definition of the diagonal we computed the Kendall rank correlation coefficient. Its value of 0.18 supports positive assortative matching.

Table 2 presents the corresponding values for the 2018-2022 period. As this period is associated with a substantial increase in wage inequality, the level of household inequality is likely to have increased even if the cell frequencies remained at the 1976-80 values. However, it appears that increased marital sorting has exacerbated the individual level inequality. For example, the two extreme cells have increased to 2.94 and 3.25. This represents a particularly large increase in the d1/d1 cell. The frequencies in the bottom 2 and top 2 have increased to 8.13 and 8.48. This reflects a large growth in the frequency in which lower (higher) paid males are married to lower (higher) paid females. The growth in the positive sorting at the bottom is concerning given the manner in which wages, relative to 1976, have fallen for males and only slightly increased for females at these locations of their distributions. The other features of Table 1 regarding sorting are also generally supported by Table 2. There is greater evidence that the lowly paid are married to the lowly paid while the highly paid are married to the highly paid. The sum of the diagonal is now 17.24 and the Kendall rank correlation coefficient is 0.27. This suggests increased positive assortative matching relative to the 1976-80 period.

The differences across the two tables are important given their potential implications for inequality. However, the differences might reflect a variety of factors. As employment rates have changed substantially over the sample period, the com-

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.191**	1.287	1.158	0.937	0.833	0.899	0.77	0.684	0.66	0.582
	d2	1.723	**1.352**	1.203	1.069	0.958	0.839	0.803	0.767	0.755	0.531
	d3	1.272	1.394	**1.099**	1.108	1.12	0.872	0.866	0.86	0.758	0.654
	d4	1.17	1.4	1.194	**1.117**	0.925	0.916	0.905	0.899	0.779	0.693
	d5	0.899	1.12	1.224	1.12	**1.105**	0.973	0.955	0.94	0.911	0.755
	d6	0.74	1.036	1.063	1.14	1.063	**1.111**	1.003	1.024	0.955	0.866
	d7	0.612	0.782	1.102	1.036	1.087	1.099	**1.042**	1.122	1.146	0.97
	d8	0.579	0.609	0.827	1.134	1.078	1.108	1.152	**1.125**	1.218	1.17
	d9	0.436	0.618	0.702	0.758	1.146	1.203	1.275	1.254	**1.224**	1.385
	d10	0.379	0.403	0.43	0.579	0.687	0.982	1.227	1.325	1.594	**2.391**

Table 1: Frequencies of the combination of deciles of married couples for the period 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10 ´
Wives' quantiles	d1	**2.946**	1.56	1.06	0.881	0.874	0.774	0.527	0.579	0.445	0.353
	d2	1.862	**1.771**	1.335	1.085	0.981	0.817	0.643	0.658	0.5	0.347
	d3	1.268	1.71	**1.408**	1.21	0.987	0.881	0.85	0.64	0.57	0.476
	d4	0.887	1.21	1.603	**1.234**	1.161	0.987	0.957	0.759	0.667	0.536
	d5	0.75	0.975	1.207	1.353	**1.122**	1.082	1.045	1.03	0.808	0.628
	d6	0.64	0.683	0.89	1.295	1.445	**1.225**	1.143	1.042	0.917	0.719
	d7	0.488	0.686	0.932	1.009	1.222	1.335	**1.2**	1.149	1.109	0.871
	d8	0.46	0.527	0.661	0.741	0.942	1.253	1.536	**1.311**	1.411	1.158
	d9	0.375	0.503	0.549	0.677	0.75	0.969	1.234	1.509	**1.774**	1.661
	d10	0.326	0.375	0.354	0.518	0.515	0.677	0.866	1.323	1.798	**3.249**

Table 2: Frequencies of the combination of deciles of married couples for the period 2018-2022.

position, in terms of observed and unobserved characteristics, of both husbands and wives may have changed and this may have affected the observed patterns of sorting. It is also possible that the prices of these observed and unobserved characteristics have changed. Finally, the nature of marital sorting, as measured by the joint distribution of the couples observed and unobserved characteristics, may have changed.

Many couples share characteristics which are determinants of wages, such as race, geographical location and age, although this does not necessarily reflect sorting on wages. Another important characteristic is education although this might be more reasonably interpreted as indicative of productivity. To examine how these characteristics are allocated across cells, Tables B1-B15 report the average value of some measure of each of these characteristics for the same time periods.

Tables B1-B4 report the ages of wives and husbands. As age is a determinant of wages and spouses generally are close in age, it is possible that the patterns in Tables 1 and 2 simply reflect age differences. For the first period the average age for both husbands and wives increases by about 2 years as one goes from the d1/d1 to d10/d10. This seems a remarkably small difference given the large wage discrepancies across these cells. The highest husband age is associated with the highest male decile and the highest ages of wives are for the lowest paid women marrying these men. While this is an interesting result it is beyond the scope of the paper to investigate it further. There are differences across the cells, but it does not appear that age differences are the factors driving the observed sorting. For the later period the average ages of the spouses increase by approximately 2.5 years as one goes from d1/d1 to d10/d10. In percentage terms this is a notable increase compared to the earlier period although it does not appear to be the driving force of increased inequality across the two periods. The oldest males continue to be the highest paid married to the lowest paid women and the oldest wives are those married to these men. The age differences across cells is larger than in the earlier period but are unlikely to explain the observed wage differences.

Tables B5 to B9 report the fraction of wives and husbands with university education in the two time periods. Unsurprisingly, educational levels in the higher wage cells are higher than those in the lower. The large increase in individuals obtaining college education is reflected in the substantially higher averages for the later period. We also report the percentage of households in each cell for which both spouses have university education. For the earlier period the product of the two d10 cells is .353(.552*.64) although the corresponding cell in Table B7 is .467. This difference is supportive of couples sorting on education at high wages. For the later period the corresponding numbers are .868(.938*.926) and .897. Thus while there are more married couples with both spouses university educated, there appears to be more positive sorting on education at high wages in the earlier period. The product of the d1 cells for the first period is .004(.064*.069) while for Table B7 it is .034. For the later period the corresponding numbers are .012(.107*.115) and .059. This also suggests there is greater positive sorting on education in the earlier period. Tables B11 to B15 unsurprisingly indicate that race has some association with an individual's location in each of the wage distribution and that the strength of this relationship has changed drastically over time. Moreover, Tables B11-B15 collectively confirm positive sorting on race. This highlights the necessity to account for these factors in estimating the sorting models below. The final characteristic we consider is the location of residence. This is also likely to affect wage differences as it may capture cost of living differences. Tables B17 and B18 suggest that the higher paid are living in metropolitan areas and that the fraction of individuals living in these areas has increased substantially over the sample period. As married individuals generally cohabitate this may also spuriously imply sorting on wages.

4 Econometric model

4.1 Determinants of Marital Sorting

We now focus on estimating the determinants of the marital sorting frequencies in Tables 1 and 2. We do so via the bivariate distribution regression (BDR) approach of Fernández-Val et al. (2023) which employs a Local Gaussian Representation (LGR) of the joint distribution of the wives' and husbands' wages (Chernozhukov, Fernández-Val, and Luo 2019). We represent this joint distribution as:

$$F_{Y_w,Y_h|X}(y_w, y_h \mid x) = \Phi_2(\mu_w(y_w, x), \mu_h(y_h, x); \rho(y_w, y_h, x)), \quad (y_w, y_h) \in \mathbb{R}^2, \tag{1}$$

where Y_w and Y_h are the observed wages of wives and husbands, X is a set of observed characteristics, $\Phi_2(\cdot, \cdot; \rho)$ is the standard bivariate normal CDF with correlation ρ, and $\mu_j(y, x)$ is formally defined below. In the LGR, the marginal conditional CDFs of Y_w and Y_h are represented by:

$$F_{Y_j|X}(y \mid x) = \Phi(\mu_j(y, x)), \quad y \in \mathbb{R}, \quad j \in \{w, h\},$$

where Φ is the standard univariate normal CDF. The parameter $\rho(y_w, y_h, x)$ is the local correlation between the unobservables influencing the spouses' wages at (y_w, y_h)

and captures sorting on unobservables. The unconditional joint distribution of the wives' and husbands' wages can be obtained from the LGR as:

$$F_{Y_w,Y_h}(y_w, y_h) = \int \Phi_2(\mu_w(y_w, x), \mu_h(y_h, x); \rho(y_w, y_h, x))dF_X(x), \quad (y_w, y_h) \in \mathbb{R}^2,$$

where F_X is the CDF of X, and the corresponding marginals are:

$$F_{Y_j}(y) = \int \Phi(\mu_j(y, x))dF_X(x), \quad y \in \mathbb{R}, \quad j \in \{w, h\}.$$

Chernozhukov, Fernández-Val, and Luo (2019) show that the LGR is non-parametric as it does not impose any restrictions on the conditional joint distribution.

The BDR model augments the LGR with two assumptions:

Assumption 1 (BDR) *(1) $\mu_j(y, x) = P_j(x)'\beta_j(y)$, $j \in \{w, h\}$, where P_w and P_h are transformations of x and $\beta_w(y)$ and $\beta_h(y)$ are vectors of coefficients; and (2) $\rho(y_w, y_h, x) = \rho(y_w, y_h)$.*

We allow different specifications for P_w and P_h. For example, P_w includes the wife's education and age while P_h does not. While we assume that the marital sorting parameter does not depend on observed characteristics, we allow it to vary by location in the joint distribution of wages. This model is semiparametric as the parameters $y \mapsto \beta_j(y)$, $j \in \{w, h\}$, and $(y_w, y_h) \mapsto \rho(y_w, y_h)$ are function-valued.

The BDR model describes the joint distribution of wages conditional on the two indices capturing the observed determinants of husbands and wives' wages. Given a random sample of (Y_w, Y_h, X), estimation is performed via a series of bivariate probits of the indicators $\mathbf{1}(Y_w \leq y_w)$ and $\mathbf{1}(Y_h \leq y_h)$ on X for multiple values of y_h and y_w. This is done in two steps. First, estimate $\beta_j(y_j)$ via univariate probit of $\mathbf{1}(Y_j \leq y_j)$ on X, for $j \in \{w, h\}$. Second, estimate $\rho(y_w, y_h)$ via bivariate probit plugging-in the estimates of $\beta_w(y_w)$ and $\beta_h(y_h)$.

We begin by examining our capacity to explain the variation in Tables 1 and 2 via 100 bivariate probits. The entries correspond to estimates of:

$$
\begin{aligned}
s(\underline{y}_w, \overline{y}_w, \underline{y}_w, \overline{y}_w) &:= \frac{\Pr(\underline{y}_w < Y_w \le \overline{y}_w, \underline{y}_h < Y_h \le \overline{y}_h)}{\Pr(\underline{y}_w < Y_w \le \overline{y}_w)\Pr(\underline{y}_h < Y_h \le \overline{y}_h)} \\
&= \frac{F_{Y_w,Y_h}(\overline{y}_w, \overline{y}_h) - F_{Y_w,Y_h}(\overline{y}_w, \underline{y}_h) - F_{Y_w,Y_h}(\underline{y}_w, \overline{y}_h) + F_{Y_w,Y_h}(\underline{y}_w, \underline{y}_h)}{[F_{Y_w}(\overline{y}_w) - F_{Y_w}(\underline{y}_w)][F_{Y_h}(\overline{y}_h) - F_{Y_h}(\underline{y}_h)]}, \quad (2)
\end{aligned}
$$

where $\underline{y}_w$ and $\overline{y}_w$ are evaluated at the sample deciles of Y_w, and $\underline{y}_h$ and $\overline{y}_h$ at the sample deciles of Y_h. The predicted probabilities are reported in Tables B19 for 1976-1980 and Table B20 for 2018-2022. A comparison of the predicted and empirical probabilities for each of the sample periods indicates that the estimated model reproduces the empirical probabilities across the 100 cells despite the restrictions imposed by Assumption 1.[3]

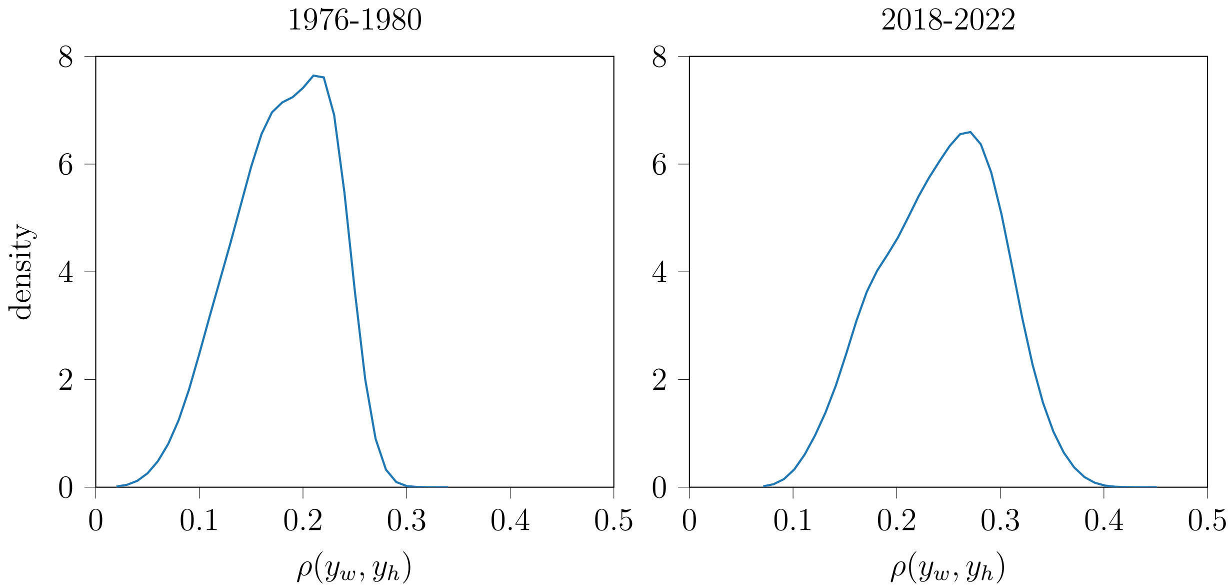
Figure 5 presents the estimated densities of the 100 estimates of $\rho(y_w, y_h)$ for each sample period. The average values for the two periods are 0.18 and 0.24 and the figure reveals a clear shift in the distribution to the right over time. Thus, despite the rich nature of X, there remains a strong and positive correlation between the unobservables driving wages for spouses and it has increased over time.[4] This supports the presence of positive and increasing assortative matching on wages. While this may partially reflect common influences on wages, such as local adjustments related to cost of living, it may also capture unobserved ability or other unobserved determinants of productivity.

While the estimate of $\rho(y_w, y_h)$ is consistent with positive sorting, it is not immediately clear from the coefficient how economically important this parameter is in determining the observed patterns. We employ the estimates from the BDR model to construct counterfactual contingency tables from setting $\rho(y_w, y_h) = 0$ for all

[3]The predicted probabilities are computed by first using univariate distribution regression models to obtain the quantiles of the marginal distribution of wives and husbands. That is, we estimate the quantiles for wives, $Q^w_\tau; \tau \in [0,1]$ solving the empirical analog of $\int \Phi(x\beta(Q^w_\tau))dF_{X_w}(x) = \tau$, where $\beta(\cdot)$ is estimated by univariate distribution regression.The quantiles for husbands are estimated similarly. Based on these quantiles, we can estimate the bivariate distribution regression model as in (1) and estimate the empirical analog of $\int \Phi_2(x_w\beta(Q^w_\tau), x_h\beta(Q^h_\tau), \rho)dF_{X_w,X_h}(x_w, x_h)$.

[4]X includes 3 dummy variables for education (high school, some college, college degree or higher), age, age^2, age interacted with 3 education dummy variables, age^2 interacted with 3 education dummy variables, a dummy variable for non-white, a dummy variable for Hispanic, 2 dummy variables for metropolitan area (central city, outside central city), and 7 regional dummy variables (middle Atlantic, east north central, west north central, south atlantic, east south central, west south central, mountain, and pacific. base: New England).

Figure 5: Estimated densities of $\rho(y_w, y_h)$



y_w, y_h.[5] This corresponds to no correlation or sorting on unobservables while retaining the sorting on the observed characteristics. We implement this counterfactual in the two periods and report the estimates in Tables 3 and 4. The predicted probabilities indicate that sorting on unobservables accounts for a large fraction of the observed positive marital sorting. This is remarkable given the positive sorting that occurs on race, age, educational attainments and location of residence highlighted above. Even though it is hard to make any conjectures about the source of the unobserved heterogeneity, Eika, Mogstad, and Zafar (2019) show for Norwegian data that the choice of college major is an important source of educational sorting. This would be captured in $\rho(y_w, y_h)$.

