Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15912

Marital Sorting and Inequality: How Educational Categorization Matters

Frederik Almar
Benjamin Friedrich
Ana Reynoso
Bastian Schulz
Rune Vejlin

JANUARY 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15912

Marital Sorting and Inequality: How Educational Categorization Matters

Frederik Almar
Aarhus University

Benjamin Friedrich
Northwestern University

Ana Reynoso
University of Michigan

Bastian Schulz
Aarhus University, Dale T. Mortensen Centre and CESifo

Rune Vejlin
Aarhus University and IZA

JANUARY 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 and Inequality: How Educational Categorization Matters*

This paper revisits the link between education-based marriage market sorting and income inequality. Leveraging Danish administrative data, we develop a novel categorization of marriage market types based on the starting wages and wage growth trajectories associated with educational programs: ambition types. We find a substantial increase in sorting by educational ambition over time, which explains more than 40% of increasing inequality since 1980. In contrast, sorting trends are flat with the commonly used level of education. Hence, the mapping between education and marriage-market types matters crucially for conclusions about the role of marital sorting in rising income inequality.

JEL Classification: D13, D31, I24

Keywords: marital sorting, inequality, education

Corresponding author:
Rune Vejlin
Aarhus University
Fuglsangs Alle 4
8240 Aarhus N
Denmark
E-mail: rvejlin@econ.au.dk

* We thank Basit Zafar, Elena Mattana, Rasmus Landersø, and seminar and workshop participants at Aarhus University, IAB Nuremberg, and the DGPE network for helpful conversations and discussions. Hans Magnus Appel Laursen provided excellent research assistance. Access to the Danish register data used in this study is provided by ECONAU, Aarhus University. Bastian Schulz thanks the Aarhus University Research Foundation (Grant no. AUFF-F-2018-7-6) and the Dale T. Mortensen Centre at the Department of Economics and Business Economics, Aarhus University, for financial support

1 Introduction

An ongoing debate questions the contribution of increasing education-based assortative matching in the marriage market to rising household income inequality. Some studies find evidence that marital sorting has strengthened over the last decades, which—because income increases with education—has contributed to rising income inequality across households (Fernández and Rogerson, 2001; Greenwood, Guner, Kocharkov and Santos, 2014, 2016; Mare, 2016; Hryshko, Juhn and McCue, 2017; Ciscato and Weber, 2020; Calvo, Lindenlaub and Reynoso, 2022); however, other papers argue against these findings (Kremer, 1997; Breen and Salazar, 2011; Breen and Andersen, 2012; Eika, Mogstad and Zafar, 2019; Gihleb and Lang, 2020).

In this paper, we use rich administrative data from Denmark to show that the common categorization of education by level (primary, secondary, bachelor's, and master's/PhD) conceals important heterogeneity that affects conclusions on whether sorting has changed and influenced inequality. To make progress, we develop a novel education-based categorization of marriage-market types based on granular *education programs*. Education programs matter for marriage market matching because they signal time commitments between career and family and affect meeting probabilities (Nielsen and Svarer, 2009; Wiswall and Zafar, 2021; Kirkeboen, Leuven and Mogstad, 2021). In particular, education programs feature remarkable heterogeneity in initial conditions and lifetime labor market outcomes (Altonji, Kahn and Speer, 2014, 2016; Kirkeboen, Leuven and Mogstad, 2016), which may serve as predictors for potential partners' expected inputs into household income, home production, and childcare: high wage growth can be associated with long working hours and limited flexibility (Goldin, 2014) or low fertility (Jones, Schoonbroodt and Tertilt, 2008).

To capture this heterogeneity in the labor market prospects of education programs, we merge Danish education registers with the labor market histories of all program graduates and compute *starting wages* and *wage growth* trajectories for each program. We group individuals based on their educational program similarity in these two dimensions using k-means clustering, a well-established and popular partitioning method in machine learning and computer science (Steinley, 2006). This method has also recently been introduced to economic research (Bonhomme and Manresa, 2015) and applied to categorize unobserved worker and firm types in the labor market (Bonhomme, Lamadon and Manresa, 2019, 2022). To our knowledge, we are the first to apply this method to construct marriage market types. Our method successfully clusters the more than 1800 education programs in Denmark into four clearly distinct groups based on whether starting wages and wage growth are high or low. We interpret individuals pursuing high-starting-wages/high-wage-growth programs as *ambitious* in their career and label our categorization *educational ambition*. In contrast, groups based on *educational levels* mask heterogeneity in both dimensions.

Our first finding is that sorting on educational ambition has increased significantly. Since 1980, an increasing number of graduates of ambitious educational programs have married someone with a similar degree. During the same period, sorting on educational levels has not changed. Thus, conclusions about sorting trends crucially depend on the categorization of types. For both categorizations, we follow Eika et al. (2019) and Chiappori, Costa Dias and Meghir (2020) and flexibly control for the changing education distributions by defining our sorting measures as the weighted sum of the frequency of equally educated couples relative to the same frequency under random matching. This measure is robust to mechanical changes that occur when the type distributions of women and men change over time; consequently, trends in sorting based on both categorizations can be compared.

Our second finding is that changes in who marries whom in terms of educational ambition explain more than 40% of the overall rise in income inequality across couples (as measured by the Gini coefficient) between 1980 and 2018. In contrast, changes in sorting on education level have a negligible effect on inequality. Methodologically, we compare the observed across-household inequality measure every year to the counterfactual measure that results from reshuffling individuals into households so that martial sorting stays at the 1980 levels—a decomposition method inspired by Eika et al. (2019), Fortin, Lemieux and Firpo (2011), and DiNardo, Fortin and Lemieux (1996). We also consider the contribution of labor market returns to ambition types and find that it has been a major determinant of inequality growth. Moreover, the increasing relative number of graduates from ambitious educational programs has amplified inequality overall.

In summary, our new classification of marriage market types considers heterogeneity in educational outcomes beyond the education level. Accounting for this heterogeneity is crucial to reveal substantial changes in who marries whom over time, and these changes have played an important role in the secular increase in inequality. Hence, our analysis more broadly suggests that richer type classifications than the commonly used level of education are critical for future research on marital sorting patterns and household labor market outcomes.

The paper is organized as follows: Section 2 introduces our data. In Section 3, we discuss differences between common educational categories and educational-ambition types and show how these differences affect the measurement of marital sorting. Section 4 presents our analysis of the drivers of changes in inequality and Section 5 concludes.

2 Data

We use Danish register data for our analysis. Unique person IDs identify individuals across registers. The population registers contain demographic variables and the person ID of the (married or cohabiting) partner for all residents between 1980 and 2018 (Statistics Denmark BEF, 1990–2018). We include both legally married and cohabiting opposite-sex couples in the analysis and, for brevity, refer to both types of couples as married in the remainder of the paper.[1] We select all residents in the age range 19-60. This corresponds to an average of 3,031,511 individuals per year. The population of couples with both partners in the age range 19-60 consists of 1,800,866 individuals per year on average. There is an upward (downward) trend in cohabitation (legal marriage), but the combined stock of couples is stable over time.[2] In the education register, four-digit educational program codes (ISCED) uniquely identify all educational programs in Denmark (Statistics Denmark UDDA, 1990–2018). Based on the income register, we calculate yearly household income by adding the labor income of spouses, which includes both wages and income from self-employment (Statistics Denmark IND, 1990–2018).[3] For the calculation of the labor market outcomes for each program, which underlie the educational-ambition types, we use the employment register (Statistics Denmark IDAN, 1990–2018). To abstract from the intensive margin, we deflate hourly wages by regressing wages on year effects with 2000 as the base year.