4.2 Marital Sorting with Employment Selection

The evidence above is based on the subpopulation of working married couples as the BDR model above does not account for selection into employment. As our sample period witnessed drastic increases in the market participation of married females it is possible that both the compositions of the subpopulations of working married

[5]As the quantiles of the marginal distributions are not affected by setting $\rho(y_w, y_h)$ to zero, we can simply estimate the corresponding distributions in (1) by setting $\rho(y_w, y_h)$ equal to zero.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**1.326**	1.233	1.08	1.065	1.008	0.944	0.914	0.871	0.828	0.697
	d2	1.271	**1.223**	1.088	1.093	1.04	0.98	0.956	0.912	0.87	0.734
	d3	1.132	1.124	**1.018**	1.039	1.005	0.949	0.936	0.906	0.871	0.746
	d4	1.097	1.135	1.046	**1.083**	1.058	1.001	0.992	0.972	0.94	0.813
	d5	1.036	1.091	1.027	1.086	**1.077**	1.032	1.023	1.024	0.993	0.9
	d6	0.914	0.977	0.925	0.994	0.995	**0.96**	0.966	0.978	0.957	0.916
	d7	0.914	0.98	0.949	1.027	1.046	1.026	**1.042**	1.058	1.058	1.059
	d8	0.814	0.892	0.877	0.977	1.005	1.01	1.026	**1.065**	1.073	1.122
	d9	0.784	0.861	0.854	0.965	1.015	1.047	1.089	1.14	**1.175**	1.28
	d10	0.656	0.708	0.723	0.844	0.905	0.977	1.069	1.156	1.266	**1.601**

Table 3: Results of estimated sorting measures in 1976-1980 at different quantiles when $\rho_{Y_w^*,Y_h^*} = 0$.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**1.525**	1.375	1.236	1.131	0.991	0.933	0.834	0.739	0.683	0.553
	d2	1.351	**1.301**	1.229	1.15	1.021	0.978	0.884	0.781	0.724	0.581
	d3	1.212	1.218	**1.189**	1.134	1.028	1.01	0.926	0.833	0.792	0.659
	d4	1.097	1.132	1.141	**1.114**	1.025	1.031	0.967	0.897	0.873	0.724
	d5	0.965	1.024	1.059	1.069	**1.02**	1.028	1.009	0.962	0.989	0.875
	d6	0.889	0.966	1.019	1.039	1.01	**1.047**	1.041	1.005	1.046	0.938
	d7	0.809	0.885	0.958	0.991	0.986	1.048	**1.061**	1.069	1.154	1.039
	d8	0.73	0.803	0.879	0.927	0.962	1.046	1.081	**1.115**	1.253	1.205
	d9	0.66	0.717	0.786	0.845	0.922	1.009	1.092	1.174	**1.386**	1.408
	d10	0.569	0.627	0.691	0.77	0.876	0.962	1.093	1.218	1.488	**1.706**

Table 4: Results of estimated sorting measures in 2018-2022 at different quantiles when $\rho_{Y_w^*,Y_h^*} = 0$.

females and the working married males with working spouses have changed. We now incorporate the participation decision of each spouse, while allowing for a relationship between these decisions, by extending the BDR model to incorporate endogenous employment decisions. This represents an application of the CFL approach with some provisos we outline below. We begin with some necessary preliminaries.

4.3 The sample selection model

Consider the vector of random variables $(D_w^*, D_h^*, Y_w^*, Y_h^*)$, where D_w^* and D_h^* are latent variables that determine the employment decision of the wife and husband, and Y_w^* and Y_h^* are the offered wages to the wife and husband. Z is a vector of observed characteristics of which X is a subset . We make the following assumptions regarding the joint CDF of $(D_w^*, D_h^*, Y_w^*, Y_h^*)$ conditional on Z:

Assumption 2 (LGR, Relevance and Exclusion) *(1) For* $(d_w, d_h, y_w, y_h) \in \mathbb{R}^4$,

$$F_{D_w^*,D_h^*,Y_w^*,Y_h^*|Z}(d_w, d_h, y_w, y_h \mid z) = \Phi_4(\boldsymbol{\mu}(d_w, d_h, y_w, y_h, z); \boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, z)),$$

where $\Phi_4(\cdot; \boldsymbol{\Sigma})$ *is the standard tetravariate normal CDF with correlation matrix* $\boldsymbol{\Sigma}$,

$$\boldsymbol{\mu}(d_w, d_h, y_w, y_h, z) = \begin{pmatrix} \mu_{D_w^*}(d_w, z) \\ \mu_{D_h^*}(d_h, z) \\ \mu_{Y_w^*}(y_w, z) \\ \mu_{Y_h^*}(y_h, z) \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, z) =$$

$$\begin{pmatrix} 1 & \rho_{D_w^*,D_h^*}(d_w, d_h, z) & -\rho_{D_w^*,Y_w^*}(d_w, y_w, z) & -\rho_{D_w^*,Y_h^*}(d_w, y_h, z) \\ \rho_{D_w^*,D_h^*}(d_w, d_h, z) & 1 & -\rho_{D_h^*,Y_w^*}(d_h, y_w, z) & -\rho_{D_h^*,Y_h^*}(d_h, y_h, z) \\ -\rho_{D_w^*,Y_w^*}(d_w, y_w, z) & -\rho_{D_h^*,Y_w^*}(d_h, y_w, z) & 1 & \rho_{Y_w^*,Y_h^*}(y_w, y_h, z) \\ -\rho_{D_w^*,Y_h^*}(d_w, y_h, z) & -\rho_{D_h^*,Y_h^*}(d_h, y_h, z) & \rho_{Y_w^*,Y_h^*}(y_w, y_h, z) & 1 \end{pmatrix}$$

is non-singular almost everywhere in z; (2) $\mu_{D_w^*}(d_w, z) \neq \mu_{D_w^*}(d_w, z')$ *and* $\mu_{D_h^*}(d_h, z'') \neq \mu_{D_h^*}(d_h, z''')$ *for some* $z = (z_1, x)$, $z' = (z_1', x)$, $z'' = (z_1'', x)$, *and* $z''' = (z_1''', x)$; *and (3)* $\mu_{Y_w^*}(y_w, z) = \mu_{Y_w^*}(y_w, x)$, $\mu_{Y_h^*}(y_h, z) = \mu_{Y_h^*}(y_h, x)$, $\rho_{D_w^*,Y_w^*}(d_w, y_w, z) = \rho_{D_w^*,Y_w^*}(d_w, y_w, x)$, *and* $\rho_{D_h^*,Y_h^*}(d_h, y_h, z) = \rho_{D_h^*,Y_h^*}(d_h, y_h, x)$, *for* $z = (z_1, x)$.

Assumption 2(1) is similar to the LGR of a joint CDF. In contrast to the bivariate case, this representation restricts some features of the joint tetravariate distribution. While the univariate and bivariate marginals remain unrestricted, some restrictions are imposed on the trivariate marginals and joint tetravariate distributions. To highlight this, note that the local dependence between any pair of random variables, as measured by the corresponding component of the matrix $\boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, z)$, does not depend on the value of the other components. For example, local pairwise independence of all the components, that is $\boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, z)$ equal to the identity matrix, implies joint local independence of all the components.

Assumption 2(3) embodies exclusion restrictions on the marginal distributions of Y_w^* and Y_h^*, and the local dependence matrix $\boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, z)$. Thus, Y_w^* and Y_h^* are independent of the components of Z not included in X. Moreover, these components do not affect the local dependence between all the components in $(D_w^*, D_h^*, Y_w^*, Y_h^*)$. Assumption 2(2) is a relevance condition of these excluded components of Z on the expectations of D_w^* and D_h^*.

The observed variables (D_w, D_h, Y_w, Y_h) are related to the latent variables as:

$$D_w = \mathbf{1}(D_w^* \leq 0),$$
$$D_h = \mathbf{1}(D_h^* \leq 0),$$

where D_w (D_h) equals 1 when the wife (husband) is working FTFY and 0 otherwise. Moreover,

$$Y_w = Y_w^* \text{ if } D_w = 1 \text{ and } D_h = 1$$
$$Y_h = Y_h^* \text{ if } D_w = 1 \text{ and } D_h = 1$$

where Y_w and Y_h are the FTFY hourly wages of the wife and husband. These are only observed for FTFY working couples.

We show in Appendix A that all the parameters of the LGR of the latent variables are identified from the distribution of the observed variables.

Theorem 1 (Identification Under Employment Selection) *Under Assumption 2, $\boldsymbol{\mu}(0, 0, y_w, y_h, z)$ and $\boldsymbol{\Sigma}(0, 0, y_w, y_h, z)$ are identified from the joint distribution of (D_w, D_h, Y_w, Y_h, Z).*

5 Measures of sorting in the presence of selection

We consider how the expressions for the contingency tables entries change when we account for endogenous sample selection. They are now written:

$$
\begin{aligned}
& s(\underline{y}_w, \overline{y}_w, \underline{y}_h, \overline{y}_h \mid D_w = 1, D_h = 1) = \\
& \frac{\Pr(\underline{y}_w < Y_w \leq \overline{y}_w, \underline{y}_h < Y_h \leq \overline{y}_h \mid D_w = 1, D_h = 1)}{\Pr(\underline{y}_w < Y_w \leq \overline{y}_w \mid D_w = 1, D_h = 1) \Pr(\underline{y}_h < Y_h \leq \overline{y}_h \mid D_w = 1, D_h = 1)},
\end{aligned} \tag{3}
$$

which corresponds to the ratio of the joint distribution of (Y_w, Y_h) to the product of the marginals in the selected population. It is equal to 1 when the wages of the wives and husbands are independent from each other in the selected population.

The values of s are identified from the wage data among couples in which both partners work and does not depend on the identification of the sample selection model. Nevertheless, it is interesting to understand how changes in the sorting measure can be attributed to changes in the model's parameters. This can be conducted via counterfactuals. Note that the numerator of this sorting measure can be written as:

$$
\int_{\mathcal{Z}} \Pr(\underline{y}_w < Y_w \leq \overline{y}_w, \underline{y}_h < Y_h \leq \overline{y}_h \mid D_w = 1, D_h = 1, Z = z) dF_{Z \mid D_w, D_h}(z \mid 1, 1) \tag{4}
$$

where the integrand equals:

$$
\begin{aligned}
& \Pr(\underline{y}_w < Y_w \leq \overline{y}_w, \underline{y}_h < Y_h \leq \overline{y}_h \mid D_w = 1, D_h = 1, Z = z) = \\
& \frac{\Pr(\underline{y}_w < Y_w \leq \overline{y}_w, \underline{y}_h < Y_h \leq \overline{y}_h, D_w = 1, D_h = 1 \mid Z = z)}{\Pr(D_w = 1, D_h = 1 \mid Z = z)} = \\
& \frac{\Pr(\underline{y}_w < Y_w^* \leq \overline{y}_w, \underline{y}_h < Y_h^* \leq \overline{y}_h, D_w^* \leq 0, D_h^* \leq 0 \mid Z = z)}{\Phi_2(\mu_{D_w^*}(0, z), \mu_{D_h^*}(0, z), \rho_{D_w^*, D_h^*}(0, 0, x))},
\end{aligned}
$$

with

$$
\begin{aligned}
& \Pr(\underline{y}_w < Y_w^* \leq \overline{y}_w, \underline{y}_h < Y_h^* \leq \overline{y}_h, D_w^* \leq 0, D_h^* \leq 0 \mid Z = z) = \\
& \Phi_4(\boldsymbol{\mu}(0, 0, \overline{y}_w, \overline{y}_h, z); \boldsymbol{\Sigma}(0, 0, \overline{y}_w, \overline{y}_h, x)) - \Phi_4(\boldsymbol{\mu}(0, 0, \underline{y}_w, \overline{y}_h, z); \boldsymbol{\Sigma}(0, 0, \underline{y}_w, \overline{y}_h, x)) \\
& - \Phi_4(\boldsymbol{\mu}(0, 0, \overline{y}_w, \underline{y}_h, z); \boldsymbol{\Sigma}(0, 0, \overline{y}_w, \underline{y}_h, x)) + \Phi_4(\boldsymbol{\mu}(0, 0, \underline{y}_w, \underline{y}_h, z); \boldsymbol{\Sigma}(0, 0, \underline{y}_w, \underline{y}_h, x)).
\end{aligned}
$$

5.1 Counterfactuals based on a specific time period

For each time period counterfactuals can be obtained by replacing $\boldsymbol{\Sigma}(d_w, d_h, y_w, y_h, x)$ with an alternative positive definite $\widetilde{\boldsymbol{\Sigma}}(d_w, d_h, y_w, y_h, x)$ reflecting different model parameters. For example, setting $\rho_{D_w^*, D_h^*}(0, 0, x)$ to zero examines the impact of making the employment decisions for husbands and wives conditionally independent. Setting $\rho_{D_w^*, Y_w^*}(0, y_w, x)$, $\rho_{D_h^*, Y_h^*}(0, y_h, x)$, $\rho_{D_w^*, Y_h^*}(0, y_h, x)$, and $\rho_{D_h^*, Y_w^*}(0, y_w, x)$ to zero eliminates the role of selection while setting $\rho_{Y_w^*, Y_h^*}(y_w, y_h, x)$ to zero eliminates sorting on the unobservables correlated with wages. Sorting may still arise here on the basis of observed characteristics. Another potential counterfactual distribution integrates (4) over the distribution of Z for the whole population of married couples in our sample rather than the population of FTFY working couples.

5.2 Counterfactuals based on different periods of time

To investigate intertemporal changes in sorting we examine counterfactuals over time. We start by rewriting the elements of the contingency matrix in (3) as:

$$
\begin{aligned}
&s(\underline{y}_w, \overline{y}_w, \underline{y}_h, \overline{y}_h \mid D_w = 1, D_h = 1) \\
&= \tfrac{F_{Y_w, Y_h \mid D_w, D_h}(\overline{y}_w, \overline{y}_h \mid 1,1) - F_{Y_w, Y_h \mid D_w, D_h}(\overline{y}_w, \underline{y}_h \mid 1,1) - F_{Y_w, Y_h \mid D_w, D_h}(\underline{y}_w, \overline{y}_h \mid 1,1) + F_{Y_w, Y_h \mid D_w, D_h}(\underline{y}_w, \underline{y}_h \mid 1,1)}{[F_{Y_w \mid D_w, D_h}(\overline{y}_w \mid 1,1) - F_{Y_w \mid D_w, D_h}(\underline{y}_w \mid 1,1)][F_{Y_h \mid D_w, D_h}(\overline{y}_h \mid 1,1) - F_{Y_h \mid D_w, D_h}(\underline{y}_h \mid 1,1)]},
\end{aligned}
$$

where

$$
F_{Y_w, Y_h \mid D_w, D_h}(y_w, y_h \mid 1, 1) = \Pr(Y_w \le y_w, Y_h \le y_h \mid D_w = 1, D_h = 1),
$$

and

$$
F_{Y_j \mid D_w, D_h}(y_j \mid 1, 1) = \Pr(Y_j \le y_j \mid D_w = 1, D_h = 1), \quad j = w, h.
$$

We calculate counterfactual joint distributions assuming employment selection is as in year q, the wage structure is as in year r, and the composition of the work force is as in year s by:

$$F^{q,r,s}_{Y_w,Y_h|D_w,D_h}(y_w, y_h \mid 1,1) := \int \Pr^{q,r}(Y_w \leq y_w, Y_h \leq y_h \mid D_w = 1, D_h = 1, Z = z) dF^s_Z(z) \quad (5)$$

where F^s_Z is the distribution of Z in year s, and

$$\Pr^{qr}(Y_w \leq y_w, Y_h \leq y_h \mid D_w = 1, D_h = 1, Z = z) = \frac{\Phi_4(\boldsymbol{\mu}^{q,r}(0,0,y_w,y_h,z); \boldsymbol{\Sigma}^{q,r}(0,0,y_w,y_h,x))}{\Phi_2(\mu^q_{D^*_w}(0,z), \mu^q_{D^*_h}(0,z), \rho^q_{D^*_w,D^*_h}(0,0,x))},$$

with

$$\boldsymbol{\mu}^{q,r}(0,0,y_w,y_h,z) = \begin{pmatrix} \mu^q_{D^*_w}(0,z) \\ \mu^q_{D^*_h}(0,z) \\ \mu^r_{Y^*_w}(y_w,x) \\ \mu^r_{Y^*_h}(y_h,x) \end{pmatrix}$$

and[6]

$$\boldsymbol{\Sigma}^{q,r}(0,0,y_w,y_h,x) = \begin{pmatrix} 1 & \rho^q_{D^*_w,D^*_h}(0,0,x) & -\rho^q_{D^*_w,Y^*_h}(0,y_h,x) & -\rho^q_{D^*_w,Y^*_h}(0,y_h,x) \\ \rho^q_{D^*_w,D^*_h}(0,0,x) & 1 & -\rho^q_{D^*_h,Y^*_w}(0,y_w,x) & -\rho^q_{D^*_h,Y^*_h}(0,y_h,x) \\ -\rho^q_{D^*_w,Y^*_w}(0,y_1,x) & -\rho^q_{D^*_h,Y^*_w}(0,y_w,x) & 1 & \rho^r_{Y^*_w,Y^*_h}(y_w,y_h,x) \\ -\rho^q_{D^*_w,Y^*_h}(0,y_h,x) & -\rho^q_{D^*_h,Y^*_h}(0,y_h,x) & \rho^r_{Y^*_w,Y^*_h}(y_w,y_h,x) & 1 \end{pmatrix}.$$

Comparable counterfactual marginal distributions can be computed for the denominator.