3 Education and Marriage Market Sorting

3.1 Education-Based Marriage Market Types

Common education-based categorizations group programs based on educational level, i.e., primary (compulsory schooling), secondary (high school, vocational degrees), and tertiary (higher) education. Following Eika et al. (2019), we divide the tertiary category into two subcategories to take the program length into account: short tertiary programs (bachelor degree programs of four years or less, at vocational colleges and universities) and long tertiary programs (master/PhD programs with study times of five years or more, graduate-level education at universities).

[1]While cohabitation is not a legal status per se, couples who are considered to be cohabiting (samlevende) enjoy some of the same status as married couples. Cohabiting couples are identified based on a number of criteria: two opposite-sex individuals who have a joint child and/or share the same address, exhibit an age difference of less than 15 years, have no family relationship, and do not share their accommodation with other adults.

[2]Figure A.1 depicts the evolution of the stocks of different couple types and their age composition.

[3]Specifically, we use the variables $ERHVERVSINDK\_13$.

The novel educational-ambition categorization we develop is also based on educational attainment, but it groups programs based on the labor market outcomes of graduates. For each educational program in Denmark, we measure two outcomes: starting wage, w_0, and wage growth, g. To compute these, we use all individuals in the data who completed their education after 1980.[4] The starting wage, w_0, is the average hourly wage of individuals during the first five years in the labor force.[5] To calculate the average growth rate g, we measure the percentage change between w_0 and w_1, where w_1 is the average hourly wage of individuals in years 9-11 in the labor force.[6] We average over years for both w_0 and w_1 to smooth out year-to-year variation that is unrelated to worker productivity. To cluster individuals based on starting wage and growth, we use the k-means algorithm. The method minimizes the within-cluster variation in the two dimensions and thus produces relatively homogeneous groups in terms of starting wages and growth.[7] For our benchmark results, we use four clusters.[8]

We compare the educational-level and educational-ambition categorizations in terms of starting wages and wage growth in Figure 1. The two panels locate all programs in the space of standardized starting wages and wage growth. There is much heterogeneity across programs in both dimensions, and the educational-level classification hardly takes this into account. In Panel (a), which depicts four groups based on the level of education, one can discern a ranking in terms of starting wages, which are on average low for compulsory schooling (blue squares) and high for long tertiary education (gray diamonds). However, the overlap is vast. Many secondary (red circles) and short tertiary (small orange diamonds) programs have higher starting wages than longer tertiary programs. In the growth dimension, there is no clear pattern.

Panel (b) shows how the k-means algorithm groups educational programs. By construction, programs within a cluster are similar in terms of starting wages and growth. To discern a ranking, note that programs in the gray (diamonds) cluster have the lowest starting wages and wage growth. This is the bottom category. The orange cluster (triangles) includes programs with low starting wages but relatively high wage growth. This is the second category. The blue cluster (squares) is delimited from above in terms of wage growth, but starting wages are high. Thus, we rank the blue cluster above the gray and orange clusters. The red cluster (circles) is the top category that includes programs with high starting wages and high wage growth.

[4]That is, the expected wage growth g_i of an individual who enters in 1990 is based on the observed wage growth trajectories of all that enter the educational program later and prior, e.g., in 1980 and 2005.

[5]We define labor market entry as the year in which individuals complete the highest education obtained before turning 35 if observed or the highest education observed at the oldest age if not turning 35 before 2018.

[6]We focus on the first 9-11 years because wage profiles stabilize later in one's career (Bhuller et al., 2017).

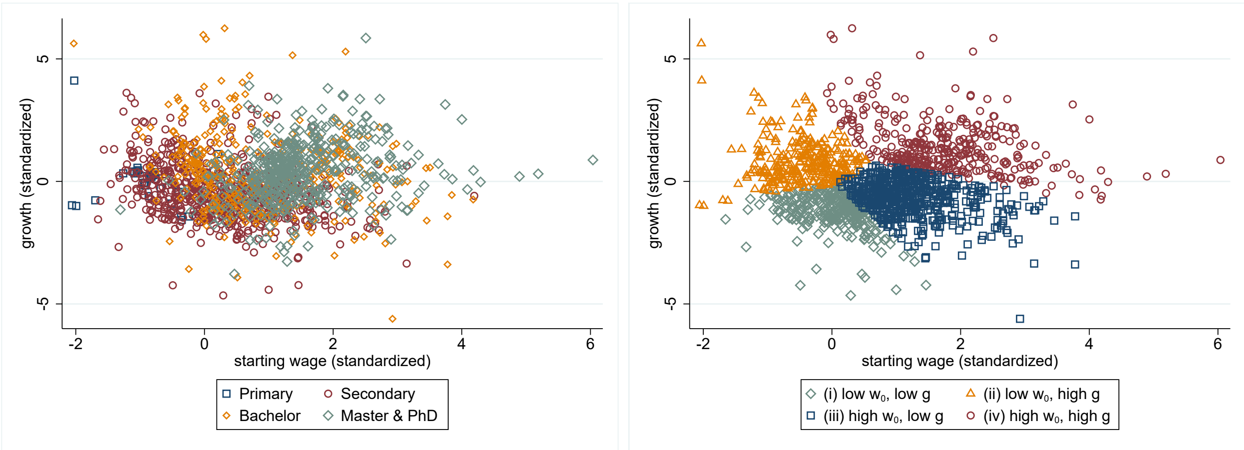
[7]For details on the algorithm and further references, see Steinley (2006).

[8]Results for three and five clusters and three levels of education are depicted in Figure A.4 in the Appendix.

Figure 1: Educational Program Categorizations with Starting Wages and Wage Growth

(a) Educational level, all programs

(b) Educational ambition, all programs



Note: Each point in the figures represents the average wage growth (g) and starting wages (w_0) in educational programs with at least ten graduates observed by 2018. The symbols of points refer to the educational-level or educational-ambition type associated with the program. Axes are standardized hourly starting wages and standardized hourly wage growth in the first ten years after graduation. The sample includes the full population as defined in Section 2.

To obtain a sense of which educational programs are included in the respective clusters, we locate 14 large[9] educational programs in the same space for both educational-level and educational-ambition categorizations (Figure A.3 isolates them). Three large 5-year tertiary programs—law, business, and medicine—as well as bachelor's degrees in business end up in the top educational-ambition category (red). Another 5-year tertiary program—architecture[10]—is, however, part of the blue cluster due to relatively low wage growth. Thus, the k-means algorithm combines architecture with programs that are similar in terms of labor market outcomes but separate according to the level of education, e.g., teachers, nurses, and carpenters.[11] The low-starting-wage/low-wage-growth category includes high school degrees and preschool teachers. The orange low-starting-wage/high-wage-growth cluster is quite heterogeneous. It includes compulsory schooling, degrees from high schools that specialize in business, and vocational degrees as office clerks or bank advisors. For all these programs, graduates have low starting wages and relatively high wage growth despite very different levels of education.

Additionally, we describe the four educational ambition categories in terms of population shares, education levels, and income (Table A.1 provides a summary table). More than 20% of the individuals in our sample graduated from programs in the category with low starting wages and growth. The majority of them are females (64.8%). 47.5% of graduates are assigned

[9]These are the 3-4 largest four-digit educational programs in terms of the number of graduates within each of the four clusters. We count graduates in the 2018 sample of couples as defined in Section 2. In total, the examples cover 21% of all graduates in this sample.

[10]Architecture is a relatively small program. We include it to illustrate heterogeneity within tertiary education.