Comparisons should be based on wage levels of husbands and wives which incorporate the changes in the wage distributions. Since the real wages of males decreased at the bottom of the distribution while the real wages increased over the whole distribution for females, using fixed wage levels is not appropriate. For example, we are more likely to find a husband earning more than 20 dollars an hour with a wife

[6] A practical problem is that there is no guarantee that $\boldsymbol{\Sigma}^{q,r}(0,0,y_w,y_h,x)$ is positive definite. Nevertheless, we did not encounter this problem in our empirical analysis.

earning more than 15 dollars an hour in 2020 than in 1976. This, however, does not imply a change in sorting behavior. Rather, it reflects the changes in the likelihoods from the respective marginal distributions. As we use quantiles rather than fixed wages to make these comparisons, we need the counterfactual quantiles of the corresponding marginal wage distributions. For example, the τ-th quantile of the marginal wage distribution of the wives is:

$$
\begin{aligned}
&Q^{q,r,s}_{Y_w}(\tau) \\
&= \inf\left\{ y \in \mathbb{R} : \int \Pr^{q,r}(Y_w \leq y, Y_h \leq \infty \mid D_w = 1, D_h = 1, Z = z) dF^s_Z(z) \geq \tau \right\}.
\end{aligned} \tag{6}
$$

6 Estimation

We consider a semiparametric BDR model with selection that imposes:

Assumption 3 (BDR with Selection) *(1) $\mu_{Y_j^*}(y, x) = P_j(x)'\beta_j(y)$, $\mu_{D_j^*}(0, z) = Q_j(z)'\gamma_j$, where P_j and Q_j are transformations of x and z, and $\beta_j(y)$ and γ_j are vectors of coefficients, $j \in \{w, h\}$; and (2) $\mathbf{\Sigma}(d_w, d_h, y_w, y_h, x) = \mathbf{\Sigma}(d_w, d_h, y_w, y_h)$.*

6.1 Estimation of the local model parameters

Using Assumption 3 with $P_j(x) = x$ and $Q_j(z) = z$ for $j \in \{w, h\}$ we estimate the parameters in 2 steps:

1. Bivariate probit to obtain γ_w, γ_h and $\rho_{D_w^*, D_h^*} := \rho_{D_w^*, D_h^*}(0, 0)$ using:

$$
\mathbb{P}(D_w = 1, D_h = 1 \mid Z = z) = \Phi_2(z'\gamma_w, z'\gamma_h; \rho_{D_w^*, D_h^*}).
$$

   We denote the estimators as $\widehat{\gamma}_w$, $\widehat{\gamma}_h$ and $\widehat{\rho}_{D_w^*, D_h^*}$.

2. Multivariate probit with sample selection correction to estimate the remaining

parameters using:

$$\mathbb{P}(Y_w \leq y_w; Y_h \leq y_w | D_w = 1; D_h = 1, X = x, Z = z) \propto \Phi_4(z'\gamma_w, z'\gamma_h, x'\beta_w(y_w), x'\beta_h(y_h); \mathbf{\Sigma}(0, 0, y_w, y_h)),$$

which can be estimated by a small adaption of the multivariate probit model after plugging in the first-stage estimators $\widehat{\gamma}_w$, $\widehat{\gamma}_h$ and $\widehat{\rho}_{D^*_w,D^*_h}$.

The first step is standard. The second step is straightforward, but the calculation of higher-order integrals in the multivariate probit is both computationally intensive and imprecise. The imprecision is especially unfortunate when combined with a numerical optimization method that assumes smoothness of the first and second order derivative of the criterion function. Therefore, we employ the GHK importance sampling simulator of Geweke (1991), Hajivassiliou and McFadden (1990) and Keane (1990) to simulate these probabilities. This importance sampling simulator uses the result that multivariate normal distributions, conditional on realizations of one or more elements of the outcome vector, are also normally distributed but with a lower dimension.

6.2 Estimation of the measures of sorting

Estimation of the sorting measures requires estimates of the quantiles of the marginal distributions of wives' and husbands' wage distributions. These are obtained via application of the generalized inverse or rearrangement operator to the plug-in estimator of the distribution. For example:

$$\widehat{Q}^{q,r,s}_{Y_w}(\tau) = \int_0^\infty \mathbf{1}\left\{ \frac{1}{n_s} \sum_{i=1}^{n_s} \widehat{\Pr}^{q,r}(Y_w \leq y \mid D_w = 1, D_h = 1, Z^s = Z_i) \leq \tau \right\} dy, \quad (7)$$

with:

$$\widehat{\Pr}^{q,r}(Y_w \leq y_w \mid D_w = 1, D_h = 1, Z^s = z) = \frac{\Phi_4(z'\widehat{\gamma}^q_w, z'\widehat{\gamma}^q_h, x'\widehat{\beta}^r_w(y_w), \infty; \widehat{\mathbf{\Sigma}}^{q,r}(0, 0, y_w, y_h))}{\Phi_2(z'\widehat{\gamma}^q_w, z'\widehat{\gamma}^q_h; \widehat{\rho}^q_{D^*_w,D^*_h})},$$

for any y_h, where $\widehat{\mathbf{\Sigma}}^{q,r}(0,0,y_w,y_h)$ is the plug-in estimator of $\mathbf{\Sigma}^{q,r}(0,0,y_w,y_h)$, and the integrals are calculated using the GHK importance sampling method.

Equation (7) is solved using bisection implying that we need to calculate the term on the right-hand side of that equation for different trial values of the quantile. The counterfactual quantiles for husbands can be estimated in a similar way.

The estimator of the measures of sorting is:

$$\widehat{s}^{q,r,s}(\underline{y}_w, \overline{y}_w, \underline{y}_h, \overline{y}_h \mid D_w = 1, D_h = 1)$$
$$= \frac{\widehat{F}^{q,r,s}_{Y_w,Y_h|D_w,D_h}(\overline{y}_w,\overline{y}_h|1,1) - \widehat{F}^{q,r,s}_{Y_w,Y_h|D_w,D_h}(\overline{y}_w,\underline{y}_h|1,1) - \widehat{F}^{q,r,s}_{Y_w,Y_h|D_w,D_h}(\underline{y}_w,\overline{y}_h|1,1) + \widehat{F}^{q,r,s}_{Y_w,Y_h|D_w,D_h}(\underline{y}_w,\underline{y}_h|1,1)}{[\widehat{F}^{q,r,s}_{Y_w|D_w,D_h}(\overline{y}_w|1,1) - \widehat{F}^{q,r,s}_{Y_w|D_w,D_h}(\underline{y}_w|1,1)][\widehat{F}^{q,r,s}_{Y_h|D_w,D_h}(\overline{y}_h|1,1) - \widehat{F}^{q,r,s}_{Y_h|D_w,D_h}(\underline{y}_h|1,1)]},$$

where $\widehat{F}^{q,r,s}_{Y_w,Y_h|D_w,D_h}$, $\widehat{F}^{q,r,s}_{Y_w|D_w,D_h}$ and $\widehat{F}^{q,r,s}_{Y_h|D_w,D_h}$ are plug-in estimators of $F^{q,r,s}_{Y_w,Y_h|D_w,D_h}$, $F^{q,r,s}_{Y_w|D_w,D_h}$ and $F^{q,r,s}_{Y_h|D_w,D_h}$ in (5), respectively; and $\underline{y}_j$ and $\overline{y}_j$ evaluated at $\widehat{Q}^{q,r,s}_{Y_j}((i-1)/10)$ and $\widehat{Q}^{q,r,s}_{Y_k}(i/10)$, $i = 1, \ldots, 10$ and $j \in \{w, h\}$.

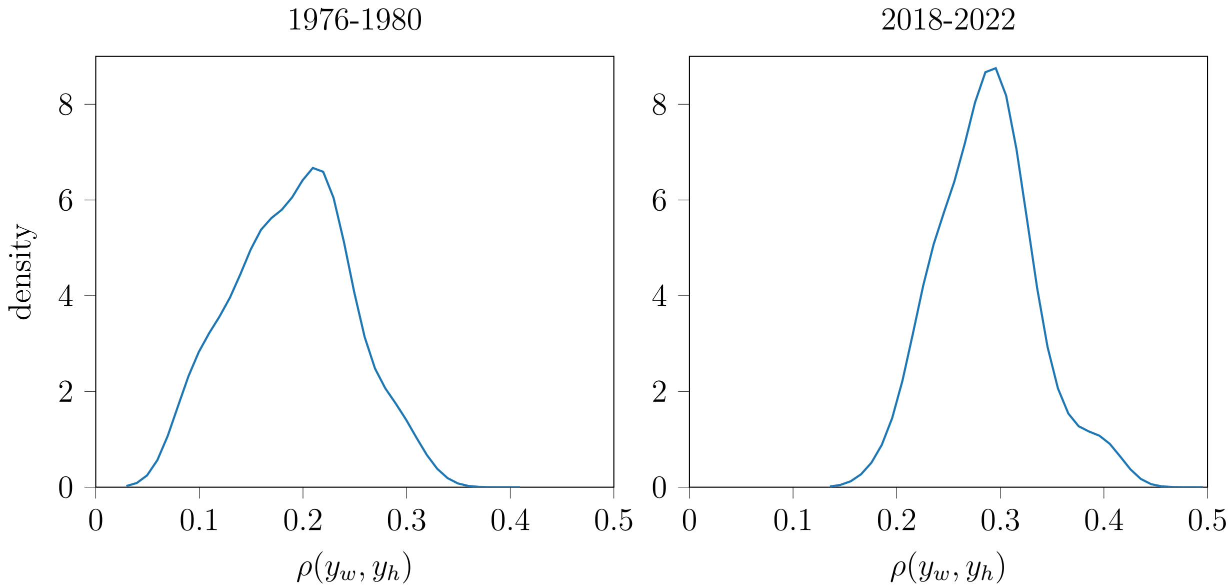
7 Empirical Results

7.1 Parameter Estimates

We estimate the model at each decile for both of the males and females wage distribution for 5 year periods starting from 1976 and ending at 2022. As the total number of years is not divisible by 5 we use two periods of 6 years. These are 1991-1996 and 1997-2002. We estimate one hundred different sets of coefficients for each time period corresponding to the different combinations of deciles for each spouse. As our primary focus is on the role of selection and sorting, we do not report the 100 sets of parameter estimates but Figures 7-9 present the estimates of the different ρ's at the same quartile of both distributions along with their bootstrapped standard errors. Results for the other quantiles are available from the authors. Note that we employ the 100 sets of parameter estimates in conducting the counterfactuals reported below and the discussion of the ρ's which follows is only to provide some insight into the behavior of these selection and sorting parameters.

We discuss our model specification before proceeding to the results. Footnote 4 listed the variables in X employed in the BDR specification. We continue to use these variable but identification now requires some variables in Z which are excluded

Figure 6: Kernel density estimates of $\rho(y_w, y_h)$.



from X. We follow Mulligan and Rubinstein (2008) and employ a dummy variable for children at home, family size, and a dummy variable for children at home under 5 years of age in this role. These variables are seen as contentious and we discuss the implications of these choices below.

Consider the correlation between the unobservables in the spouses' employment equations, $\rho_{D_w^*,D_h^*}$, recalling it is invariant to the location in the wage distributions. The positive value indicates that the unobservables affecting the spouses' work decisions are positively correlated. The coefficient is small in magnitude but is statistically significantly different from zero. It generally increases over time, except for a dip in the middle of the sample period, and reveals that the unobserved factors which make spouses both work FTFY are increasingly correlated over time. We leave a discussion of the marginal impact of changing this, and other parameters, to the following section.

The estimated selection coefficients for wives ($\rho_{D_w^*,Y_w^*}$) and husbands ($\rho_{D_h^*,Y_h^*}$) capture the correlation between the unobservables affecting the individual's respective work decision and their own wages. While in conventional selection models they capture the relationship between the selection decision and the mean wage, we evaluate this correlation at different quantiles of the wage distribution. The estimate for wives is particularly interesting and consistent with earlier evidence using similar identification approaches (see, for example, Mulligan and Rubinstein, 2008, and FVVP, 2023b) for FTFY workers. Similar to these earlier papers we find evidence of negative selection changing to positive selection. At each quantile there is negative selection in the earlier period with an estimate around -0.3 which is statistically different from zero. For the latter sample period the estimate has increased. At Q1/Q1 the estimate has increased to zero, while at Q2/Q2 and Q3/Q3 the estimates are approximately 0.2 and 0.3. As discussed in FVVP (2023b), the sign of $\rho_{D_w^*,Y_w^*}$ is contentious given its interpretation and its implication for selection. They note that the sign of the selection terms appears to reflect the impact of the variables used to explain participation which are excluded from the wage equation. The change of sign appears to capture the changing impact of these variables on the participation decision.[7] While the sign of $\rho_{D_w^*,Y_w^*}$ is controversial we explore the impact of changing its value in counterfactual exercises below.

It is not typical to account for selection when estimating male wage equations. However, it is important to do so here as we do not know apriori the impact of selection in this model. Moreover, as we account for the role of male selection on the female wage it is necessary to model the male work decision. The parameter $\rho_{D_h^*,Y_h^*}$ is very poorly estimated at each of the quantiles reported. At Q1/Q1 and Q2/Q2 it is negative and large but very imprecisely estimated. At Q3/Q3 it is both negative and positive but generally imprecisely estimated. We attribute this result to our inability to identify this parameter from these data.

The parameters $\rho_{D_w^*,Y_h^*}$ and $\rho_{D_h^*,Y_w^*}$ also vary by location in the respective wage distributions and capture the correlation between the unobservables affecting the work decision of the individual with the unobservables affecting the wage of their spouse at a certain quantile in the spouses wage distribution. First consider $\rho_{D_w^*,Y_h^*}$. Figures 7-9 suggest that at each quartile the estimate is both negative and positive at each quantile depending on the time period. However, the confidence bands

[7] FVVP (2023b) show that this negative coefficient for the earlier period is likely to be due to the exclusion restrictions. The reader is referred to that paper for a detailed discussion of how the exclusion restrictions may be generating the selection related results. One could reproduce the counterfactual that follows using the FVVP (2023b) approach which employs an alternative identification strategy but that is not feasible without making a number of additional assumptions.

suggest that we cannot reject that the estimate is zero for many of the periods with the exception of some middle years at the first quartile. $\rho_{D_h^*,Y_w^*}$ captures how the unobservables driving the husband's participation decision affects the wife's wage. Given the evidence above on male selection it is unsurprising that this parameter is imprecisely estimated although it appears to be generally negative and statistically different from zero.

$\rho_{Y_w^*,Y_h^*}$ is an economically important parameter as it captures sorting via the correlation between the unobservables generating the husband's and wife's wages. If individuals are sorting with respect to these unobservables we expect this parameter to be positive. Moreover, as this parameter also captures the prices of these unobservables it will capture similar structural effects associated with the implicit prices of observed characteristics. As noted above, this parameter will capture factors related to influences, such as cost of living or wage premia, which are not captured by the conditioning variables but which are shared by spouses. For example, while the data measure various aspects of the location in which the spouses live we are unable to distinguish between those living in costly urban areas. It will also capture other factors specific to unobserved issues shared by the spouses. For example, it captures that both may work for the same firm and or went to the same college and share the costs and/or benefits associated with their wages. The estimate of this parameter is positive and generally ranges between 0.2 and 0.3. Figure 6 plots the estimated densities for this parameter for the starting and ending time periods for all 100 models while Figures 7-9 present the estimates at the various quantiles we consider. There is some evidence that it is increasing at these different quantiles over time but the confidence intervals do not appear to reject that the impact is constant.

With respect to sorting the only clear and consistent evidence is associated with $\rho_{Y_w^*,Y_h^*}$. The evidence from the counterfactuals based on the BDR estimates revealed that setting $\rho_{Y_w^*,Y_h^*}$ to zero drastically reduced the presence of positive sorting. However, this may now be offset by allowing for selection even though the estimates of $\rho_{D_w^*,D_h^*}$, $\rho_{D_w^*,Y_w^*}$, $\rho_{D_h^*,Y_h^*}$, $\rho_{D_w^*,Y_h^*}$ and $\rho_{D_h^*,Y_w^*}$ do not individually or collectively present a clear story regarding sorting.

Figure 7: Results for covariance matrix at Q1.

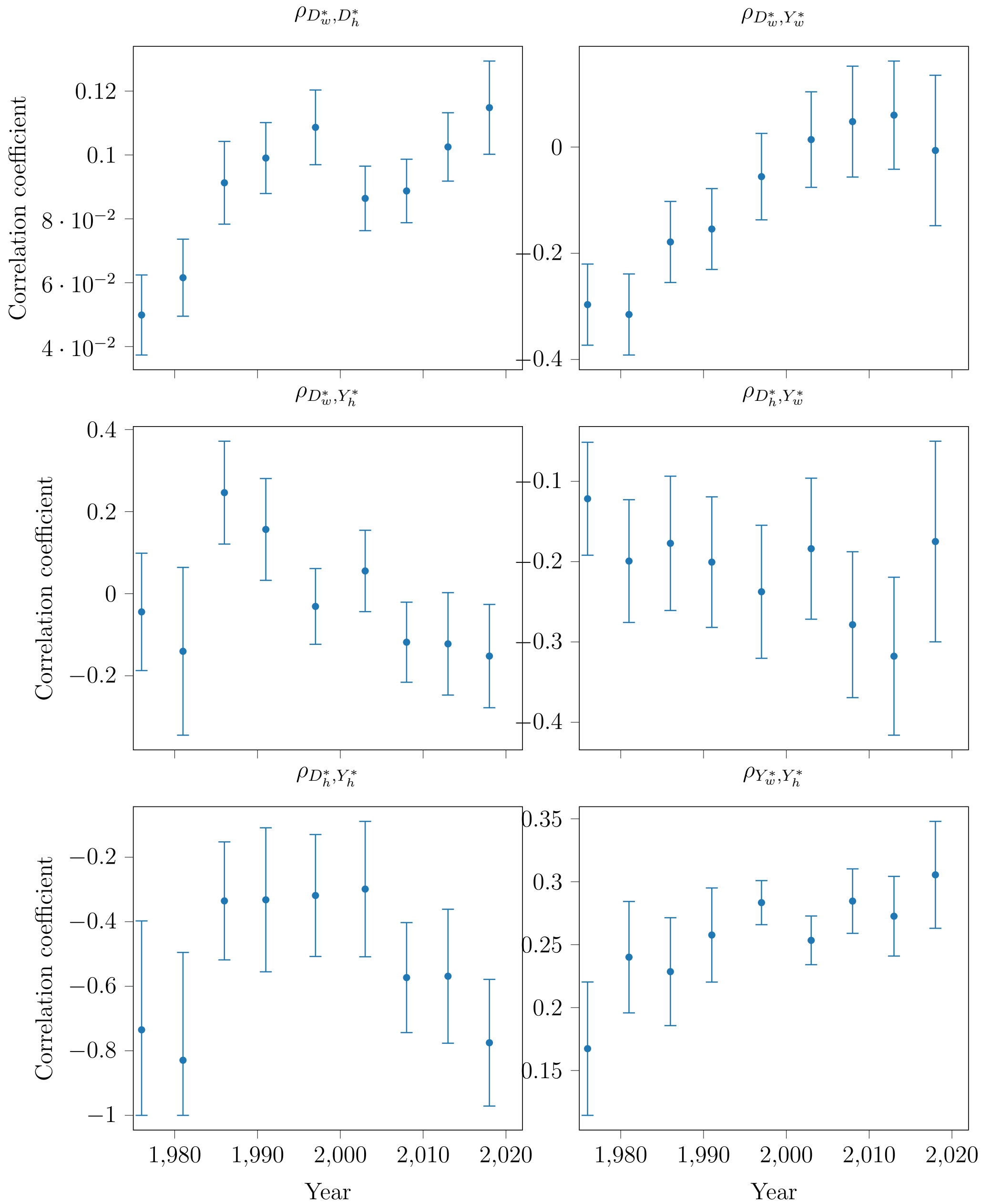


Figure 8: Results for covariance matrix at Q2.

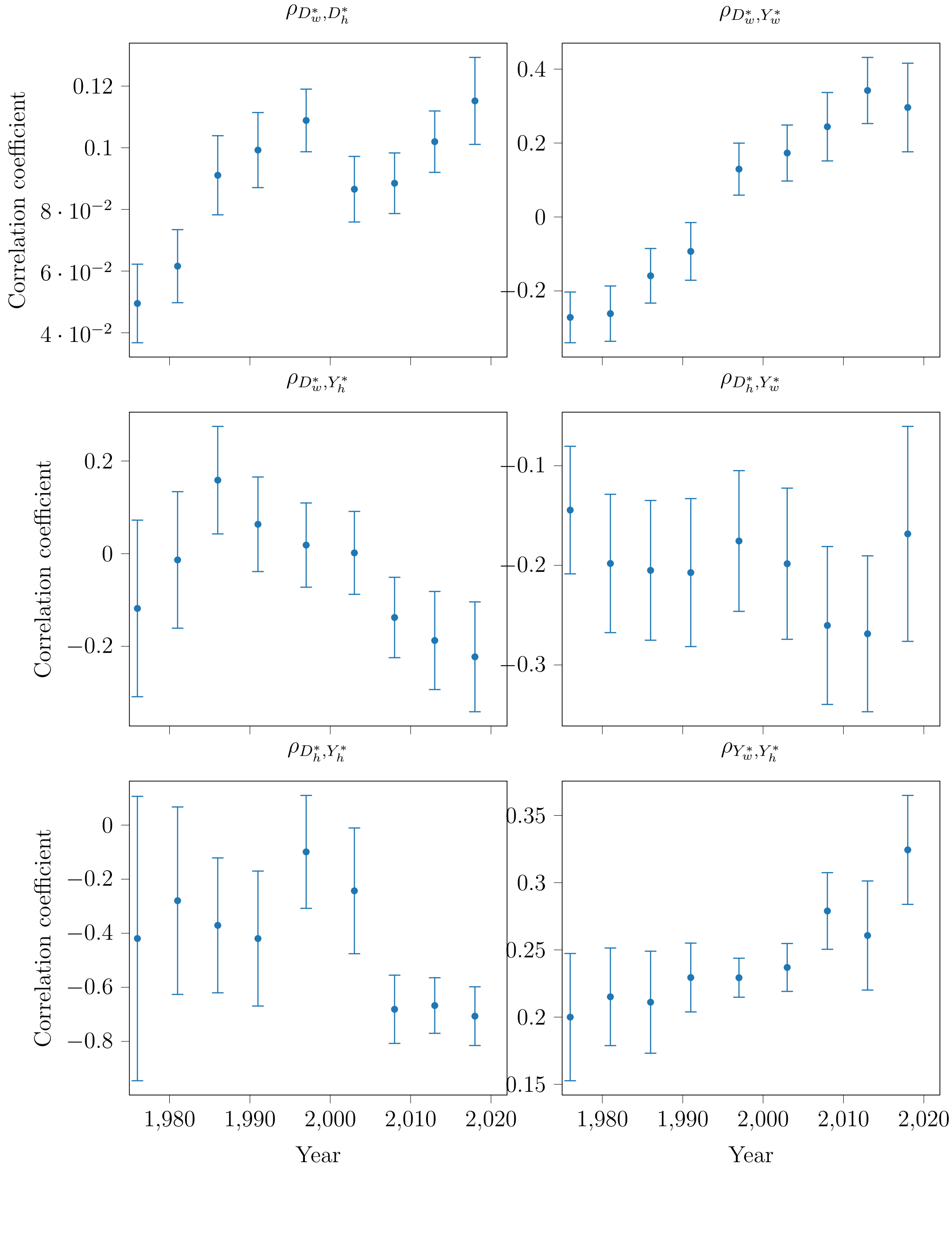
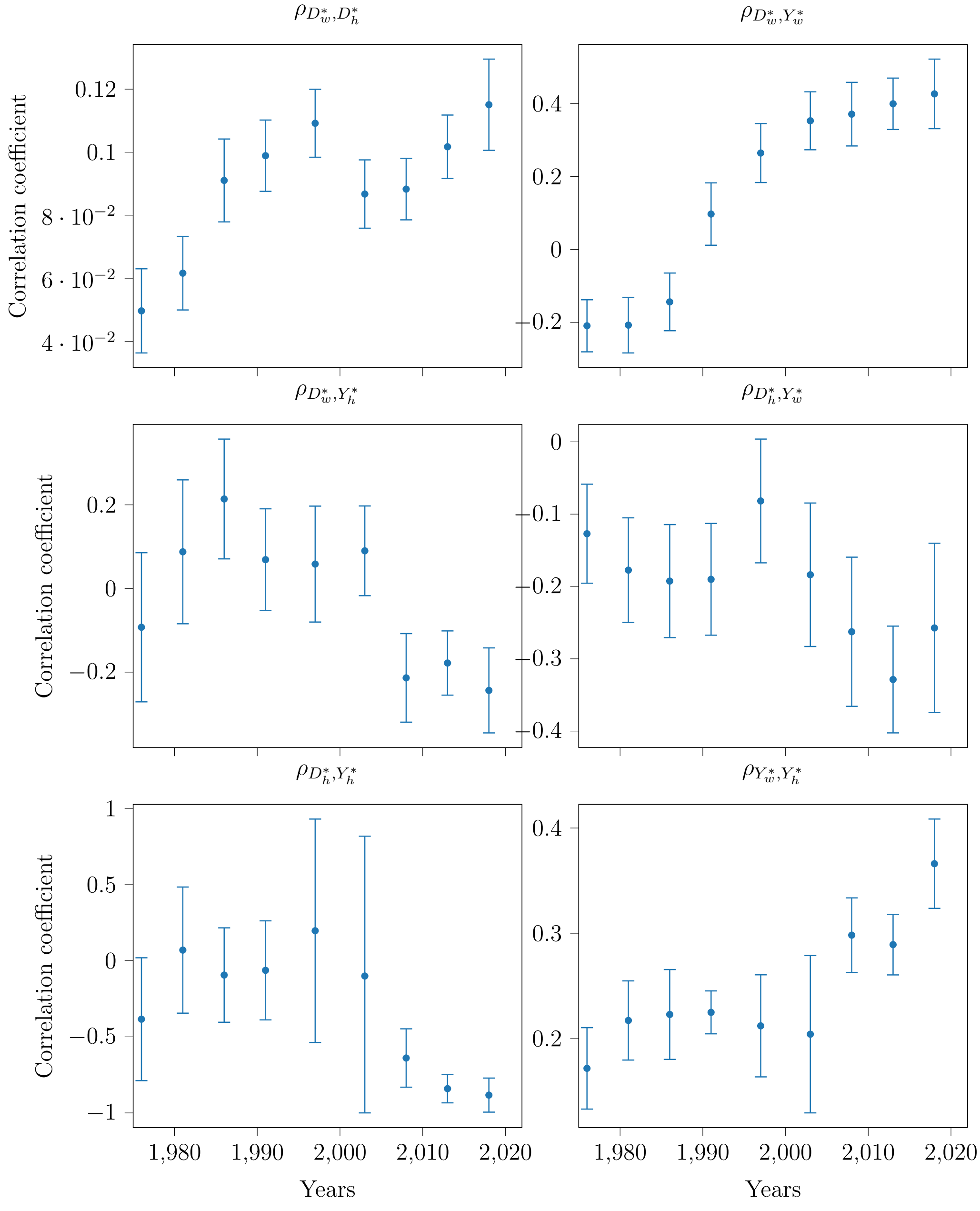


Figure 9: Results for covariance matrix at Q3.



The estimates of the parameters related to the husband's selection decision warrant further discussion. As male participation is typically treated as exogenous there is little empirical work on the impact of male selection on wages. Moreover, as methods to analyze the role of selection at different quantiles have only recently been developed there is little existing evidence on how it varies across the wage distribution. However, the parameters are not precisely estimated in this setting and in some instances may be unreasonable in magnitude. This is likely to reflect the form of identification employed. Alternatively, husbands' work decisions may be exogenous. Given the restricted version of the model is a special case of the richer model we decided to proceed with the results with this full model but we also estimated the model treating the husbands' work decisions as exogenous and conducted the corresponding counterfactuals. We comment on those results below although to anticipate the main finding, the treatment of husbands' work decisions as exogenous did not alter the substantive results. We report the estimates of $\rho_{D_w^*,Y_w^*}$, $\rho_{D_w^*,Y_h^*}$ and $\rho_{D_h^*,Y_w^*}$ from this restricted model in Figures 10-12.

7.2 Counterfactual Sorting Patterns

The estimated model's capacity to reproduce the frequencies in Tables 1 and 2 is reflected in Tables B21 and B22 and, although the large number of cells makes it difficult to draw a conclusion by a visual inspection, the estimated cells appear close to the true values. We now examine the role of various model parameters in generating the observed data by changing selected model parameters and examining the predicted allocation across cells. We compare these counterfactuals to Tables B21 and B22 as they represent the predictions of our estimated model.

We begin by setting $\rho_{D_w^*,D_h^*}$ to zero. Table 5 presents the results for 1976-1980 and Table 9 those for 2018-2022. They appear similar to Table B21 and Table B22 respectively and we conclude that this parameter is unimportant for determining the joint wage distribution.

Tables 6 and 10 present the results when $\rho_{D_w^*,Y_w^*}$, $\rho_{D_h^*,Y_h^*}$, $\rho_{D_w^*,Y_h^*}$ and $\rho_{D_h^*,Y_w^*}$ are also set to zero. We characterize these parameters as collectively capturing the selection process. For the earlier period the d10/d10 cell increases from 2.29 to

Figure 10: Results of covariance matrix at Q1 (exogenous selection of husbands).

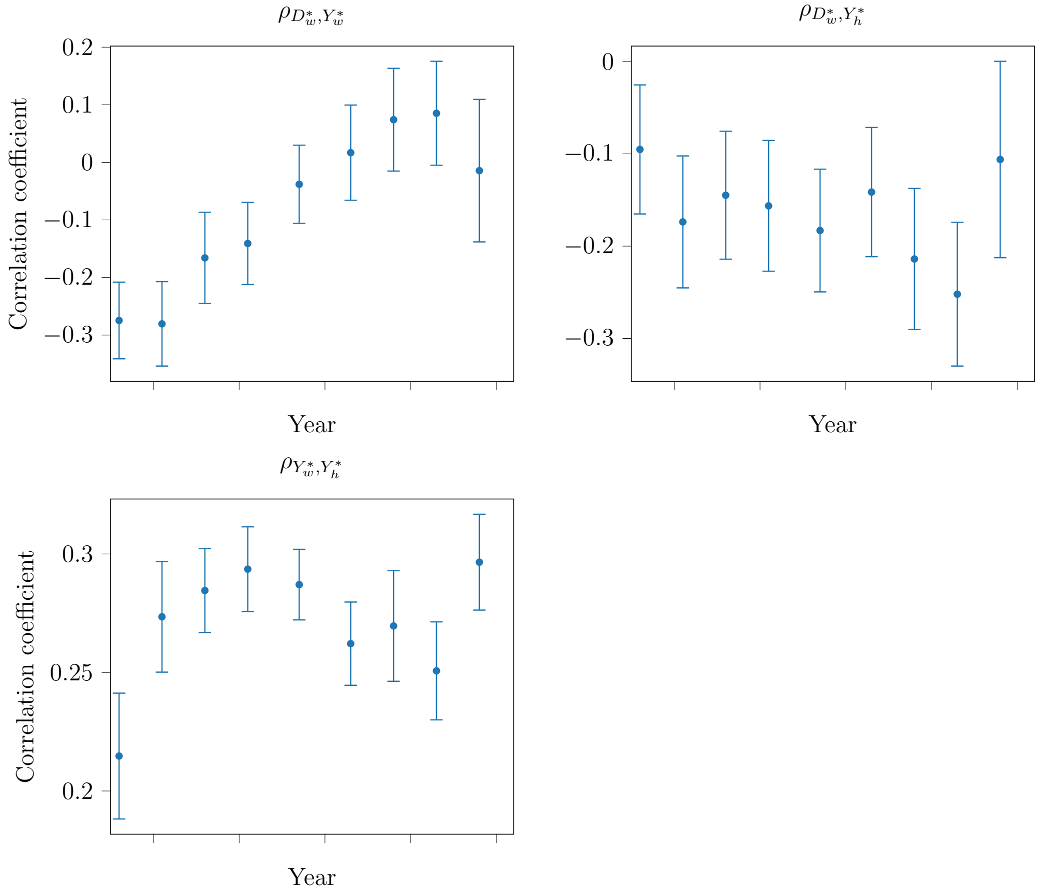


Figure 11: Results of covariance matrix at Q2 (exogenous selection of husbands).

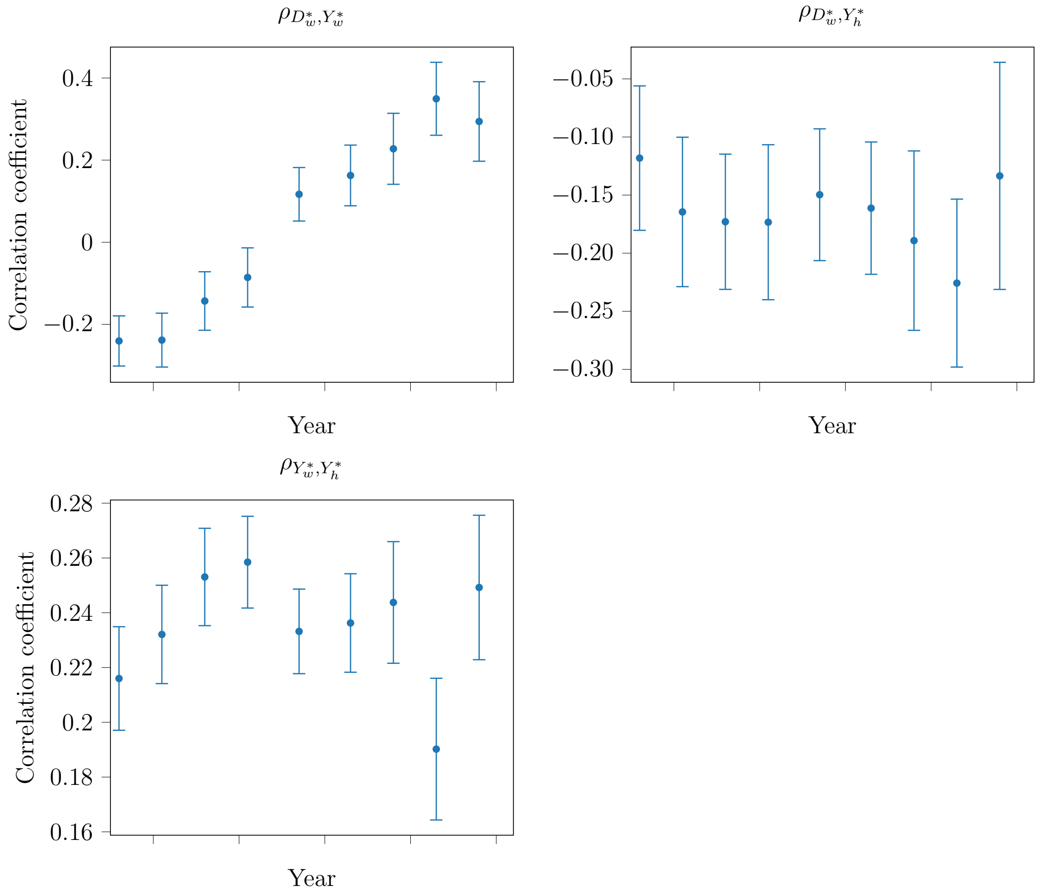
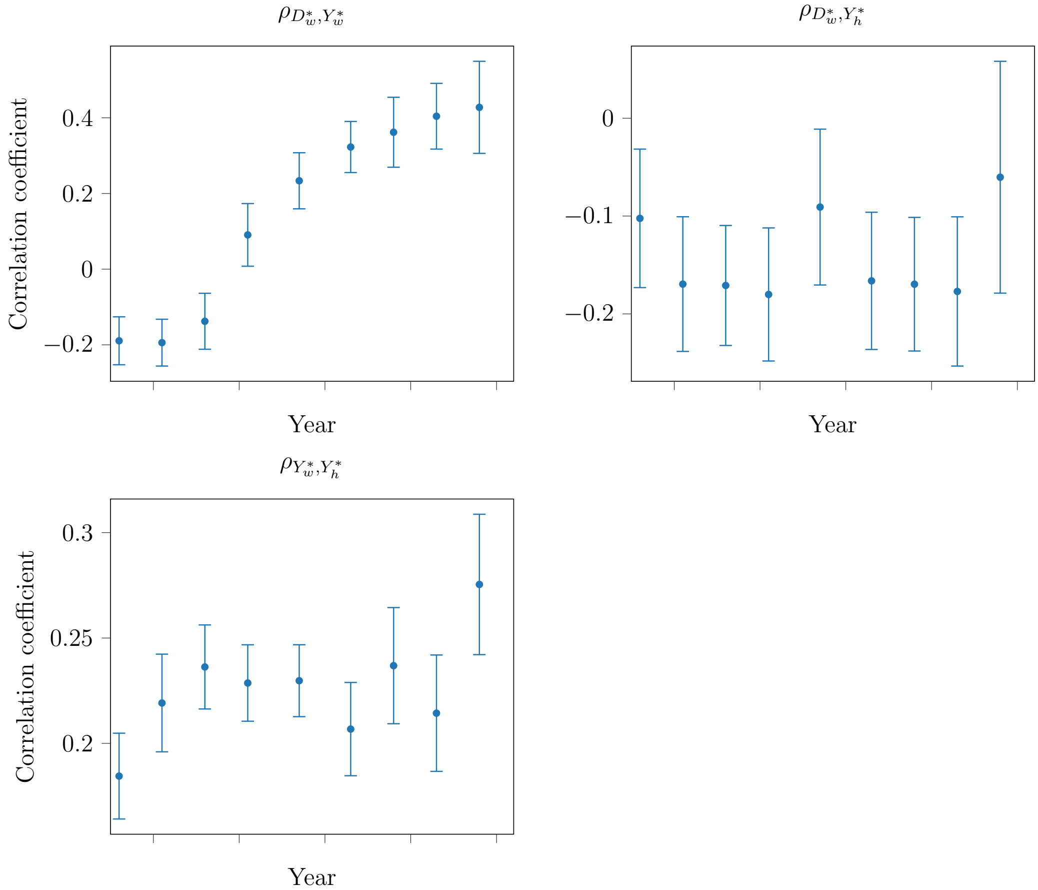