[11]Teachers and nurses have bachelor's degrees. Carpenters have a secondary, vocational degree.

to the second category with similar low starting wages but higher wage growth. Here, 56% are females. Programs in the second highest category with high starting wages but relatively low wage growth generate 22.7% of all graduates. The top group with high starting wages and high wage growth is the smallest with 9.7% of graduates. Interestingly, only approximately one-third of graduates are female in the two top categories. In categories one and two, the majority of individuals have received primary or secondary education. The share of graduates with a tertiary degree increases in the ambition type. However, even in the top group, we see a sizable minority (10.3%) with secondary degrees. Finally, note that the ranking we describe aligns well with annual income, which clearly increases with increasing ambition.

3.2 Measurement of Sorting

From an empirical perspective, positive assortative matching (PAM) manifests itself as a positive association of spousal types in the cross section. This association can in principle be measured based on correlation coefficients, distance measures, or the frequency distribution of couples' types, which is often represented by a contingency table. Determining whether sorting changes over time has proven elusive because the marginal distributions of types in the marriage market change as well. We follow Eika et al. (2019) and Chiappori et al. (2020) who focus on the frequency-based approach. A benefit of this approach is that it takes changing marginal distributions into account by constructing weights for couples with different type combinations.

Assume that every couple i consists of two individuals, (i, m) and (i, f), where m and f indicates males and females, respectively. Each individual has a one-dimensional type t. Let this type be a categorical variable with $t \in \{1, \ldots, j, \ldots, N\}$, where N is the number of categories. For example, these categories may represent levels of education (primary, secondary, tertiary, i.e., $N = 3$). For each category, we compute a likelihood index that relates the observed frequency of couples to the expected frequency under random matching (the denominator):

$$s(j, j') = \frac{P(t_{i,m} = j, t_{i,f} = j')}{P(t_{i,m} = j)\, P(t_{i,f} = j')}. \tag{1}$$

Here, j (j') is the male (female) type. Note that a ratio above one for $j = j'$ is indicative of PAM. By summing across the categories in which the male and female types are identical, $j = j'$, we obtain the sorting measure

$$\mathcal{S} = s(1,1) \times w_1 + s(2,2) \times w_2 + \cdots + s(N,N) \times w_N, \tag{2}$$

where $\{w_1 \dots w_N\}$ are the weights for the respective categories in which male and female types are identical ($j = j'$). We follow Eika et al. (2019) and construct the weights based on the expected frequencies under random matching:

$$w_j = \frac{P(t_{i,m} = j)\, P(t_{i,f} = j)}{\sum_{k=1}^{N} P(t_{i,m} = k)\, P(t_{i,f} = k)}. \tag{3}$$

Greenwood et al. (2016) divides the sum of the diagonal elements (trace) of the matrix formed by the contingency table by the trace of the counterfactual matrix under random matching. Similar to (1), this measure assesses the extent of sorting by relating it to the expected frequency distribution under random matching, so a measure above one indicates PAM, but it does so for all cells along the diagonal without an explicit weighting scheme. Notably, therefore, the Greenwood et al. (2016) measure is mathematically equivalent to the weighted sum of likelihood indices in Eika et al. (2019), which uses (3) as weights. Consider an example with $N = 2$, although the point holds for any number of categories. We multiply the weight in Equation (3) with the likelihood index in Equation (1). The product of the two marginal distributions for the respective category cancels out, and one is left with the sorting measure

$$\mathcal{S} = \frac{P(t_{i,m} = 1, t_{i,f} = 1) + P(t_{i,m} = 2, t_{i,f} = 2)}{P(t_{i,m} = 1)\, P(t_{i,f} = 1) + P(t_{i,m} = 2)\, P(t_{i,f} = 2)}, \tag{4}$$

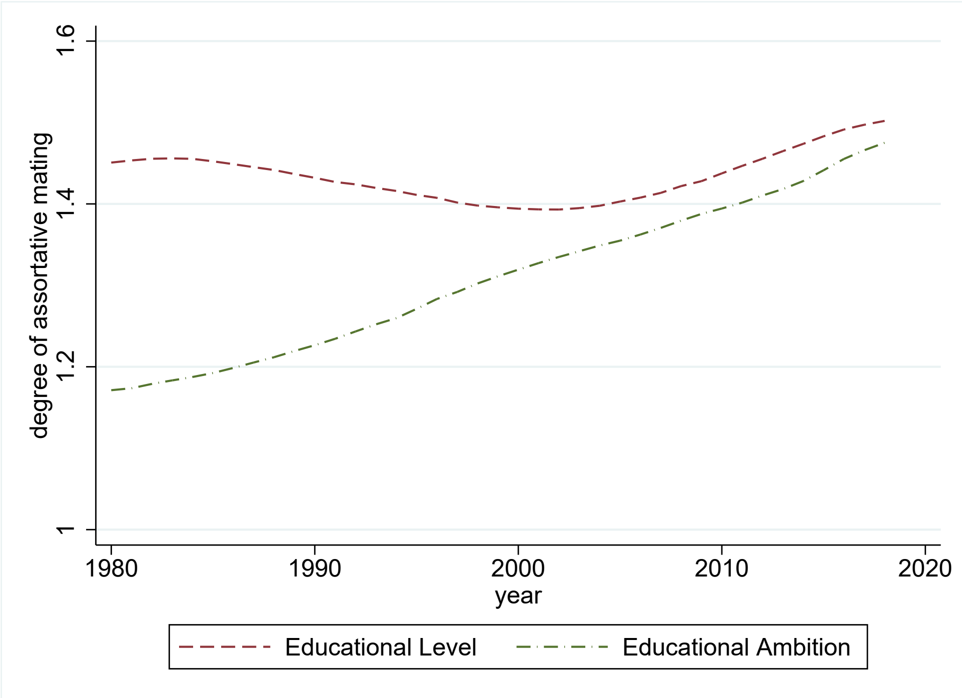
which is exactly the Greenwood et al. (2016) measure.[12] This measure relates the observed frequencies to the frequencies under random matching for all type combinations along the diagonal. Although the weighting is not explicit, the random matching counterfactual in the denominator takes the marginal distributions and their changes over time into account. Therefore, we can interpret the sorting measure $\mathcal{S}$ similarly to (1): a value above one indicates PAM.

Figure 2 depicts the sorting measure $\mathcal{S}$ for educational-level types (dashed red) and educational-ambition types (dashed-dotted green). Overall, sorting is positive, as both measures are consistently greater than one. Based on the educational-level categories, we find that sorting has not increased since 1980. This is consistent with what Eika et al. (2019) find for the US. Based on the educational-ambition types, however, we find a strong increase in sorting. Relative to random matching, the likelihood of observing couples with the same educational-ambition type has increased from below 1.2 to approximately 1.5.

The striking difference in the sorting trend between the two categorizations is explained by the different evolution of marginal distributions (Figure A.5) and likelihood ratios (Figure A.2)

[12] See Table 1 and its discussion in Greenwood et al. (2016).

Figure 2: Significant Increase in Marital Sorting based on Educational-Ambition Types



Note: The sorting measures $\mathcal{S}$ (equation (2)) for educational-level types and educational-ambition types both rely on categorizations with 4 types. Weights are defined in Equation (3). The sample includes all couples as defined in Section 2.

over time. In the educational-level categorization, the "secondary" category is large and relatively stable for both men and women. The likelihood ratio in this category is just slightly above one and flat, indicating that PAM among individuals with secondary education is neither pronounced nor increasing over time. Due to the size of this group, its trend dominates the red dashed line in Figure 2. In addition, the tendency to match assortatively decreases in the growing tertiary categories and increases in the shrinking primary education category, consistent with Eika et al. (2019) for the US. Initially, the primary category shrinks faster than the tertiary categories grow. This explains the wave-like pattern for educational-level sorting.

For educational-ambition sorting, the picture is quite different. In terms of likelihood ratios, we obtain a clear distinction between the top group with high starting wages and wage growth, in which sorting is consistently positive but falls over time. The likelihood ratio decreases from more than 7 to approximately 3. In contrast, there is very little PAM in the other three educational-ambition categories. Consequently, as the top group grows in size, its weight increases, and the overall sorting measure reflects PAM within this group to a larger extent. This explains the increasing trend for educational-ambition types in Figure 2. In sum, the flat trend for educational level is the result of stable sorting in the large and heterogeneous (recall Figure 1, Panel (a)) group with secondary education and opposing sorting trends in the groups with primary and tertiary education. The increasing trend for educational ambition, on the other hand, is driven by the increasing size of the ambitious top group in which sorting is positive relative to the bottom groups with little sorting.

4 Marriage Market Sorting and Inequality

To study the link between marriage market sorting and inequality, we follow Eika et al. (2019) and implement a decomposition exercise inspired by DiNardo et al. (1996). We construct a stochastic matching algorithm that rematches married individuals under three different counterfactual scenarios: (i) fixed marriage market sorting; (ii) fixed labor market returns to educational type; and (iii) fixed composition of the population in terms of educational types. For each scenario, we compare differences in between-household inequality for actual and rematched households and analyze how sensitive the results are to alternative categorizations of marriage-market types. We use the Gini coefficient as a measure of overall inequality and zoom in on the upper and lower halves of the income distribution using percentile ratios.

The matching algorithm works as follows. At the initialization, we split all married couples into singles. The algorithm samples pairs of potential spouses based on the male and female marginal type distributions and forms new couples based on counterfactual matching probabilities, p, for different type combinations. To calculate the matching probabilities, we map the likelihood indices $s(j, j')$ (Equation 1) onto $p \in [0, 1]$ by dividing each index by the sum of indices for all potential partner types. This number may differ for men and women, so we take the average to compute the matching probability of a (j, j') couple. We then draw from a binomial distribution with parameter p to determine whether a match is formed. For individuals who remain unmatched in the first round, we repeat the process until all individuals are matched with a new partner.

Before we discuss the counterfactual scenarios (i)–(iii), we address a potential problem. Our matching algorithm is one-dimensional, i.e., it takes only educational types into account, However, the data-generating process could feature multidimensional matching. If one of the dimensions that we do not model was correlated with individual earnings potential, sorting within cells could arise and bias the counterfactual inequality measures.

To investigate if this is a problem, we use the algorithm to rematch all married individuals randomly within couple-type-combination cells ($t_{i,m} = j, t_{i,f} = j'$). Comparing the resulting inequality measures to the actual ones gives us an indication of how restrictive the one-dimensional matching algorithm is. The Gini coefficient, the 90/50 percentile ratio, and the 50/10 percentile ratio in the real data in 2018 are 0.307, 1.688, 2.518, respectively. With the random reshuffling of educational-ambition types within cells, we obtain 0.294, 1,693, and 2.167. The same conclusions hold if we rematch couples based on educational-level types. In this case, the numbers are 0.291, 1.675, and 2.178. Thus, we conclude that the matching algorithm works well, especially for the upper half of the income distribution.

Next, we analyze a random matching scenario in which the matching probability p is set to 50%. This scenario corresponds to no sorting, i.e., all likelihood indices are 1. Without sorting, the level of income inequality in 2018 is significantly lower. The Gini coefficient is 0.281 instead of 0.307, i.e., inequality would have grown by 18% instead of 28% between 1980 and 2018 if marriage market matching were random. Interestingly, for 1980, the level of inequality implied by random matching is not far from the truth. The reason is that sorting was not very pronounced in 1980 outside the top group, which was relatively small in 1980.[13] Thus, in terms of the implied inequality, random matching is a good approximation of marriage market matching in Denmark for 1980 but a bad approximation for 2018, which is another reflection of the fact that positive sorting has increased.

(i) Fixed marriage market sorting

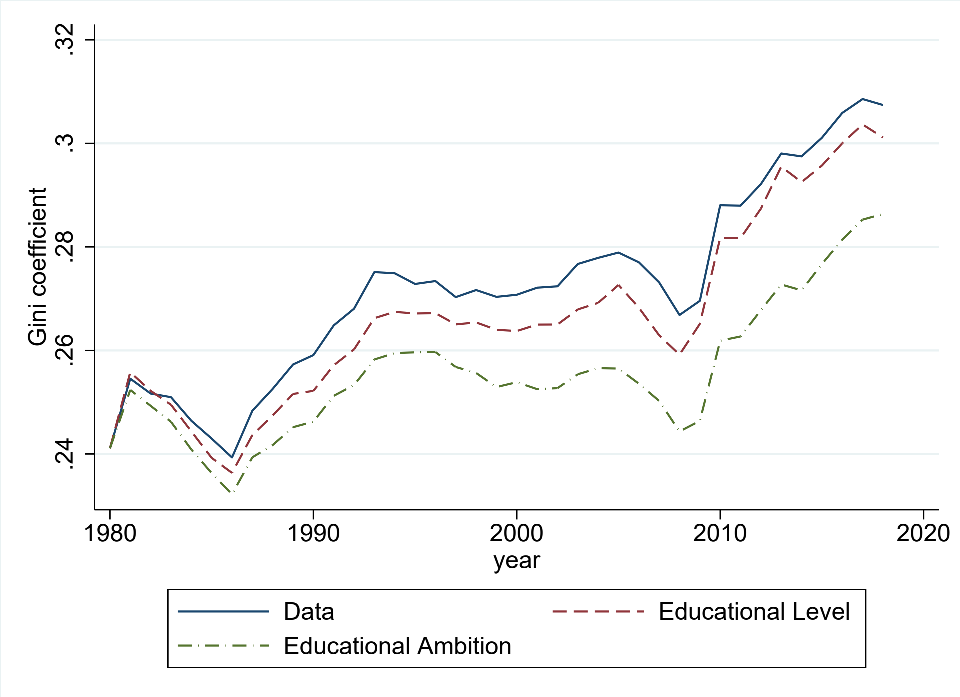
We construct scenario (i) by rematching couples based on the matching probabilities from 1980. To keep sorting fixed, we fix both the likelihood indices (1) and the marginal type distributions that they depend on at their 1980 levels. The year-τ counterfactual data is constructed by resampling year-τ individuals to obtain the same marginal distributions as in 1980. Then, we use the matching algorithm to create new couples based on the 1980 matching probabilities.

Figure 3 shows the resulting gaps in inequality for all years τ between 1981 and 2018. Based on both educational-level and educational-ambition types, we find that inequality would have increased less with fixed sorting. All lines are below the trend in the data (solid blue line). However, the red dashed line for fixed educational-level sorting is much closer to the trend in the data than the green dashed line for educational-ambition sorting. That is, increasing positive sorting by educational ambition amplifies the inequality trend, while flat sorting by educational level contributes little.

In Table 1, we zoom in on the end points in Figure 3 and decompose the total change in household-level income inequality between 1980 and 2018 for the three scenarios. Column (a) shows the results for the Gini coefficient, which summarizes inequality in the entire distribution; column (b) and (c) show the 90/50 and 50/10 percentile ratios, respectively. For each

[13]See the marginal type distributions in Figure A.5 and the likelihood indices in Figure A.2.