Figure 12: Results of covariance matrix at Q3 (exogenous selection of husbands).



		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.135**	1.37	1.14	1.006	0.848	0.939	0.757	0.691	0.728	0.386
	d2	1.623	**1.43**	1.2	1.034	0.974	0.845	0.758	0.735	0.724	0.678
	d3	1.368	1.184	**1.122**	1.067	1.092	0.898	0.842	0.8	0.737	0.89
	d4	1.154	1.344	1.193	**1.111**	0.925	0.976	0.87	0.897	0.756	0.774
	d5	0.955	1.185	1.225	1.178	**1.105**	1.012	0.982	0.969	0.936	0.452
	d6	0.756	1.042	1.086	1.147	1.108	**1.151**	1.007	1.005	0.946	0.751
	d7	0.597	0.776	1.074	1.033	1.078	1.131	**1.031**	1.096	1.096	1.088
	d8	0.601	0.603	0.832	1.131	1.093	1.132	1.135	**1.13**	1.197	1.144
	d9	0.427	0.597	0.676	0.756	1.136	1.209	1.245	1.239	**1.167**	1.546
	d10	0.381	0.544	0.374	0.535	0.655	0.696	1.373	1.434	1.714	**2.294**

Table 5: Results of estimated sorting measures in 1976-1980 at different quantiles when $\rho_{D^*_w,D^*_h} = 0$.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.189**	1.235	0.987	0.848	0.986	0.705	0.935	0.675	0.646	0.795
	d2	1.752	**1.525**	1.067	1.087	0.813	0.865	0.844	0.755	0.684	0.61
	d3	1.272	1.354	**1.218**	1.1	0.932	1.081	0.809	0.912	0.699	0.623
	d4	1.068	1.316	1.286	**1.103**	1.024	0.916	1.041	0.885	0.837	0.524
	d5	0.859	1.226	1.272	1.137	**1.045**	0.995	1.022	0.915	0.909	0.619
	d6	0.648	1.041	1.241	1.117	1.156	**1.075**	1.029	1.032	0.912	0.749
	d7	0.658	0.625	1.207	1.16	1.13	1.17	**1.015**	1.187	1.121	0.726
	d8	0.484	0.707	0.764	1.2	1.182	1.169	1.153	**1.098**	1.237	1.005
	d9	0.539	0.573	0.512	0.722	1.129	1.248	1.16	1.215	**1.493**	1.41
	d10	0.523	0.403	0.442	0.53	0.6	0.766	1.01	1.249	1.538	**2.94**

Table 6: Results of estimated sorting measures in 1976-1980 at different quantiles when $\rho_{D^*_w,D^*_h} = 0, \rho_{D^*_w Y^*_w} = 0, \rho_{D^*_w,Y^*_h} = 0, \rho_{D^*_h,Y^*_w} = 0, \rho_{D^*_h,Y^*_h} = 0$.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.209**	1.217	0.94	0.826	0.981	0.731	0.956	0.601	0.733	0.806
	d2	1.722	**1.521**	1.155	1.019	0.924	0.844	0.837	0.723	0.631	0.623
	d3	1.388	1.327	**1.139**	1.185	0.842	1.065	0.844	0.762	0.915	0.533
	d4	1.058	1.407	1.31	**1.11**	0.982	0.903	1.017	0.934	0.7	0.58
	d5	0.822	1.237	1.282	1.112	**1.088**	1.056	0.98	1.007	0.792	0.624
	d6	0.673	1.024	1.266	1.118	1.161	**1.013**	1.117	1.02	0.893	0.716
	d7	0.641	0.594	1.182	1.181	1.095	1.155	**0.973**	1.363	1.02	0.796
	d8	0.452	0.722	0.769	1.188	1.214	1.202	1.095	**1.201**	1.191	0.967
	d9	0.571	0.511	0.496	0.75	1.118	1.236	1.173	1.368	**1.299**	1.477
	d10	0.476	0.435	0.458	0.519	0.579	0.785	1.014	1.022	1.818	**2.894**

Table 7: Results of estimated sorting measures in 1976-1980 at different quantiles when $\rho_{D^*_w,D^*_h} = 0, \rho_{D^*_w,Y^*_w} = 0, \rho_{D^*_w,Y^*_h} = 0, \rho_{D^*_h,Y^*_w} = 0, \rho_{D^*_h,Y^*_h} = 0$ and distribution of Z as in total population.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**1.353**	1.2	1.095	0.973	1.045	0.919	0.946	0.856	0.82	0.795
	d2	1.328	**1.255**	1.125	1.001	1.051	0.941	0.942	0.827	0.779	0.752
	d3	1.209	1.048	**1.02**	1.124	0.832	1.024	0.904	0.912	0.998	0.93
	d4	1.051	1.055	1.046	**1.04**	0.994	0.988	0.993	0.99	0.944	0.901
	d5	0.985	1.02	1.031	1.046	**1.011**	1.007	0.998	1.01	0.962	0.93
	d6	0.944	0.996	1.011	1.013	1.014	**1.012**	1.024	1.012	1.006	0.968
	d7	0.86	0.935	0.974	1.002	1.026	1.018	**1.03**	1.076	1.035	1.043
	d8	0.786	0.884	0.946	0.98	1.029	1.045	1.06	**1.102**	1.075	1.093
	d9	0.734	0.797	0.878	0.938	0.975	1.017	1.067	1.171	**1.182**	1.242
	d10	0.759	0.809	0.872	0.892	1.008	1.022	1.041	1.048	1.191	**1.357**

Table 8: Results of estimated sorting measures in 1976-1980 at different quantiles when all correlations are zero and distribution of Z as in total population.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.896**	1.573	1.044	0.891	0.886	0.82	0.576	0.646	0.523	0.146
	d2	1.636	**1.702**	1.254	1.02	0.884	0.818	0.632	0.655	0.502	0.897
	d3	1.161	1.721	**1.373**	1.123	1.004	0.872	0.842	0.608	0.588	0.708
	d4	0.85	1.295	1.604	**1.26**	1.078	1.023	0.954	0.808	0.7	0.428
	d5	0.779	0.998	1.269	1.336	**1.122**	1.13	1.054	1.042	0.797	0.472
	d6	0.663	0.766	0.924	1.302	1.413	**1.27**	1.144	1.071	0.897	0.549
	d7	0.549	0.761	0.964	1.062	1.234	1.386	**1.294**	1.254	1.211	0.285
	d8	0.454	0.465	0.609	0.635	0.812	1.125	1.445	**1.111**	1.25	2.093
	d9	0.472	0.575	0.604	0.761	0.748	1.031	1.274	1.575	**1.803**	1.156
	d10	0.537	0.144	0.354	0.609	0.816	0.526	0.789	1.23	1.73	**3.265**

Table 9: Results of estimated sorting measures in 2018-2022 at different quantiles when $\rho_{D^*_w,D^*_h} = 0$.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.757**	1.275	0.984	0.945	0.919	0.679	0.777	0.619	0.446	0.599
	d2	1.633	**1.959**	1.287	1.074	0.911	0.862	0.793	0.597	0.524	0.36
	d3	0.776	1.885	**1.385**	1.212	1.067	1.143	0.904	0.701	0.545	0.381
	d4	0.867	1.28	1.517	**1.061**	1.191	1.076	0.901	0.9	0.668	0.54
	d5	0.853	0.762	1.267	1.637	**1.198**	1.103	1.146	0.717	0.697	0.619
	d6	0.584	0.81	1.075	1.148	1.336	**1.181**	1.057	1.068	1.019	0.721
	d7	0.735	0.625	0.804	1.018	1.19	1.188	**1.238**	1.254	1.082	0.865
	d8	0.709	0.512	0.609	0.774	0.946	1.304	1.196	**1.41**	1.242	1.299
	d9	0.524	0.53	0.604	0.699	0.712	0.82	1.131	1.652	**1.813**	1.515
	d10	0.558	0.36	0.467	0.436	0.529	0.638	0.862	1.086	1.961	**3.104**

Table 10: Results of estimated sorting measures in 2018-2022 at different quantiles when $\rho_{D^*_w,D^*_h} = 0, \rho_{D^*_w Y^*_w} = 0, \rho_{D^*_w,Y^*_h} = 0, \rho_{D^*_h,Y^*_w} = 0, \rho_{D^*_h,Y^*_h} = 0$.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.88**	1.375	0.903	0.819	0.958	0.582	0.741	0.621	0.453	0.669
	d2	1.771	**1.803**	1.329	1.116	0.961	0.792	0.812	0.59	0.486	0.339
	d3	0.76	1.978	**1.486**	1.138	1.108	1.117	0.834	0.651	0.538	0.391
	d4	0.767	1.195	1.65	**1.104**	1.184	1.095	1.04	0.743	0.659	0.564
	d5	0.762	0.847	1.133	1.609	**1.273**	1.101	1.068	0.905	0.73	0.572
	d6	0.521	0.766	1.075	1.197	1.328	**1.217**	1.048	1.074	1.017	0.758
	d7	0.793	0.571	0.792	1.0	1.115	1.399	**1.184**	1.236	1.038	0.873
	d8	0.579	0.536	0.583	0.805	0.949	1.201	1.324	**1.44**	1.305	1.278
	d9	0.523	0.546	0.611	0.637	0.744	0.849	1.122	1.668	**1.813**	1.488
	d10	0.638	0.38	0.445	0.494	0.457	0.651	0.831	1.076	1.95	**3.077**

Table 11: Results of estimated sorting measures in 2018-2022 at different quantiles when $\rho_{D^*_w,D^*_h} = 0, \rho_{D^*_w,Y^*_w} = 0, \rho_{D^*_w,Y^*_h} = 0, \rho_{D^*_h,Y^*_w} = 0, \rho_{D^*_h,Y^*_h} = 0$ and distribution of Z as in total population.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**1.542**	1.265	1.058	0.96	0.945	0.879	0.841	0.817	0.8	0.894
	d2	1.215	**1.204**	1.108	1.056	1.029	0.966	0.908	0.896	0.839	0.78
	d3	1.089	1.153	**1.12**	1.072	1.068	1.0	0.942	0.893	0.867	0.794
	d4	1.031	1.077	1.096	**1.066**	1.064	1.012	0.977	0.941	0.901	0.836
	d5	0.908	0.984	1.042	1.05	**1.065**	1.045	1.023	1.005	0.972	0.907
	d6	0.859	0.926	0.991	1.004	1.039	**1.041**	1.061	1.049	1.035	0.995
	d7	0.825	0.881	0.95	0.992	1.014	1.051	**1.067**	1.081	1.075	1.064
	d8	0.803	0.864	0.939	0.975	1.014	1.051	1.094	**1.103**	1.096	1.062
	d9	0.797	0.821	0.88	0.909	0.954	1.007	1.073	1.13	**1.186**	1.242
	d10	0.926	0.821	0.822	0.836	0.887	0.953	1.016	1.088	1.218	**1.433**

Table 12: Results of estimated sorting measures in 2018-2022 at different quantiles when all correlations are zero and distribution of Z as in total population.

2.94 while the sum in the (d9+d10)/(d9+d10) cells increases from 6.7 to 7.4. This suggests that negative selection bias operating in both labor markets is reducing household inequality. Negative selection is removing the relatively higher paid males and females from FTFY employment. Setting these parameters to zero inserts more higher paid workers into the market. This results in a higher propensity of the higher paid males and females to marry. This is an interesting result although it may reflect the identification strategy. Interestingly the frequencies in Table 10 provide a somewhat comparable story although the changes are less drastic. Setting the other parameters to zero appears to have little affect on the lower cells. However, while the d10/d10 cell decreases marginally the sum in the (d9+d10)/(d9+d10) cells increases from 7.9 to 8.4. This is similar to the earlier period although the estimated positive selection effects for females at upper quantiles is offsetting the negative selection effects for males throughout the period.

We also examine the counterfactual in which the above parameters are set to zero and we use the Z's of the whole sample rather than only those who are working. The results are in Tables 7 and 11 and a comparison with Tables 6 and 10 indicates they do not affect the observed sorting patterns. Although previous studies (see, for example, Chernozhukov, Fernández-Val, and Luo 2019) have analyzed the impact of selection on the distribution of wages and have found it to have some impact, our results indicate it would not change the sorting patterns even if it is changing the distribution of both, or either of, the male and female wages.

Finally we set $\rho_{Y_w^*,Y_h^*}$ to zero. For the BDR model the corresponding counterfactual produced a substantial reduction in sorting. Table 8 reports the results for the earlier time period. Given the large number of cells it is useful to focus on the extreme cases as these are the outcomes more closely associated with inequality. First consider the d1/d1 and d10/d10 cells. The former decreases from 2.21 to 1.35 while the latter decreases even more dramatically from 2.89 to 1.36. Expanding the 4 lower and upper combinations to include d2 and d9 respectively produces reductions from 6.6 and 7.4 to 5.1 and 4.96 respectively. It is interesting that the more substantial reductions appear to occur at the cells for the higher wage deciles. Turning to the 2018-2022 time period we see a similar pattern. The d1/d1 cell decreases from 2.88 to 1.54 and d10/d10 decreased from 3.07 to 1.43. Expanding to the four lowest and highest we see reductions of 6.82 to 5.21 and 8.3 to 5.07 respectively. This clearly suggests that the correlation in the unexplained components of wages is largely driving the observed sorting patterns. That is, sorting on unobservables is an important factor in driving inequality.

To capture the sorting behavior which includes the off diagonals we compute the Kendall rank correlation coefficient for these counterfactuals. These are reported in Table 13 and confirm the patterns described above. For the 1976-1980 period neither the selection parameters nor the use of the working or total population composition of Z's affect the rank correlation coefficient. For each of these experiments the

Table 13: Kendall rank correlation coefficient.