Figure 3: Growth in Educational-Ambition Sorting Amplified Inequality



Note: The 1980–2018 development in the Gini coefficient for the joint income of spouses observed in the data and in the counterfactual scenario (i) of keeping marriage market sorting fixed at the 1980 level. The sample includes all couples as defined in Section 2.

inequality measure, the first row contains inequality changes in the data (Δ_{Data}). Household-level income inequality between couples has increased according to all three measures. The Gini coefficient has increased by 0.066, the 90/50 percentile ratio has increased by 0.165, and the 50/10 percentile ratio has even increased by 0.573. These changes correspond to 100%, which is our baseline.

With fixed sorting, see Panel (i), the counterfactual 2018 Gini coefficient amounts to 57% of the true coefficient with educational-ambition types. The corresponding number with educational-level types is 91%. For the 90/50 ratio in (b), fixed educational-level sorting leads to a counterfactual inequality measure that is even slightly above the benchmark (110%). That is, sorting based on educational levels has somewhat mitigated inequality in the upper half of the income distribution. However, the change in the 90/50 ratio under fixed educational-ambition sorting is only 54% of the true change, which suggests that increasing positive sorting by educational ambition amplifies inequality in the upper half of the income distribution. In the lower half, captured by the 50/10 ratio in column (c), both educational-level and educational-ambition sorting have amplified inequality, but the amplification we see is stronger with educational ambition types.

Table 1: Changes in Income Inequality (1980–2018)

	(a) Gini		(b) $P90/P50$		(c) $P50/P10$	
Factual change (Δ_{Data})	0.066	100%	0.165	100%	0.573	100%
(i) Fixed sorting	Δ_{Gini}	$\frac{\Delta_{Gini}}{\Delta_{Data}}$	$\Delta_{P90/P50}$	$\frac{\Delta_{P90/P50}}{\Delta_{Data}}$	$\Delta_{P50/P10}$	$\frac{\Delta_{P50/P10}}{\Delta_{Data}}$
Educational Level	0.060	91%	0.182	110%	0.390	68%
Educational Ambition	0.038	57%	0.089	54%	0.187	33%
(ii) Fixed returns						
Educational Level	0.010	15%	0.127	77%	-0.060	-10%
Educational Ambition	0.007	11%	0.080	49%	-0.029	-5%
(iii) Fixed marginals (both)						
Educational Level	0.094	142%	0.197	119%	1.731	302%
Educational Ambition	0.062	93%	0.110	67%	0.750	131%
(iiia) Fixed marginals (male)						
Educational Level	0.060	91%	0.109	66%	0.719	125%
Educational Ambition	0.058	87%	0.121	74%	0.592	103%
(iiib) Fixed marginals (female)						
Educational Level	0.093	141%	0.218	133%	1.125	196%
Educational Ambition	0.067	101%	0.146	89%	0.633	110%

Note: Panel (a) reports the Gini coefficient, while Panels (b) and (c) report the ratio of the 90th and 50th percentile and the ratio of the 50th and 10th percentile in the income distribution. The first row shows the inequality changes in the data (Δ_{Data}). For each of the counterfactual scenarios (i)-(iiib), we first report the counterfactual change, e.g., Δ_{Gini}, and then the counterfactual change relative to the change in the data, e.g., $\frac{\Delta_{Gini}}{\Delta_{Data}}$. The sample includes all couples as defined in Section 2.

(ii) Fixed labor market returns to educational type

In this scenario, we analyze how income inequality would have developed had the income premia of high-type (i.e., tertiary or ambitious) educational programs remained unchanged. To this end, we introduce a household reweighting factor, $\widehat{\psi_y}$, to construct the counterfactual household income distribution:

$$\widehat{F}(y|\tau_y = 1980, \tau_x = 2018, \tau_s = 2018) = \int F_{Y|X}(y|x, \tau_y = 1980)\psi_y dF(x|\tau_x = 1980), \quad (5)$$

where subscripts y denote household income, x is the couple combination of $(t_{i,m} = j, t_{i,f} = j')$, s matching probabilities, and τ time and

$$\widehat{\psi_y} = \frac{P(\tau_x = 2018|x, \tau_s = 2018)}{P(\tau_x = 1980|x, \tau_s = 2018)} \frac{P(\tau_x = 1980)}{P(\tau_x = 2018)} \quad (6)$$

is the reweighting factor. We obtain $\widehat{\psi_y}$ following Fortin et al. (2011). To obtain the conditional probability of being in 1980 for couple combination x under the matching probabilities at

$\tau_s = 2018$, we use the rematching algorithm on 1980 households with implied 2018 probabilities. Intuitively, couple combinations x that are relatively more present in 2018 are weighted greater than one in the counterfactual income distribution, and vice versa.

Overall, we find that changing returns to educational types are the major source of increasing income inequality; see Panel (ii) in Table 1. Without changing returns, the increase in the Gini is only 15% of the true increase for educational-level types and 11% for educational-ambition types. That is, without the rising income premium that highly educated individuals receive relative to less educated individuals, inequality would have barely changed, and this conclusion holds irrespective of how we categorize marriage market types. However, there are some interesting differences between the upper and lower halves of the income distribution; see columns (b) and (c). In the upper half (90/50 ratio), we see that increasing returns contributed less to the inequality trend. For educational-level (educational-ambition) types, the 90/50 ratio still increases by 77% (49%) relative to the data. That is, increasing returns to education amplified inequality less in the upper half of the distribution than overall, but changing returns to "ambitious" education programs are more important than changing returns to broadly defined tertiary education. In the lower half of the income distribution (50/10 ratio), we find that absent increasing returns to education, inequality would even have decreased with both categorizations.

(iii) Fixed composition in terms of educational types

To obtain the last counterfactual scenario (iii) in which we hold the marginal distributions fixed, our approach is similar to (ii). We reweight households in the 2018 income distribution based on the changes in the marginal distributions of $t_{i,m}$ and $t_{i,f}$. In this case, the reweighting factor is $\widehat{\psi_x} = (\widehat{\psi_y})^{-1}$. We first keep the type distributions for both genders fixed at the 1980 level, see Panel (iii) in Table 1, and repeat the exercise keeping either the male or female marginal type distribution fixed, see Panels (iiia) and (iiib). Recall that the marginal distributions shifted such that the numbers of individuals in the top categories increased; see Figure A.5. That is, we have more men and women who graduate with tertiary degrees and/or from ambitious educational programs in 2018 compared to 1980, and these changes have been more pronounced for women.

Based on the Gini coefficient in column (a) and educational levels, we find that increasing educational attainment had a mitigating effect on inequality. Without the shift, inequality would have increased to 142% of the true 2018 value. The mitigating effect manifests itself mainly in the lower half of the income distribution. For the 50/10 percentile ratio in column (c), inequality would have been three times higher without changing marginal distributions (302%). For the 90/50 ratio in column (d), we see a modest mitigating effect for educational-level types (119%). In contrast, based on educational-ambition types, we overall find a slight amplification effect due to changing marginals (93%). This effect consists of an amplification in the upper half (67%) and a mitigating effect (131%) in the lower half of the distribution, which is much less pronounced than for educational-level types. The difference in findings across categorizations reflects the fact that top and bottom educational ambition categories are distinct in terms of wage growth while the top and bottom level categories are not (recall Figure 1). Therefore, inequality rises for educational ambition as the number of individuals in the top category increases, but this is not true for educational-level types.

Interestingly, if we keep only the female marginal distributions fixed at the 1980 level, the conclusions from this counterfactual exercise hardly change. The results in Panels (iii) and (iiib) in Table 1 are similar for both categorizations. Only the mitigating effect in the lower half of the distribution is less pronounced, especially for educational levels. This implies that changes to the female marginal distributions are the main reason behind the mitigating effects we find.

If we instead keep only the male marginal distributions fixed (iiia), the mitigating effect on inequality is gone. The resulting inequality is relatively close to the true 2018 value for both educational-level (91%) and educational-ambition (87%) types based on the Gini coefficient. We do see a modest mitigating effect of the male marginal distribution in the lower half of the distribution (50/10 ratio) for educational-level types (125%). In the upper half of the distribution, changing male types amplify income inequality considerably: the counterfactual 90/50 ratio would be only 66% of the 2018 value for educational-level types and 74% for educational-ambition types.

In summary, the importance of distinguishing between the two categorizations is particularly evident from the distinct effects that the changing marginal distributions of female types had on inequality between married households. While the move of females into tertiary education overall had a considerable mitigating effect, their entry into (ambitious) high-wage programs amplified inequality in the upper half of the distribution, and this has offset a mitigating effect in the lower half. For men, the shifts in both marginal distributions amplified inequality overall, but the difference between the two classifications is smaller. Notably, the relative size of the top group increased more for women than for men in both categorizations.[14]

[14]The share of men with long-cycle tertiary education more than tripled between 1980 and 2018. For women,

5 Conclusions

We provide new insights into the relationship between education-based marital sorting and across-household inequality by showing that conclusions depend on how education categories are defined. Using detailed data from Danish education and labor market registers, we cluster education *programs* by average starting wages and wage growth of graduates to define four *educational-ambition* types. Because educational ambition reflects the labor market prospects of individuals significantly better than the usual categorization of education by levels, they are better suited to study marriage market sorting and its effect on inequality.

Our first main result shows an increase of more than 25% in sorting based on the educational-ambition categorization between 1980 and 2018. In contrast, sorting based on the level of education remained close to its 1980 levels throughout this period. This result contributes to the ongoing debate on whether sorting on education has increased over the last few decades. We highlight the previously overlooked fact that the definition of types is crucial.

Our second main result reveals that changes in who marries whom in terms of educational ambition had a large and significant impact on the increase in across-household inequality in Denmark between 1980 and 2018. Had the configuration of couples in terms of educational ambition stayed at their 1980 levels over the last four decades, across-household inequality growth would have been mitigated by approximately 40%. In contrast, marriage market sorting trends based on education level contributed minimally to income inequality growth. This result is independent of how aggregate sorting (and its trend) is measured (a topic that has received much attention in the recent literature) because the counterfactual analysis does not depend on a specific sorting measure.

To motivate future research, we explore how our classification method can be applied when data at the level of the educational program are unavailable. We argue that categorizations based purely on educational level mask relevant heterogeneity to studying marriage market trends and inequality. In our data, part of this heterogeneity within educational levels is captured by the field of study.[15] Thus, combining the level of education, the field of study, and labor market outcomes—variables available in, e.g., the US Panel Study of Income Dynamics and the American Community Survey—can allow researchers to create similar classifications to ours and to gain valuable new insights into the link between marriage sorting and inequality.

this share increased by a factor of 13. Similarly, for educational-ambition types, the share of men in the top category doubled, while there were more than eight times as many highly ambitious women in 2018 compared to 1980. Figure A.5 provides more details.

[15]For example, the field of study separates natural sciences and humanities, two fields that differ greatly in terms of labor market outcomes.

References

Altonji, Joseph G., Lisa B. Kahn, and Jamin D. Speer (2014) "Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs," *American Economic Review*, Vol. 104, pp. 387–93.

——— (2016) "Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success," *Journal of Labor Economics*, Vol. 34, pp. S361–S401.

Bhuller, Manudeep, Magne Mogstad, and Kjell G. Salvanes (2017) "Life-Cycle Earnings, Education Premiums, and Internal Rates of Return," *Journal of Labor Economics*, Vol. 35, pp. 993–1030.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019) "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, Vol. 87, pp. 699–739.

——— (2022) "Discretizing Unobserved Heterogeneity," *Econometrica*, Vol. 90, pp. 625–643.

Bonhomme, Stéphane and Elena Manresa (2015) "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, Vol. 83, pp. 1147–1184.

Breen, Richard and Signe Hald Andersen (2012) "Educational Assortative Mating and Income Inequality in Denmark," *Demography*, Vol. 49, pp. 867–887.

Breen, Richard and Leire Salazar (2011) "Educational Assortative Mating and Earnings Inequality in the United States," *American Journal of Sociology*, Vol. 117, pp. 808–843.

Calvo, Paula A, Ilse Lindenlaub, and Ana Reynoso (2022) "Marriage market and labor market sorting," Working Paper 28883, National Bureau of Economic Research.

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir (2020) "Changes in Assortative Matching: Theory and Evidence for the US," Working Paper 26932, National Bureau of Economic Research.

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

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux (1996) "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, Vol. 64, pp. 1001–1044.

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

Fernández, Raquel and Richard Rogerson (2001) "Sorting and Long-Run Inequality," *The Quarterly Journal of Economics*, Vol. 116, pp. 1305–1341.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (2011) "Decomposition methods in economics," in *Handbook of labor economics*, Vol. 4: Elsevier, pp. 1–102.

Gihleb, Rania and Kevin Lang (2020) "Educational homogamy and assortative mating have not increased," in *Change at Home, in the Labor Market, and On the Job*: Emerald Publishing Limited.

Goldin, Claudia (2014) "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, Vol. 104, pp. 1091–1119.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos (2014) "Marry Your Like: Assortative Mating and Income Inequality," *American Economic Review*, Vol. 104, pp. 348–53.

——— (2016) "Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation," *American Economic Journal: Macroeconomics*, Vol. 8, pp. 1–41.

Hryshko, Dmytro, Chinhui Juhn, and Kristin McCue (2017) "Trends in earnings inequality and earnings instability among U.S. couples: How important is assortative matching?" *Labour Economics*, Vol. 48, pp. 168–182.

Jones, Larry E, Alice Schoonbroodt, and Michele Tertilt (2008) "Fertility theories: can they explain the negative fertility-income relationship?"Technical report, National Bureau of Economic Research.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad (2016) "Field of Study, Earnings, and Self-Selection," *The Quarterly Journal of Economics*, Vol. 131, pp. 1057–1111.

Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad (2021) "College as a Marriage Market," Working Paper 28688, National Bureau of Economic Research.

Kremer, Michael (1997) "How Much does Sorting Increase Inequality?," *The Quarterly Journal of Economics*, Vol. 112, pp. 115–139.

Mare, Robert D. (2016) "Educational Homogamy in Two Gilded Ages: Evidence from Intergenerational Social Mobility Data," *The ANNALS of the American Academy of Political and Social Science*, Vol. 663, pp. 117–139.

Nielsen, Helena Skyt and Michael Svarer (2009) "Educational Homogamy: How Much is Opportunities?" *Journal of Human Resources*, Vol. 44, pp. 1066–1086.

Statistics Denmark BEF (1990–2018) "BEF - Befolkningen, Statistics Denmark: Population and Education, Social statistics.," URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/the-population, Data accessed through the ECONAU project database.

Statistics Denmark IDAN (1990–2018) "IDAN - IDA ansættelser, Statistics Denmark: Labour and Income, Social statistics," URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/labour-market-account, Data accessed through the ECONAU project database.