	1976-1980	2018-2022
Data	0.1833	0.2758
Estimated	0.1885	0.2588
Participation	0.1877	0.2582
Selection	0.1853	0.2421
X as in whole population	0.1869	0.2483
$\rho_{Y_w^*,Y_h^*} = 0$	0.0743	0.0768

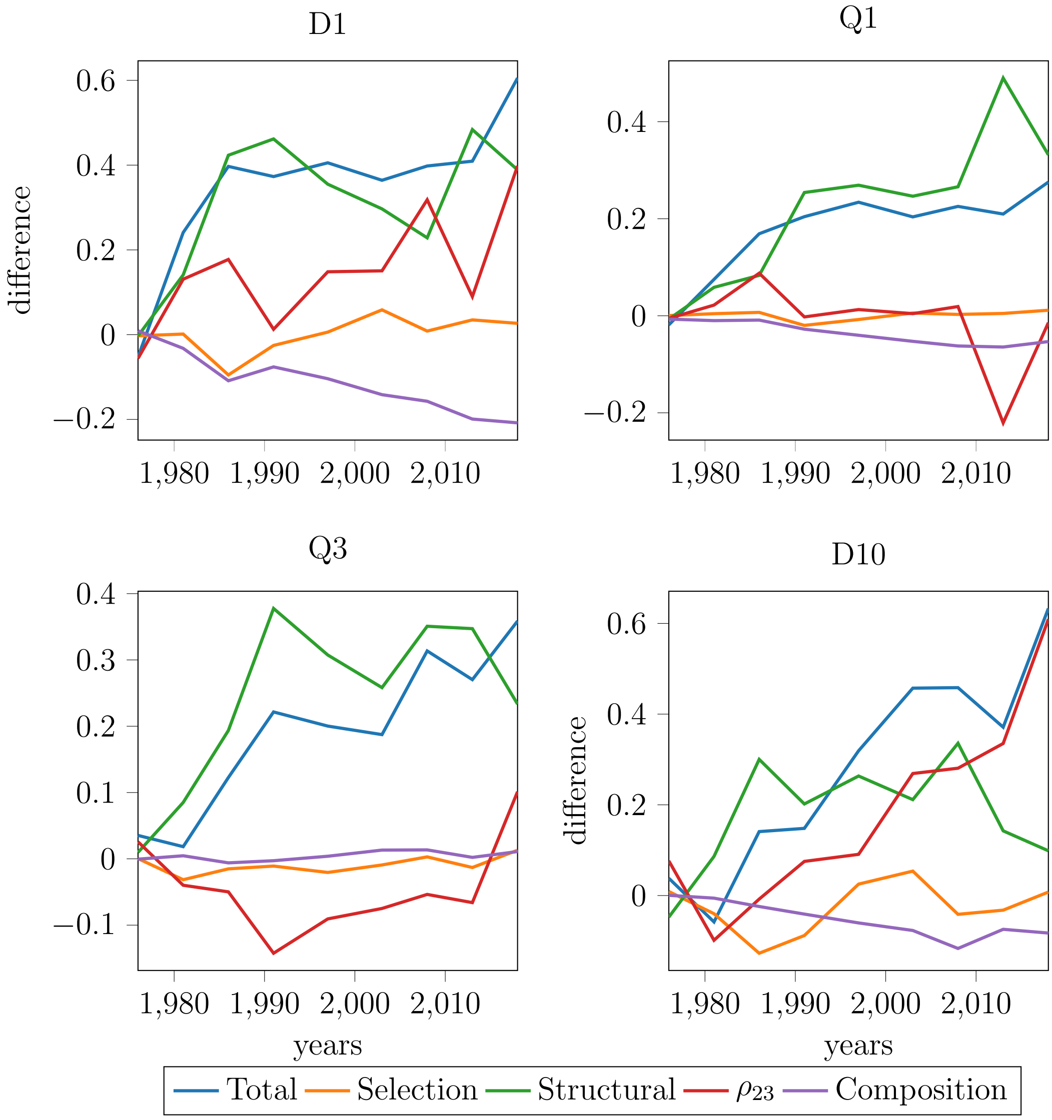
rank correlation coefficient does not differ greatly from its value for the data of .18. However, when we additionally set the value of $\rho_{Y_w^*,Y_h^*}$ to zero the value falls to .07. For the 2018-2022 period the results are similar despite the rank correlation coefficient of .27 being 50% higher than the 1976-1980 value. For this period the only counterfactual which generates a different pattern of sorting is that corresponding to $\rho_{Y_w^*,Y_h^*} = 0$. For this latter period the reduced rank correlation coefficient is also .07. This is consistent with Tables 8 and 12.

We noted that we estimated the model treating the male employment decision as exogenous. Although we do not report the results here, we conducted the corresponding counterfactuals to those for the full model. The general flavor of the simulations were similar although there was stronger evidence of selection affecting the observed sorting for the earlier period. However, the result that setting $\rho_{Y_w^*,Y_h^*}$ to zero reduced the observed level of positive sorting was also clearly supported.

7.3 Decomposing Changes in the Sorting Patterns

We now focus on the changes in the sorting patterns in the earlier tables and examine how the probability of both spouses being in a specific quantile of their wage distribution changes over our sample period. We decompose the total change into structural, selection and composition effects. Note that unlike the previous section, we now also examine the impact of the changes in the β's over time. As we employ a multivariate normal approximation for each point of the four dimensional partition of the data we can use the model parameter estimates to decompose the observed changes into the various components. While the composition effects are defined as the changes in the conditioning variables, the allocation of the other parameters

Figure 13: Decompositions of changes in the sorting measures



are less straightforward. The changes in the β's can be assigned the interpretation of conventional structural effects. We assign $\rho_{D_w^*,Y_w^*}$, $\rho_{D_h^*,Y_h^*}$, $\rho_{D_w^*,Y_h^*}$ and $\rho_{D_h^*,Y_w^*}$ to the selection effects. We isolate the impact of $\rho_{Y_w^*,Y_h^*}$ parameter as it captures the correlation between the unobservables which impact the spouses' wages. While this is a type of structural effect we separate it from the structural effects operating through the β's. Although we can evaluate the sorting behavior observed in 100 cells it appears that the interesting behavior occurs in the top and bottom cells. Accordingly we conduct the decompositions for both members of the couple being located in the bottom and top deciles and quartiles. These plots are presented in Figure 13. The decompositions represent how the change in the measures can be explained by changes in the various components. If a specific component appears to be close to zero then this indicates that its contribution has remained unchanged and not that its contribution is zero.

We start with the probability of both spouses being located in the bottom decile and quartile. Over the 47 years of our sample the former increases by almost 60 percent and the latter by almost 40 percent. For d1 the large positive change is driven by changes in the structural effect and $\rho_{Y_w^*,Y_h^*}$. Changes in the selection effect are unimportant and the changes in the composition effect are negative. For Q1 the results are essentially the same although the changes in the contribution of $\rho_{Y_w^*,Y_h^*}$ are generally unimportant except for a decrease towards the end of the sample. At the top of the distribution there is a similar finding. The probability of both in the top decile and quartile increases by around 60% and 35% respectively. The change in the probability of both spouses being in d10 is again driven by changes in the structural effect and $\rho_{Y_w^*,Y_h^*}$ although the change in the structural effect diminishes towards the end of the sample. The change in the probability of both spouses being in Q3 is due to changes in the structural effects. Changes in the composition and selection effects appear unimportant for either d10 or Q3. The results on the composition effects are consistent with the evidence as noted above in Breen and Salazar (2011), Gihleb and Lang (2018), Eika, Mogstad, and Zafar (2019) and Chiappori, Costa-Dias and Meghir (2020,2023). Education is a component of the composition effect and there is no impact at either Q1 or Q3. At D1 and D9 it is less clear as both probabilities decline.

The increasing presence of married females into the FTFY in the 1970's and 1980's had little impact on the sorting process in the marriage market. Overall the selection effects are small. Second, composition effects are generally unimportant and have reduced positive sorting. This reflects that the increased level of acquired education has increased everyone's probability of marrying a relatively highly educated spouse. This is consistent with the findings of others noted above regarding the impact of composition effects on positive sorting as measured by education. The clearest evidence is associated with the structural effects. These are the most important contributors to the total effect and their interpretation is also quite clear. Sorting behavior has not greatly changed in terms of the observed characteristics of the spouses. However, the higher skill premia, and the heterogeneity of the skill premia, have resulted in the highly (lowly) paid being even more highly (lowly) paid. This is consistent with empirical evidence on decompositions of changes in wage inequality. The evidence on the role of the $\rho_{Y_w^*,Y_h^*}$ provides a similar story. While this parameter captures the impact of unobservables it also reflects how they are priced. Our evidence suggests that the increased observed level of positive assortative sorting reflects the increases in prices of observables and unobservables. Individuals appear to marry the "same" spouses but the value of their characteristics has changed due to the increased skill premia. It is also possible that $\rho_{Y_w^*,Y_h^*}$ partially captures how an individual's wage is influenced by the value of their spouse's characteristics. For example, individuals with highly successful spouses may benefit from exposure to their spouse's network and this may produce some relationship between their wage and, for example, their spouse's education level. Some preliminary investigations of the data suggest that this relationship exists, and that it changes over time, but we do not pursue it here.

7.4 Household Income Inequality

We now explore the implication of the model's estimates for household income equality. We assume that all individuals comprising the FTFY couples are working the same number of hours and we employ the household wage, defined as the sum of the

Table 14: Decomposition of changes in inequality.

	1976-1980	2018-2022
Data	1.7987	2.3198
Estimated	1.8002	2.2797
Participation	1.7997	2.2841
Selection	2.0661	2.2776
Z as in whole population	2.066	2.2692
$\rho_{Y_w^*,Y_h^*} = 0$	1.735	2.1192
Random sorting	1.6812	2.0505

spouses' hourly wages, as our measure of family income. We examine how inequality has evolved by examining the D8/D2 ratio under different counterfactuals. We focus on this ratio as the manner in which its components have changed is clear from the tables above. The results are shown in Table 14. Note that we do not isolate the role of structural effects operating through the β's nor the composition effects reflecting the changes in the X's over time. As listed at the start of our Introduction, there is a large existing literature which has clearly established the impact of these factors on wage inequality. We repeat the same exercise as in the counterfactual sorting exercises of Section 7.2 in which we incrementally set the parameters capturing different features of the model to zero. We also explore the impact of employing the population X's rather than those of the working sample.

The first row of Table 14 presents the D8/D2 ratio for our two extreme sample periods. There is a large increase of 29.6%, from 1.79 to 2.32, noting that it is similar to the growth in comparable measures of inequality presented in the discussion of the data. The second row presents the comparable estimates from the predictions of our model and indicates that the level of inequality observed in the data are maintained. To examine the role of sorting we first estimate the corresponding measure in a setting of unconditional random sorting in which husbands are assigned to wives in a random manner. This is shown in the table's bottom row. The results represent the D8/D2 ratio based on 100,000 draws. The estimates are 1.68 and 2.05 and reflect an increase of 22% over the sample period. This large impact reflects the large increase in wage inequality which occurs for both husbands and wives. More interestingly, comparing the top and bottom rows of this table indicates that sorting increases inequality by 7.0% in the earlier period and 13.1% in the latter.

This suggests that sorting has substantially increased household inequality.

To investigate the role of the various model parameters, rows 3 to 6 repeat the counterfactuals conducted above. As with the tables above, each of the rows reflects the incremental change in the inequality measure as the parameters associated with some feature of the model are set to zero. For the earlier period selection has reduced inequality although we have highlighted our concerns regarding this result as it is reliant on the form of identification. The results regarding the use of the population Z's rather than the working sample Z's confirms our earlier results that this is does not appear to affect the results regarding sorting. Row 6 also highlights the recurring result that positive sorting is operating through the correlation of the spouses' unobservables. For the 2018-2022 period we have similar findings although that related to selection does not hold. However the importance of sorting on unobservables for inequality is again highlighted. Setting $\rho_{Y_w^*,Y_h^*} = 0$ decreases the D8/D2 ratio for the earlier period by 16% from 2.06 to 1.73 and for the later period by 7% from 2.26 to 2.12.

While there is an important role for $\rho_{Y_w^*,Y_h^*} = 0$, the remaining large increase in inequality in the household wage appears due to changes in composition and structural effects. However, the existing evidence, here and in the existing literature, suggests the sorting on observables has not changed for this time period and composition effects are unimportant. This suggests that the increase in wage inequality across households is due to structural effects capturing changes in the skill premia.

8 Conclusion

We examine the role of marital sorting on household income inequality in a period of substantial and increasing wage inequality. We do so by examining a sample of married couples from the CPS for the years 1976 to 2022 in which each of the spouses works full time/full year. To investigate the determinants of marital sorting and its impact on inequality we estimate a model explaining the spouse's employment decisions and their location in the gender specific wage distributions. This is done via a multivariate distribution regression approach which incorporates selection. We provide a number of important empirical findings.

First, we confirm that wage inequality has increased for both spouses in dual FTFY couples. Changes in inequality for these groups do not precisely correspond to those for all males and females but the trends are similar.

Second, we find clear evidence of positive sorting defined as higher incidences of high (low) wage males marrying high (low) wage females than expected under random sorting. However, we acknowledge that random sorting is not a realistic alternative to positive sorting as couples are frequently similar in terms of age, race and location of residence and each of these are known to determine wages. However, the disproportionate fraction of males in the extremes of the wage distribution marrying females in the corresponding extremes of their distribution is supportive of positive sorting. Moreover, this positive sorting has increased over our time period. Combining this increasing sorting with increasing spouse specific inequality has increased inequality across couples.

Third, a series of counterfactual experiments do not provide any evidence that unobservables related to either of the spouses' work decisions have any implications for the observed sorting behavior. However, there is compelling evidence that the correlation between the unobservables influencing the spouses wages is substantially contributing to the observed sorting behavior. While some of this correlation may be due to factors, such as cost of living premia, which are incurred by both members of the couple it is also possible that this captures factors corresponding to unobserved ability.

Finally, we find that positive sorting associated with the increased share of extreme cells has increased over the 47 years of our sample. A decomposition of the changes in these cells shares over the period examined reveals that their growth is almost entirely due to the prices of observable and unobservable characteristics. This suggests that the increase in observed sorting pattern in not due to different behavior nor the increase of more females in the labor market. Rather as the skill premia has increased this has increased the wages of both members of some couples while couples without these skills appear to be both negatively affected.

References

Autor, D.H., L.F. Katz , and M. Kearney (2008), "Trends in U.S. wage inequality: Revising the revisionists", *Review of Economics and Statistics* **90** 300–23.

Autor, D.H., A. Manning, and C.L. Smith (2016), "The contribution of the minimum wage to US wage inequality over three decades: A reassessment." *American Economic Journal: Applied Economics* **8**, 58–99.

Blau, F. and L.M. Kahn (2009), "Inequality and earnings distribution", In: W. Salverda, B. Nolan, and T.M. Smeeding, "Oxford Handbook on Economic Inequality", Oxford University Press, 177–203.

Blau, F.D., L.M. Kahn, N. Boboshko, and M.L. Comey (2021), "The impact of selection into the labor force on the gender wage gap", working paper, Cornell University.

Breen, R. and L. Salazar (2011), "Educational assortative mating and earnings inequality in the United States", *American Journal of Sociology* **117**, 808-43.

Cancian, M. and D. Reed (1998a), "Assessing the effects of wives' earnings on family income inequality", *The Review of Economics and Statistics* **80**, 73–79.

Cancian, M. and D. Reed (1998b), "The impact of wives' earnings on income inequality: issues and estimates", *Demography* **36**, 173–84.

Chiappori, P.A., B. Salanié and Y.Weiss, (2017), "Partner choice, investment in children, and the marital college premium", *American Economic Review* **107**, 2109–67.

Chiappori, P.A., Costa-Dias, M. and C.Meghir, (2019), "The marriage market, labor supply, and education choice", *Journal of Political Economy* **126**, s26-s72.

Chiappori, P.A., Costa-Dias, M., Crossman, S. and C.Meghir (2020), "Changes in assortative matching and inequality in income: Evidence for the UK", *Fiscal Studies* **41**, 39–63.

Chiappori, P.A., Costa-Dias, M. and C.Meghir, (2020). "Changes in assortative matching: Theory and evidence for the US", NBER Working Paper 26932.

Chiappori, P.A., Costa-Dias, M. and C.Meghir, (2023), "The measuring of assortativeness in marriage", working paper, Yale University.

Chernozhukov, V., I. Fernández-Val, and S. Luo (2019), "Distribution regression with sample selection, with an application to wage decompositions in the UK", working paper, MIT, Cambridge (MA).

De Rock, B., M. Kolvaleva, and T. Potoms (2023), "A spouse and a house are all we need? Housing demand, labor supply and divorce over the lifecycle", working paper, Free University of Brussels.

Eika, L., M. Mogstad, and B. Zafar (2019), "Educational assortative mating and household income Inequality", *Journal of Political Economy* **127**, 2795–2835.

Fernández-Val I., A. van Vuuren, and F. Vella (2018), "Decomposing Real Wage Changes in the United States", *IZA Discussion paper* **12044**, Bonn.

Fernández-Val I., A. van Vuuren, F. Vella and F. Peracchi (2023a), "Hours worked and the U.S distribution of real annual earnings 1976–2019", *Journal of Applied Econometrics*, forthcoming.

Fernández-Val I., A. van Vuuren, F. Vella and F. Peracchi (2023b) ""Selection and the distribution of female real hourly wages in the US"", *Quantitative Economics* **14**, 571-607.

Fernández-Val I., J. Meier, A. van Vuuren and F. Vella (2023) "A bivariate distribution regression model for intergenerational mobility", working paper, Boston University.

Fernández, R. and R. Rogerson (2001), "Sorting and long-run inequality", *Quarterly Journal of Economics* **116**, 1305–41.

Flood, S., M. King, S. Ruggles, and J.R. Warren (2015), "Integrated public use microdata series, Current Population Survey: Version 4.0 [Machine-readable database]", working paper, University of Minnesota.

Geweke, J. (1991), "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints", *Computing Science and Statistics* **23**, 571–78.

Gihleb, R. and K. Lang (2020), "Educational homogamy and assortative mating have not increased", *Research in Labor Economics* **48**, 1–26.

Greenwood, J., N. Guner, G. Kocharkov and C. Santos (2014), "Marry your like: assortative mating and income inequality", *American Economic Review* **104**, 348–54.

Hajivassiliou, V. (1990) "Smooth simulation estimation of panel data LDV models", working paper, Yale University.

Heckman J.J. (1974), "Shadow prices, market wages and labor supply", *Econometrica* **42**, 679–94.

Heckman J.J. (1979), "Sample selection bias as a specification error", *Econometrica* **47**, 153–61.

Juhn C., K.M. Murphy, and B. Pierce (1993), "Wage inequality and the rise in returns to skill", *Journal of Political Economy* **101**, 410–42.