Statistics Denmark IND (1990–2018) "IND - Indkomst, Statistics Denmark: Labour and Income, Social statistics.," URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/income-statistics, Data accessed through the ECONAU project database.

Statistics Denmark UDDA (1990–2018) "UDDA - Uddannelser (BUE), Statistics Denmark: Population and Education, Social statistics.," URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/highest-education-attained, Data accessed through the ECONAU project database.

Steinley, Douglas. (2006) "K-means clustering: A half-century synthesis," *British Journal of Mathematical and Statistical Psychology*, Vol. 59, pp. 1–34.

Wiswall, Matthew and Basit Zafar (2021) "Human Capital Investments and Expectations about Career and Family," *Journal of Political Economy*, Vol. 129, pp. 1361–1424.

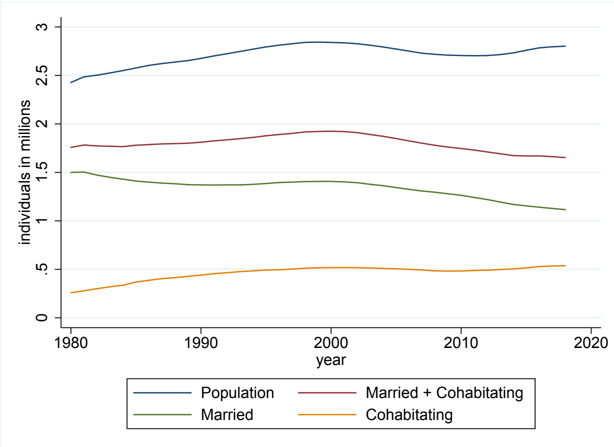
Online Appendix

(not for publication)

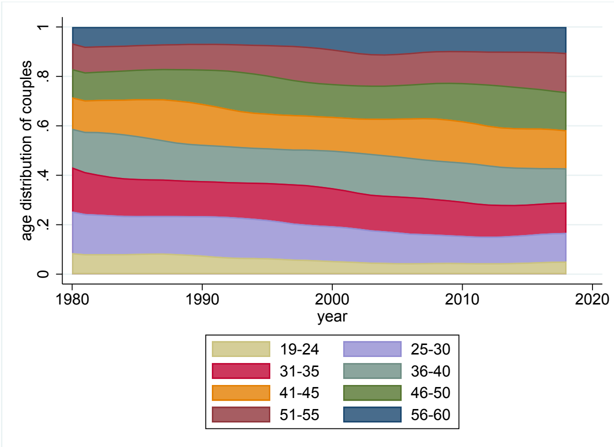
A Complementary Exhibits

Figure A.1: Marriage, Cohabitation, Age Composition

(a) Stocks of Couples



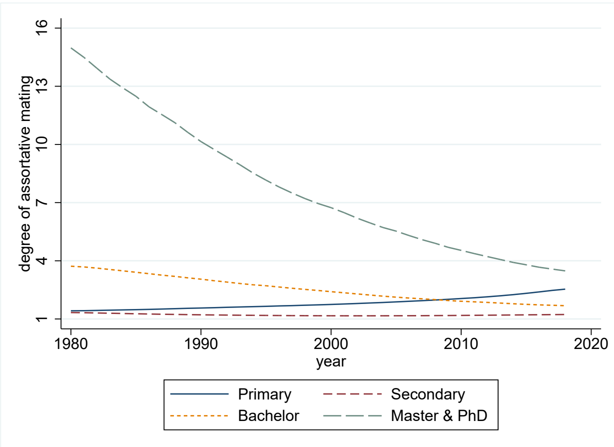
(b) Age Composition of Couples



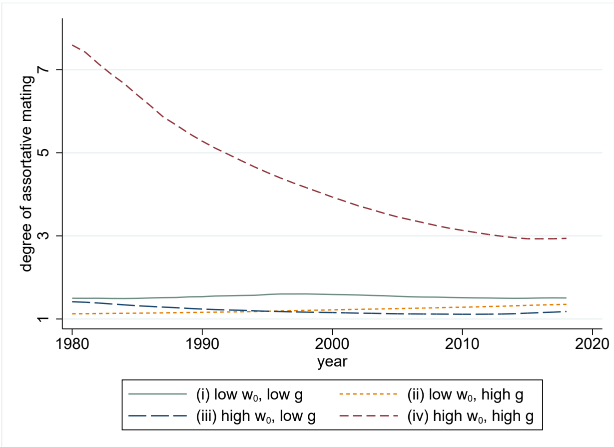
Note: Panel (a) reports the development in numbers of individuals by marital status. Panel (b) plots the age distribution of individuals who are either legally married or cohabiting. Panel (a) includes all individuals with an assigned educational-ambition type. Panel (b) includes all couples as defined in Section 2.

Figure A.2: Likelihood Indices

(a) Educational Level, 4 types



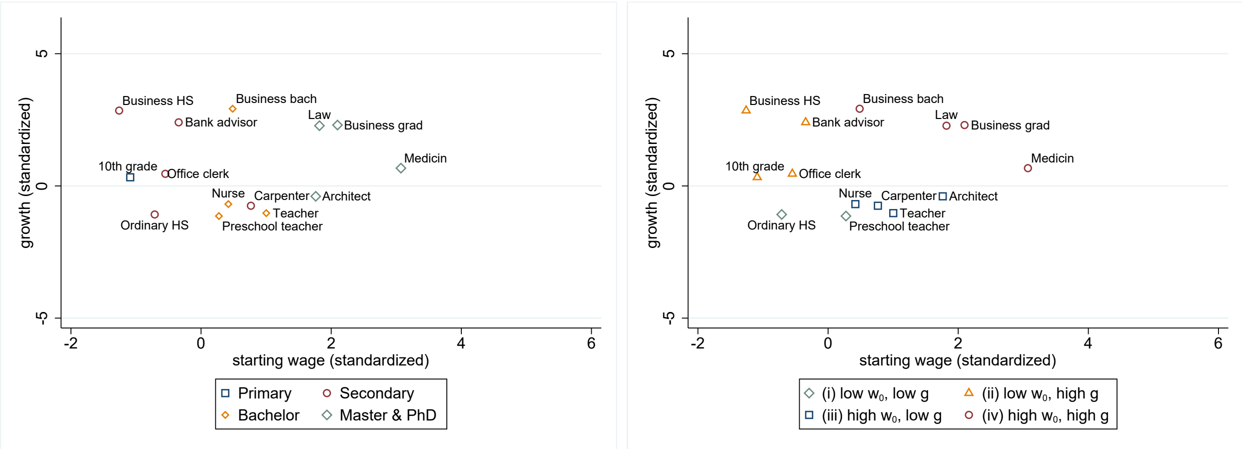
(b) Educational Ambition, 4 types



Note: Likelihood indices for assortatively matched couples cf. equation (1) for educational-level and educational-ambition categorizations. The sample includes all couples as defined in Section 2.

Figure A.3: Examples of Educational Programs

(a) Educational Level, 4 types, large programs (b) Educational Ambition, 4 types, large programs



Note: Each point in the figures represents the average wage growth (g) and starting wages (w_0) in educational programs with at least ten graduates observed by 2018. The symbols of points refer to the educational level or educational-ambition type associated with the program. Axes are standardized hourly starting wages and standardized hourly wage growth in the first ten years after graduation. The sample includes the full population as defined in Section 2.

Table A.1: Descriptive Statistics for Educational Ambition Types

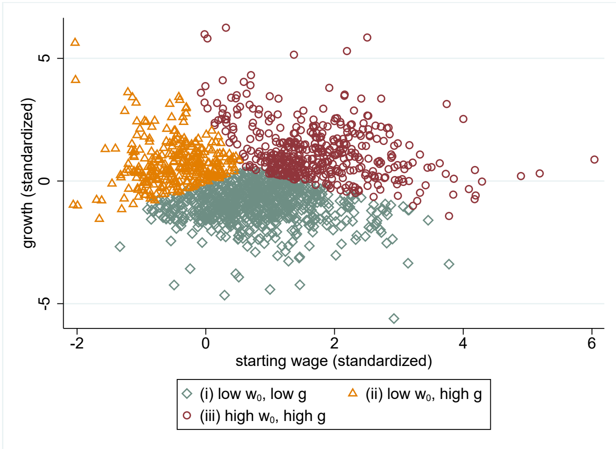
	(i) low w_0, low g	(ii) low w_0, high g	(iii) high w_0, low g	(iv) high w_0, high g	Population
Population share	20.2%	47.5%	22.7%	9.7%	100%
Female share	64.8%	56.0%	31.0%	33.4%	50%
Primary	8.3%	56.2%	0.5%	0.2%	28.5%
Secondary	66.2%	40.1%	57.3%	10.3%	46.4%
Bachelor	24.1%	3.1%	29.4%	30.3%	15.9%
Master & PhD	0.8%	0.5%	12.7%	59.0%	9.0%
Starting wage	4.841	4.728	5.015	5.181	4.860
	(0.0613)	(0.0488)	(0.0775)	(0.134)	(0.170)
Wage growth	0.0807	0.211	0.118	0.301	0.172
	(0.0339)	(0.0436)	(0.0756)	(0.0574)	(0.0862)
Annual Income	176,232.7	209,710.5	278,260.8	391,739.0	236,095.4
	(137,540.3)	(161,715.9)	(330,649.1)	(256,057.0)	(235,020.0)

Note: The four first columns correspond to each of the four educational-ambition types (Figure 1b). The final column reports descriptive statistics for the entire population of couples as defined in Section 2. Starting wages are measured in logs and wage growth are growth rates in hourly wages in the first ten years after graduation. Income is measured in DKK and deflated with base year 2000. Standard deviations in parentheses.

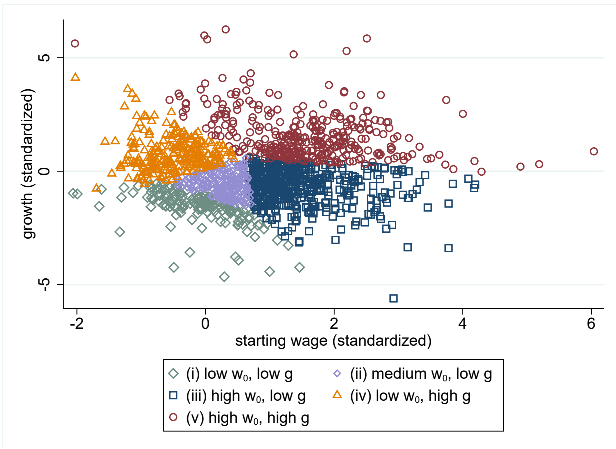
Figure A.4: Other Categorizations

(a) Educational Level, 3 types

(b) Educational Ambition, 3 types



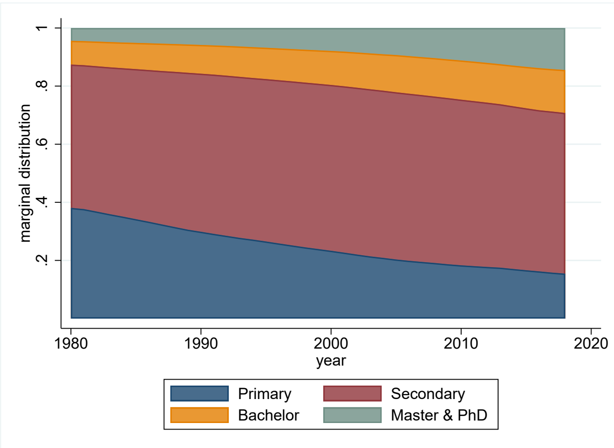
(c) Educational Ambition, 5 types



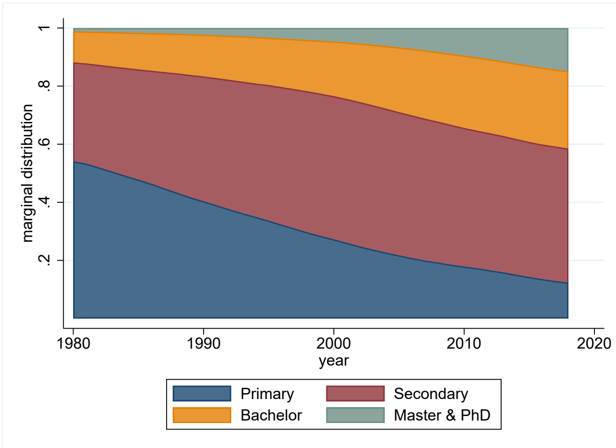
Note: Each point in the figures represents the average wage growth (g) and starting wages (w_0) in educational programs with at least ten graduates observed by 2018. The symbols of points refer to the educational level or educational-ambition type associated with the program. Axes are the standardized hourly starting wages and standardized hourly wage growth in the first ten years after graduation. The sample includes the full population as defined in Section 2.

Figure A.5: Marginal Type Distributions

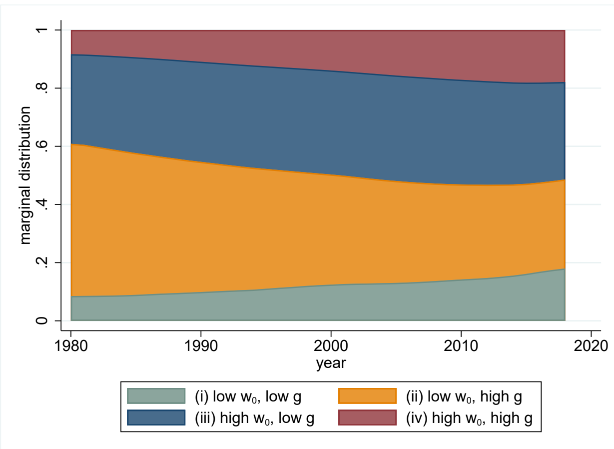
(a) Educational Level, 4 types, Men



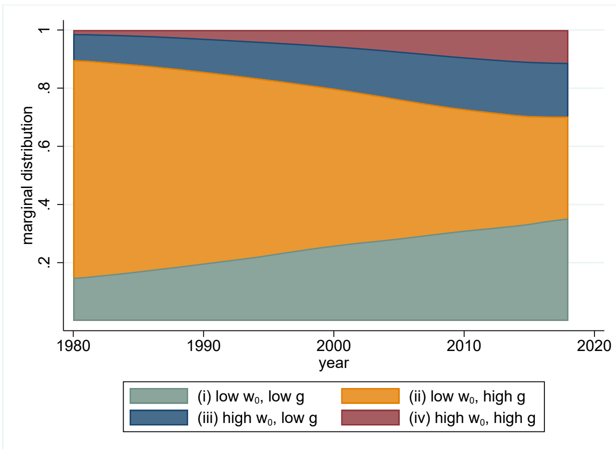
(b) Educational Level, 4 types, Women



(c) Educational Ambition, 4 types, Men



(d) Educational Ambition, 4 types, Women



Note: Marginal distributions for men and women over time by educational level and educational ambition. The sample includes all couples as defined in Section 2.