Katz L. F., and K. Murphy (1992), "Changes in relative wages, 1963–1987: Supply and demand factors" *Quarterly Journal of Economics* **107**, 35–78.

Keane, M. (1994), "A computationally practical simulation estimator for panel data", *Econometrica* **62**, 95-116.

Kremer, M. (1997), "How much does sorting increase inequality?", *Quarterly Journal of Economics* **112**, 115–39.

Lise, J. and S. Seitz (2011), "Consumption Inequality and Intra-household Allocations", *Review of Economic Studies* **78**, 328–55.

Lise, J. and K. Yamada (2018),"Household sharing and commitment: Evidence from panel data on individual expenditures and time use", *Review of Economic Studies* **86**, 2184-2219.

Mulligan, C. and Y. Rubinstein (2008), "Selection, investment, and women's relative wages over time", *Quarterly Journal of Economics* **123**, 1061–1110.

Murphy, K.M. and F. Welch (1992), "The structure of wages", *Quarterly Journal of Economics* **107**, 285–326.

Sibuya, M. (1959), "Bivariate extreme statistics, I," *Annals of the Institute of Statistical Mathematics* **11**, 195–210.

Welch F. (2000), "Growth in women's relative wages and inequality among men: One phenomenon or two?" *American Economic Review Papers & Proceedings* **90**, 444–49.

A Proof of Theorem 1

We start with a useful result about the properties of the multivariate normal CDF.

Lemma 1 *Let $\Phi_N(\cdot;\boldsymbol{\Sigma})$ denote the CDF of the multivariate standard normal distribution of dimension $N \geq 3$ with nonsingular correlation matrix $\boldsymbol{\Sigma}$ and ρ_{12} denote the $(1,2)$-element of $\boldsymbol{\Sigma}$. Then,*

$$\frac{\partial \Phi_N(\boldsymbol{x};\boldsymbol{\Sigma})}{\partial \rho_{12}} = \frac{\mathrm{E}[\phi_2(z_{1:3}, z_{2:3}; \rho_{12:3}) \mid \boldsymbol{X}_3 \leq \boldsymbol{x}_3]\Phi_{N-2}(\boldsymbol{x}_3;\boldsymbol{\Sigma}_{33})}{\sigma_{1:3}\sigma_{2:3}} > 0,$$

where $\boldsymbol{X} = (X_1, X_2, \boldsymbol{X}_3')'$ is a multivariate standard normal random variable with correlation matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \boldsymbol{\Sigma}_{13}' \\ \rho_{12} & 1 & \boldsymbol{\Sigma}_{23}' \\ \boldsymbol{\Sigma}_{13} & \boldsymbol{\Sigma}_{23} & \boldsymbol{\Sigma}_{33} \end{pmatrix},$$

$\boldsymbol{x} = (x_1, x_2, \boldsymbol{x}_3')'$, $\phi_2(\cdot;\rho)$ is the PDF of the bivariate standard normal distribution with correlation coefficient ρ, $z_{j:3} := (x_j - \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{j3}'\boldsymbol{X}_3)/\sigma_{j:3}$ and $\sigma_{j:3}^2 := 1 - \boldsymbol{\Sigma}_{j3}'\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{j3}$ for $j = 1, 2$, and

$$\rho_{12:3} := \frac{\sigma_{12:3}}{\sigma_{1:3}\sigma_{2:3}} := \frac{\rho_{12} - \boldsymbol{\Sigma}_{13}'\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{23}}{\sigma_{1:3}\sigma_{2:3}}.$$

Hence, $\rho_{12} \mapsto \Phi_N(\boldsymbol{x};\boldsymbol{\Sigma})$ is strictly increasing on $(-1, 1)$.

Proof. By the definition of conditional probability and iterated expectations

$$\begin{aligned}\Phi_N(\boldsymbol{x};\boldsymbol{\Sigma}) &= \mathrm{E}[\Pr(X_1 \leq x_1, X_2 \leq x_2 \mid \boldsymbol{X}_3) \mid \boldsymbol{X}_3 \leq \boldsymbol{x}_3] \Pr(\boldsymbol{X}_3 \leq \boldsymbol{x}_3) \\ &= \mathrm{E}[\Phi_2(z_{1:3}, z_{2:3}; \rho_{12:3}) \mid \boldsymbol{X}_3 \leq \boldsymbol{x}_3]\Phi_{N-2}(\boldsymbol{x}_3;\boldsymbol{\Sigma}_{33}), \quad (8)\end{aligned}$$

where the second equality uses that

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid \boldsymbol{X}_3 = \boldsymbol{x}_3 \sim \mathcal{N}_2\left(\begin{pmatrix} \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{13}'\boldsymbol{x}_3 \\ \boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{\Sigma}_{23}'\boldsymbol{x}_3 \end{pmatrix}, \begin{pmatrix} \sigma_{1:3}^2 & \sigma_{12:3} \\ \sigma_{12:3} & \sigma_{2:3}^2 \end{pmatrix}\right),$$

by the properties of the multivariate normal distribution. Note that the variance-covariance matrix is non-singular because $\boldsymbol{\Sigma}$ is non-singular.

The result then follows from noting that the RHS of (8) only depends on ρ_{12} through $\rho_{12:3}$ in $\Phi_2(z_{1:3}, z_{2:3}; \rho_{12:3})$, $\partial\Phi_2(\cdot;\rho)/\partial\rho = \phi_2(\cdot;\rho)$ (e.g., Sibuya, 1959), $\partial\rho_{12:3}/\partial\rho_{12} = 1/(\sigma_{1:3}\sigma_{2:3})$, and the chain rule. ■

We now proceed with the proof of Theorem 1. The argument is conditional on all the covariates Z in steps (1), (3) and (4), and the common covariates X in step (2) below. We consider the worst case where the covariates Z_1 that satisfy the exclusion restriction only include a binary variable. To lighten the notation, we drop the arguments of a function when they are 0. For example, $\mu_{D_w^*}(z) := \mu_{D_w^*}(0, z)$. The proof shows identification of all the parameters sequentially:

1. $\mu_{D_w^*}(z)$, $\mu_{D_h^*}(z)$ and $\rho_{D_w^*,D_h^*}(z)$: these parameters are identified from the distribution of (D_w, D_h) conditional on $Z = z$. In particular,

$$\mu_{D_j^*}(z) = \Phi^{-1}\left(\Pr(D_j = 1 \mid Z = z)\right), \quad j = w, h,$$

and $\rho_{D_w^*,D_h^*}(z)$ is identified as the solution in ρ to the equation

$$\Pr(D_w = 1, D_h = 1 \mid Z = z) = \Phi_2\left(\mu_{D_w^*}(z), \mu_{D_h^*}(z); \rho\right).$$

The solution exists and is unique by the same argument as in Lemma 1 of CFL.

2. $\mu_{Y_j^*}(y_j, x)$ and $\rho_{D_j^*,Y_j^*}(y_j, x)$, $j = w, h$: these parameters are identified by Theorem 1 of CFL using the relevance conditions and exclusion restrictions in Assumption 2(2)–(3).

3. $\rho_{D_w^*,Y_h^*}(y_h, z)$: this parameter is identified from the distribution of (D_w, D_h, Y_h) conditional on $Z = z$ as the solution in ρ_{12} to the equation

$$\Pr(Y_h \leq y_h, D_w = 1, D_h = 1 \mid Z = z) = \Phi_3(\mu_{Y_h^*}(y_h, x), \mu_{D_w^*}(z), \mu_{D_h^*}(z); \mathbf{\Sigma}(y_h, z)),$$

where

$$\mathbf{\Sigma}(y_h, z) = \begin{pmatrix} 1 & \rho_{12} & \rho_{D_h^*,Y_h^*}(y_h, x) \\ \rho_{12} & 1 & \rho_{D_w^*,D_h^*}(z) \\ \rho_{D_h^*,Y_h^*}(y_h, x) & \rho_{D_w^*,D_h^*}(z) & 1 \end{pmatrix}.$$

The solution exists by Assumption 2(1) and is unique by Lemma 1. A similar argument shows that $\rho_{D_h^*,Y_w^*}(y_w, x)$ is identified.

4. $\rho_{Y_w^*,Y_h^*}(y_w, y_h, z)$: this parameter is identified from the distribution of (D_w, D_h, Y_w, Y_h) conditional on $Z = z$ as the solution in ρ_{12} to the equation

$$\begin{aligned}\Pr(Y_w \leq y_w, Y_h \leq y_h, D_w = 1, D_h = 1 \mid Z = z)\\ = \Phi_4(\mu_{Y_w^*}(y_w, x), \mu_{Y_h^*}(y_h, x), \mu_{D_w^*}(z), \mu_{D_h^*}(z); \mathbf{\Sigma}(y_w, y_h, z)),\end{aligned}$$

where

$$\mathbf{\Sigma}(y_w, y_h, z) = \begin{pmatrix} 1 & \rho_{12} & \rho_{D_w^*,Y_w^*}(y_w,x) & \rho_{D_h^*,Y_w^*}(y_w,x) \\ \rho_{12} & 1 & \rho_{D_w^*,Y_h^*}(y_h,x) & \rho_{D_h^*,Y_h^*}(y_h,x) \\ \rho_{D_w^*,Y_w^*}(y_w,x) & \rho_{D_w^*,Y_h^*}(y_h,x) & 1 & \rho_{D_w^*,D_h^*}(z) \\ \rho_{D_h^*,Y_w^*}(y_w,x) & \rho_{D_h^*,Y_h^*}(y_h,x) & \rho_{D_w^*,D_h^*}(z) & 1 \end{pmatrix}.$$

The solution exists by Assumption 2(1) and is unique by Lemma 1.

■

B Tables

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**39.2**	38.2	39.4	38.9	37.5	39.0	40.6	40.0	41.3	42.8
	d2	39.9	**38.9**	37.9	38.6	39.3	39.6	40.8	41.0	41.4	43.3
	d3	38.8	38.2	**38.2**	37.9	38.9	39.1	40.1	41.2	41.2	42.1
	d4	38.5	37.8	38.3	**37.0**	38.3	39.8	40.1	39.5	41.9	42.1
	d5	38.2	36.9	36.9	37.4	**36.8**	38.5	39.0	38.7	40.4	42.7
	d6	37.5	36.4	37.8	37.2	36.6	**37.8**	38.5	39.8	40.3	42.3
	d7	36.9	39.0	36.1	37.2	37.1	38.4	**40.2**	40.2	40.8	42.4
	d8	38.7	37.2	36.9	37.8	37.7	38.6	38.7	**39.7**	40.1	42.1
	d9	37.0	38.0	37.7	37.5	37.5	38.2	39.0	39.7	**40.2**	41.8
	d10	41.9	39.2	40.2	39.3	39.5	39.4	40.1	38.6	40.9	**41.2**

Table B1: Average age of the wives for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**42.3**	41.0	42.0	41.7	40.5	41.7	43.4	42.7	43.9	46.0
	d2	42.9	**41.7**	40.7	41.6	41.7	42.5	43.7	43.9	44.0	46.2
	d3	41.3	40.8	**40.6**	40.1	41.4	41.7	42.5	44.4	43.7	44.9
	d4	41.6	40.1	40.8	**39.8**	40.7	42.6	42.4	42.0	44.3	44.9
	d5	40.7	39.6	39.3	39.7	**39.4**	40.9	41.9	41.4	43.2	45.6
	d6	39.9	38.8	40.1	39.9	39.0	**40.3**	41.0	42.3	43.0	45.3
	d7	39.0	41.2	38.4	40.0	39.8	40.9	**42.9**	42.7	43.5	45.3
	d8	41.3	39.9	38.8	39.7	40.1	41.0	41.2	**42.2**	42.7	44.9
	d9	39.4	40.1	40.2	40.2	40.1	40.1	41.3	42.1	**42.9**	44.5
	d10	44.3	41.3	42.3	40.9	41.4	41.3	42.5	40.6	43.4	**44.4**

Table B2: Average age of the husbands for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**41.0**	40.6	41.1	41.6	41.0	44.5	42.5	43.9	44.3	47.1
	d2	40.6	**41.2**	41.0	41.7	41.7	41.0	43.0	43.4	44.5	44.5
	d3	40.5	40.2	**40.0**	40.1	41.0	41.6	42.7	44.4	44.1	46.6
	d4	40.3	40.9	40.3	**41.1**	41.6	42.1	41.8	43.1	44.0	44.0
	d5	41.2	39.6	40.6	40.3	**41.3**	41.2	42.8	42.0	44.4	45.5
	d6	40.9	40.8	39.3	41.1	41.4	**42.5**	42.0	42.9	42.6	43.9
	d7	41.3	41.0	40.7	41.7	42.3	42.5	**42.2**	41.8	41.5	44.5
	d8	43.0	43.5	43.2	42.2	42.6	42.6	42.9	**42.6**	43.4	44.4
	d9	43.9	44.7	43.4	42.9	43.3	43.5	42.7	42.6	**43.6**	44.1
	d10	43.1	43.7	41.3	43.2	44.3	43.5	42.0	43.6	43.5	**43.8**

Table B3: Average age of the wives for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**42.7**	42.9	43.6	43.2	43.4	46.8	44.4	46.2	46.6	50.4
	d2	42.1	**43.3**	42.7	43.6	43.8	43.1	44.8	45.8	47.1	47.7
	d3	42.3	42.2	**42.1**	42.6	43.0	43.4	44.7	46.8	46.4	48.7
	d4	42.4	42.9	42.4	**42.9**	43.4	43.6	43.7	45.0	46.0	46.0
	d5	43.2	41.0	42.2	42.3	**43.1**	43.4	44.7	44.2	46.4	47.8
	d6	42.3	42.6	41.0	42.2	43.3	**44.5**	43.8	44.7	44.4	45.5
	d7	43.4	42.2	42.2	43.2	44.1	44.3	**44.0**	43.8	43.9	46.5
	d8	44.9	44.8	44.5	43.7	44.1	43.9	44.6	**44.4**	45.1	46.6
	d9	45.7	45.9	44.8	44.3	45.0	45.3	43.9	44.5	**45.6**	46.0
	d10	45.1	44.8	42.7	44.7	45.2	44.9	43.4	45.2	45.0	**45.5**

Table B4: Average age of the husbands for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.064**	0.053	0.049	0.045	0.05	0.056	0.058	0.048	0.063	0.149
	d2	0.05	**0.04**	0.057	0.053	0.062	0.05	0.041	0.062	0.095	0.079
	d3	0.066	0.069	**0.073**	0.094	0.064	0.075	0.083	0.059	0.059	0.078
	d4	0.082	0.109	0.09	**0.115**	0.106	0.098	0.106	0.073	0.077	0.129
	d5	0.143	0.149	0.124	0.096	**0.151**	0.117	0.128	0.124	0.118	0.19
	d6	0.194	0.193	0.177	0.173	0.149	**0.169**	0.17	0.155	0.125	0.186
	d7	0.21	0.21	0.26	0.205	0.198	0.198	**0.221**	0.194	0.224	0.243
	d8	0.361	0.25	0.249	0.295	0.28	0.224	0.277	**0.241**	0.267	0.332
	d9	0.384	0.309	0.379	0.323	0.357	0.35	0.302	0.331	**0.346**	0.403
	d10	0.402	0.378	0.389	0.407	0.413	0.435	0.44	0.484	0.507	**0.552**

Table B5: Percentage of university education among wives for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.069**	0.084	0.093	0.096	0.136	0.12	0.124	0.135	0.149	0.385
	d2	0.095	**0.073**	0.102	0.103	0.118	0.139	0.104	0.113	0.202	0.36
	d3	0.089	0.092	**0.133**	0.148	0.12	0.154	0.166	0.153	0.189	0.352
	d4	0.12	0.115	0.108	**0.15**	0.19	0.173	0.152	0.206	0.172	0.397
	d5	0.173	0.157	0.188	0.179	**0.205**	0.199	0.159	0.244	0.266	0.435
	d6	0.177	0.21	0.216	0.207	0.171	**0.231**	0.262	0.259	0.269	0.434
	d7	0.205	0.191	0.255	0.207	0.201	0.242	**0.215**	0.261	0.32	0.446
	d8	0.273	0.211	0.242	0.263	0.255	0.251	0.29	**0.284**	0.316	0.526
	d9	0.281	0.275	0.23	0.28	0.299	0.328	0.281	0.379	**0.4**	0.547
	d10	0.283	0.296	0.285	0.304	0.283	0.389	0.348	0.473	0.496	**0.64**

Table B6: Percentage of university education among husbands for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.034**	0.032	0.034	0.025	0.039	0.033	0.027	0.035	0.041	0.133
	d2	0.038	**0.026**	0.032	0.031	0.04	0.028	0.026	0.027	0.071	0.067
	d3	0.04	0.043	**0.046**	0.067	0.037	0.048	0.062	0.045	0.043	0.073
	d4	0.046	0.066	0.052	**0.061**	0.081	0.062	0.069	0.063	0.061	0.116
	d5	0.1	0.096	0.093	0.064	**0.092**	0.08	0.075	0.083	0.085	0.15
	d6	0.121	0.115	0.124	0.11	0.093	**0.113**	0.128	0.108	0.084	0.134
	d7	0.117	0.107	0.168	0.121	0.124	0.13	**0.135**	0.13	0.172	0.191
	d8	0.216	0.127	0.152	0.187	0.169	0.137	0.161	**0.167**	0.186	0.258
	d9	0.205	0.174	0.17	0.189	0.227	0.236	0.204	0.24	**0.268**	0.325
	d10	0.197	0.207	0.229	0.222	0.209	0.286	0.275	0.365	0.408	**0.467**

Table B7: Percentage of households with both university education for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.115**	0.125	0.147	0.173	0.188	0.248	0.295	0.316	0.438	0.509
	d2	0.146	**0.177**	0.192	0.211	0.245	0.269	0.227	0.315	0.36	0.465
	d3	0.243	0.264	**0.253**	0.275	0.296	0.256	0.341	0.414	0.497	0.506
	d4	0.323	0.282	0.331	**0.338**	0.396	0.327	0.424	0.414	0.484	0.642
	d5	0.472	0.397	0.485	0.405	**0.451**	0.485	0.542	0.615	0.581	0.675
	d6	0.471	0.54	0.579	0.544	0.582	**0.555**	0.581	0.684	0.708	0.729
	d7	0.575	0.551	0.601	0.619	0.633	0.598	**0.683**	0.729	0.706	0.787
	d8	0.662	0.584	0.668	0.671	0.702	0.698	0.756	**0.781**	0.819	0.845
	d9	0.724	0.794	0.722	0.811	0.752	0.767	0.842	0.836	**0.863**	0.901
	d10	0.813	0.78	0.776	0.8	0.822	0.883	0.87	0.912	0.915	**0.938**

Table B8: Percentage of university education among wives for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.107**	0.107	0.101	0.194	0.233	0.224	0.312	0.416	0.562	0.681
	d2	0.111	**0.158**	0.164	0.135	0.205	0.25	0.27	0.38	0.524	0.728
	d3	0.123	0.175	**0.182**	0.209	0.262	0.263	0.348	0.49	0.513	0.654
	d4	0.196	0.169	0.226	**0.262**	0.339	0.309	0.392	0.402	0.543	0.739
	d5	0.224	0.256	0.275	0.275	**0.361**	0.414	0.452	0.577	0.604	0.801
	d6	0.195	0.254	0.281	0.341	0.418	**0.43**	0.432	0.573	0.688	0.763
	d7	0.269	0.298	0.297	0.344	0.436	0.47	**0.553**	0.592	0.662	0.825
	d8	0.364	0.283	0.323	0.428	0.443	0.467	0.589	**0.684**	0.739	0.832
	d9	0.366	0.406	0.417	0.405	0.459	0.506	0.657	0.737	**0.789**	0.886
	d10	0.449	0.39	0.414	0.482	0.562	0.599	0.658	0.749	0.856	**0.926**

Table B9: Percentage of university education among husbands for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.059**	0.047	0.057	0.087	0.101	0.138	0.168	0.247	0.384	0.457
	d2	0.062	**0.084**	0.082	0.081	0.134	0.149	0.161	0.218	0.305	0.447
	d3	0.082	0.127	**0.106**	0.136	0.17	0.125	0.204	0.3	0.38	0.429
	d4	0.148	0.113	0.154	**0.2**	0.244	0.194	0.271	0.277	0.352	0.58
	d5	0.175	0.181	0.225	0.207	**0.288**	0.307	0.347	0.45	0.464	0.602
	d6	0.152	0.214	0.257	0.308	0.342	**0.351**	0.336	0.485	0.595	0.64
	d7	0.231	0.24	0.265	0.302	0.379	0.411	**0.467**	0.528	0.558	0.685
	d8	0.305	0.231	0.276	0.358	0.392	0.416	0.546	**0.623**	0.68	0.758
	d9	0.325	0.37	0.4	0.383	0.431	0.462	0.615	0.691	**0.741**	0.822
	d10	0.43	0.366	0.405	0.453	0.533	0.577	0.63	0.714	0.824	**0.897**

Table B10: Percentage of university education among husbands and wives for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.222**	0.144	0.152	0.111	0.093	0.12	0.124	0.087	0.068	0.056
	d2	0.184	**0.159**	0.141	0.089	0.09	0.103	0.115	0.089	0.055	0.051
	d3	0.209	0.126	**0.092**	0.092	0.109	0.134	0.086	0.087	0.102	0.041
	d4	0.166	0.128	0.095	**0.12**	0.074	0.072	0.083	0.083	0.073	0.03
	d5	0.14	0.12	0.112	0.107	**0.081**	0.086	0.1	0.06	0.066	0.063
	d6	0.169	0.15	0.14	0.107	0.098	**0.089**	0.089	0.099	0.109	0.052
	d7	0.151	0.137	0.108	0.133	0.096	0.092	**0.112**	0.053	0.081	0.071
	d8	0.119	0.113	0.097	0.137	0.136	0.102	0.132	**0.095**	0.098	0.077
	d9	0.151	0.169	0.132	0.126	0.12	0.132	0.143	0.107	**0.102**	0.078
	d10	0.142	0.156	0.167	0.18	0.113	0.134	0.175	0.146	0.144	**0.104**

Table B11: Percentage of non-white wives 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.223**	0.139	0.155	0.118	0.097	0.123	0.124	0.087	0.063	0.046
	d2	0.185	**0.161**	0.141	0.084	0.093	0.1	0.115	0.086	0.059	0.045
	d3	0.197	0.122	**0.092**	0.084	0.112	0.13	0.079	0.083	0.098	0.032
	d4	0.163	0.126	0.095	**0.115**	0.058	0.072	0.079	0.083	0.073	0.026
	d5	0.15	0.131	0.105	0.099	**0.081**	0.086	0.1	0.067	0.069	0.071
	d6	0.153	0.159	0.143	0.113	0.104	**0.089**	0.089	0.087	0.116	0.059
	d7	0.166	0.149	0.108	0.138	0.107	0.095	**0.112**	0.056	0.081	0.068
	d8	0.134	0.113	0.108	0.134	0.136	0.097	0.14	**0.098**	0.088	0.071
	d9	0.144	0.198	0.14	0.122	0.13	0.134	0.141	0.1	**0.098**	0.073
	d10	0.15	0.163	0.167	0.186	0.109	0.131	0.165	0.146	0.144	**0.102**

Table B12: Percentage of non-white husbands 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.217**	0.137	0.147	0.105	0.09	0.12	0.116	0.087	0.059	0.046
	d2	0.175	**0.155**	0.141	0.084	0.084	0.093	0.112	0.082	0.047	0.045
	d3	0.197	0.116	**0.087**	0.073	0.107	0.127	0.079	0.083	0.091	0.027
	d4	0.158	0.124	0.088	**0.115**	0.055	0.068	0.079	0.08	0.069	0.026
	d5	0.133	0.117	0.1	0.091	**0.078**	0.08	0.094	0.06	0.066	0.063
	d6	0.149	0.144	0.132	0.105	0.096	**0.089**	0.086	0.087	0.106	0.052
	d7	0.141	0.137	0.103	0.127	0.093	0.09	**0.106**	0.051	0.076	0.065
	d8	0.119	0.103	0.097	0.124	0.127	0.089	0.13	**0.093**	0.083	0.069
	d9	0.13	0.164	0.128	0.122	0.115	0.129	0.133	0.098	**0.095**	0.071
	d10	0.142	0.148	0.167	0.165	0.104	0.128	0.163	0.14	0.139	**0.091**

Table B13: Percentage of non-white husbands and wives for the years 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.238**	0.207	0.184	0.201	0.143	0.177	0.214	0.137	0.178	0.207
	d2	0.239	**0.208**	0.196	0.149	0.168	0.172	0.147	0.208	0.128	0.149
	d3	0.219	0.219	**0.184**	0.149	0.173	0.104	0.161	0.171	0.144	0.186
	d4	0.23	0.204	0.198	**0.178**	0.147	0.145	0.115	0.116	0.132	0.188
	d5	0.215	0.175	0.149	0.153	**0.152**	0.158	0.137	0.163	0.189	0.199
	d6	0.205	0.188	0.158	0.155	0.179	**0.154**	0.197	0.164	0.15	0.169
	d7	0.212	0.169	0.173	0.172	0.17	0.164	**0.15**	0.154	0.165	0.192
	d8	0.192	0.116	0.171	0.165	0.142	0.168	0.171	**0.195**	0.197	0.221
	d9	0.228	0.152	0.2	0.198	0.142	0.173	0.183	0.251	**0.242**	0.268
	d10	0.224	0.236	0.172	0.159	0.207	0.189	0.232	0.212	0.242	**0.268**

Table B14: Percentage of non-white wives 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.253**	0.207	0.19	0.183	0.146	0.161	0.214	0.147	0.137	0.155
	d2	0.244	**0.212**	0.192	0.166	0.168	0.153	0.156	0.181	0.128	0.14
	d3	0.233	0.246	**0.206**	0.166	0.176	0.128	0.14	0.162	0.15	0.199
	d4	0.244	0.207	0.196	**0.183**	0.155	0.176	0.111	0.137	0.114	0.136
	d5	0.24	0.184	0.167	0.164	**0.163**	0.163	0.134	0.18	0.147	0.17
	d6	0.21	0.237	0.151	0.148	0.167	**0.142**	0.2	0.146	0.169	0.144
	d7	0.219	0.16	0.183	0.166	0.165	0.169	**0.14**	0.146	0.162	0.157
	d8	0.219	0.116	0.194	0.193	0.133	0.156	0.179	**0.195**	0.19	0.197
	d9	0.236	0.152	0.244	0.189	0.167	0.17	0.198	0.192	**0.208**	0.239
	d10	0.262	0.236	0.155	0.176	0.195	0.171	0.218	0.2	0.212	**0.249**

Table B15: Percentage of non-white husbands 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.222**	0.162	0.19	0.176	0.026	0.176	0.026	0.067	0.114	0.04
	d2	0.165	**0.138**	0.063	0.054	0.137	0.043	0.067	0.075	0.026	0.0
	d3	0.192	0.116	**0.071**	0.054	0.083	0.14	0.077	0.06	0.071	0.027
	d4	0.183	0.132	0.17	**0.113**	0.021	0.056	0.053	0.106	0.0	0.0
	d5	0.152	0.136	0.072	0.09	**0.019**	0.108	0.048	0.038	0.0	0.098
	d6	0.143	0.136	0.145	0.115	0.188	**0.081**	0.106	0.086	0.081	0.039
	d7	0.241	0.094	0.091	0.163	0.103	0.108	**0.183**	0.016	0.0	0.038
	d8	0.071	0.115	0.146	0.186	0.143	0.111	0.148	**0.094**	0.045	0.091
	d9	0.111	0.138	0.139	0.087	0.115	0.13	0.116	0.076	**0.031**	0.06
	d10	0.043	0.286	0.15	0.25	0.116	0.13	0.078	0.069	0.133	**0.048**

Table B16: Percentage of non-white husbands and wives for the years 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.391**	0.394	0.407	0.408	0.502	0.472	0.496	0.559	0.534	0.646
	d2	0.354	**0.413**	0.432	0.453	0.474	0.548	0.535	0.49	0.542	0.612
	d3	0.437	0.422	**0.38**	0.48	0.485	0.562	0.572	0.562	0.567	0.639
	d4	0.406	0.446	0.435	**0.457**	0.532	0.56	0.558	0.638	0.598	0.612
	d5	0.498	0.437	0.524	0.525	**0.554**	0.598	0.644	0.632	0.597	0.692
	d6	0.492	0.435	0.542	0.573	0.593	**0.621**	0.631	0.618	0.669	0.693
	d7	0.493	0.531	0.52	0.571	0.569	0.601	**0.653**	0.657	0.664	0.671
	d8	0.479	0.564	0.574	0.576	0.62	0.642	0.692	**0.721**	0.721	0.681
	d9	0.514	0.618	0.634	0.63	0.609	0.64	0.693	0.712	**0.732**	0.737
	d10	0.583	0.622	0.639	0.686	0.657	0.696	0.745	0.755	0.753	**0.752**

Table B17: Percentage in metro 1976-1980.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**0.622**	0.598	0.549	0.543	0.564	0.594	0.584	0.6	0.616	0.655
	d2	0.578	**0.537**	0.559	0.584	0.54	0.582	0.578	0.606	0.634	0.746
	d3	0.56	0.565	**0.561**	0.531	0.571	0.581	0.588	0.59	0.631	0.776
	d4	0.584	0.579	0.551	**0.56**	0.612	0.565	0.592	0.602	0.671	0.778
	d5	0.642	0.603	0.578	0.586	**0.584**	0.662	0.633	0.636	0.698	0.782
	d6	0.624	0.571	0.599	0.619	0.641	**0.624**	0.707	0.687	0.731	0.826
	d7	0.588	0.582	0.663	0.637	0.661	0.687	**0.701**	0.695	0.75	0.822
	d8	0.662	0.601	0.631	0.654	0.654	0.657	0.73	**0.779**	0.775	0.818
	d9	0.569	0.594	0.589	0.707	0.679	0.739	0.741	0.808	**0.844**	0.875
	d10	0.757	0.756	0.716	0.618	0.716	0.77	0.739	0.864	0.878	**0.904**

Table B18: Percentage in metro 2018-2022.

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.127**	1.302	1.095	0.978	0.858	0.886	0.789	0.686	0.683	0.56
	d2	1.712	**1.374**	1.196	1.115	0.983	0.868	0.828	0.779	0.761	0.549
	d3	1.202	1.386	**1.033**	1.083	1.11	0.851	0.851	0.844	0.773	0.594
	d4	1.188	1.454	1.137	**1.144**	0.964	0.948	0.928	0.915	0.784	0.676
	d5	0.927	1.182	1.179	1.173	**1.153**	0.996	0.974	0.988	0.949	0.767
	d6	0.714	1.0	1.003	1.094	1.016	**1.041**	0.972	0.993	0.936	0.814
	d7	0.627	0.834	1.055	1.082	1.116	1.101	**1.063**	1.139	1.153	0.99
	d8	0.577	0.622	0.784	1.146	1.065	1.082	1.106	**1.126**	1.2	1.153
	d9	0.484	0.657	0.688	0.773	1.172	1.208	1.296	1.301	**1.228**	1.402
	d10	0.386	0.412	0.419	0.583	0.716	0.946	1.205	1.31	1.563	**2.364**

Table B19: Estimated sorting measures for the years 1976-1980 at different quantiles

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.799**	1.546	1.067	0.901	0.895	0.799	0.548	0.604	0.472	0.368
	d2	1.789	**1.745**	1.353	1.109	0.94	0.833	0.656	0.681	0.541	0.352
	d3	1.21	1.683	**1.429**	1.195	0.977	0.877	0.875	0.636	0.626	0.493
	d4	0.861	1.204	1.594	**1.244**	1.127	1.003	0.935	0.776	0.708	0.548
	d5	0.749	0.964	1.244	1.346	**1.087**	1.075	1.049	1.022	0.85	0.614
	d6	0.662	0.693	0.883	1.336	1.408	**1.234**	1.139	0.999	0.93	0.715
	d7	0.476	0.72	0.982	1.056	1.141	1.369	**1.142**	1.158	1.119	0.836
	d8	0.48	0.55	0.695	0.745	0.961	1.264	1.512	**1.22**	1.499	1.072
	d9	0.421	0.551	0.566	0.687	0.755	0.971	1.229	1.435	**1.801**	1.583
	d10	0.366	0.397	0.388	0.553	0.544	0.673	0.882	1.27	1.829	**3.098**

Table B20: Estimated sorting measures for the years 2018-2022 at different quantiles

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.135**	1.366	1.133	1.009	0.824	0.953	0.755	0.698	0.744	0.382
	d2	1.626	**1.421**	1.201	1.061	0.94	0.85	0.748	0.742	0.744	0.667
	d3	1.365	1.173	**1.142**	1.031	1.159	0.879	0.834	0.833	0.663	0.921
	d4	1.155	1.333	1.192	**1.124**	0.921	0.971	0.886	0.893	0.749	0.776
	d5	0.957	1.176	1.245	1.163	**1.143**	0.989	0.988	0.959	0.923	0.457
	d6	0.752	1.039	1.088	1.147	1.088	**1.132**	1.001	1.024	0.933	0.798
	d7	0.614	0.771	1.107	1.028	1.087	1.132	**1.036**	1.112	1.112	1.002
	d8	0.597	0.596	0.826	1.152	1.082	1.15	1.125	**1.128**	1.177	1.166
	d9	0.429	0.603	0.684	0.738	1.148	1.194	1.253	1.231	**1.177**	1.543
	d10	0.373	0.527	0.379	0.544	0.609	0.747	1.378	1.379	1.779	**2.285**

Table B21: Results of estimated sorting measures for the years 1976-1980 at different quantiles

		Husbands' quantiles									
		d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Wives' quantiles	d1	**2.891**	1.574	1.043	0.893	0.903	0.812	0.572	0.628	0.523	0.161
	d2	1.625	**1.709**	1.261	1.03	0.926	0.802	0.626	0.652	0.516	0.852
	d3	1.164	1.726	**1.381**	1.164	0.997	0.882	0.839	0.596	0.586	0.666
	d4	0.86	1.299	1.613	**1.238**	1.134	1.028	0.957	0.811	0.701	0.358
	d5	0.774	1.031	1.25	1.377	**1.141**	1.106	1.056	1.063	0.809	0.394
	d6	0.662	0.719	0.913	1.298	1.427	**1.23**	1.152	1.038	0.912	0.65
	d7	0.546	0.789	0.968	1.062	1.243	1.45	**1.279**	1.247	1.192	0.226
	d8	0.417	0.458	0.608	0.627	0.829	1.095	1.421	**1.189**	1.253	2.103
	d9	0.453	0.59	0.598	0.741	0.763	1.011	1.279	1.557	**1.799**	1.208
	d10	0.605	0.109	0.363	0.575	0.771	0.451	0.814	1.221	1.709	**3.382**

Table B22: Results of estimated sorting measures for the years 2018-2022 at different quantiles