Marital Sorting and Inequality: How Educational Categorization Matters*

Frederik Almar† Benjamin Friedrich‡ Ana Reynoso§

Bastian Schulz¶ Rune Vejlin‖

June 6, 2023

Abstract

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.

*First draft: October 2022. We thank Basit Zafar, Elena Mattana, Rasmus Landersø, and seminar and workshop participants at Aarhus University, IAB Nuremberg, the DGPE network, and DAEiNA 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.

†Aarhus University, Department of Economics and Business Economics, almar@econ.au.dk

‡Northwestern University, Kellogg School of Management, benjamin.friedrich@kellogg.northwestern.edu

§University of Michigan, Department of Economics, areynoso@umich.edu

¶Corresponding author. Aarhus University, Department of Economics and Business Economics, the Dale T. Mortensen Centre, IZA, and CESifo, bastian.schulz@econ.au.dk

‖Aarhus University, Department of Economics and Business Economics and IZA, rvejlin@econ.au.dk

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*. The idea is that individuals observe the educational degree that potential mates (expect to) obtain, and this degree carries a signal about future time commitments to career and family. The signal arises from knowledge about the expected career paths and time flexibility associated with different programs, e.g., lawyers have more convex returns to hours than pharmacists (Goldin, 2014).

We quantify the signal using *average* labor market outcomes of program graduates. Previous work supports this idea by documenting that that education programs matter for marriage market matching (Nielsen and Svarer, 2009; Wiswall and Zafar, 2021; Kirkeboen, Leuven and Mogstad, 2022) and feature remarkable heterogeneity in initial conditions and lifetime labor market outcomes (Altonji, Kahn and Speer, 2014, 2016; Kirkeboen, Leuven and Mogstad, 2016). To calculate the labor market outcomes of graduates and capture heterogeneity across 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 programs and graduates based on 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 takes into account across-program heterogeneity in labor market outcomes, which is masked by the commonly-used level of education. Accounting for this heterogeneity is crucial to reveal changes in who marries whom over time, and furthers our understanding of how such changes contribute to trends in inequality. Our main findings are robust to using alternative measures of labor market outcomes (categorizations based on life-time earnings reveal similar trends). Furthermore, they are similar when aggregating programs by field of study, which is more readily available in other data sources. Thus, our analysis suggests that it is promising to consider richer type classifications than the level of education in future research on marital sorting.

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 assesses the sensitivity of our findings. Section 6 concludes.

2 Data

We use Danish register data that cover all residents between 1980 and 2018. Unique person IDs identify individuals across registers. The population register contains demographic variables and the person ID of the married or cohabiting partner (Statistics Denmark BEF, 1990–2018).[1]

For the measurement of between-household income inequality, we include individuals in the age range 19–60 who are either married to or cohabit with an individual of the opposite sex in the same age range. We follow the literature (Eika et al., 2019) and exclude one-person households to isolate the effect of marital sorting on between-household inequality. To compute household income, we add up yearly labor income from both regular employment and self-employment for both individuals based on the income register (Statistics Denmark IND, 1990–2018).

The population of (married or cohabiting) couples with both partners in the considered age range 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] For brevity, we refer to both types of couples as married in the remainder of the paper.

For the measurement of labor market outcomes by educational program, we use the education register in which four-digit codes (ISCED) uniquely identify around 1,800 educational programs in Denmark (Statistics Denmark UDDA, 1990–2018). We identify all program graduates in the age range 19-60 irrespective of marital status and use their *hourly wages* according to the employment register (Statistics Denmark IDAN, 1990–2018) to calculate labor market outcomes at the program level.[3] To abstract from the increasing mean and variance of wages, we regress log hourly wages and on year effects with 2000 as the base year and use the residuals.

[1]Legally, couples who are considered to be cohabiting (samlevende) largely have the same status as married couples. In the data, cohabiting couples are identified based on a number of criteria: two opposite-sex individuals who have a joint child and/or share an address without other adults, exhibit an age difference of less than 15 years, have no family relationship.

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

[3]Hourly wages are imputed from administrative information on labor income and hours worked in the employment register. Hours worked reflect contractual hours, i.e., part-time work is captured but not overtime. Before 2008, the hours are reported in four discrete bins. For further details, see Lund and Vejlin (2016).

3 Education and Marriage Market Sorting

3.1 Education-Based Marriage Market Types

The novel educational-ambition categorization we develop groups programs based on the labor market outcomes of graduates. We use the residualized hourly wages to measure average starting wages, w_0, and wage growth, g, for every educational program in Denmark. We include all individuals 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 average wage growth 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.

In contrast, common education-based categorizations group programs based on educational level, i.e., primary (compulsory schooling), secondary (high school, vocational degrees), and tertiary (higher) education. We further 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) and long tertiary programs (master/PhD programs with study times of five years or more).

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.

[4] That is, the expected wage growth g_i of a 1990 graduate is based on the observed wage growth trajectories of both previous and later graduates.

[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.

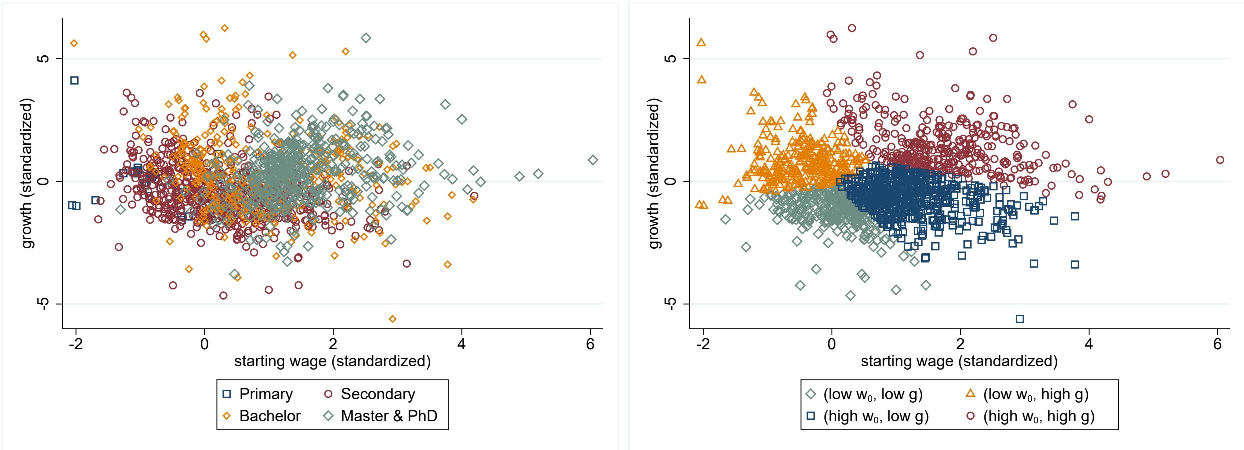
[6] We focus on the first 9-11 years because wage profiles stabilize later in one's career (Bhuller et al., 2017).

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

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) of educational programs with at least ten graduates in 2018. The symbols refer to the educational-level or educational-ambition category that a program belongs to. Axes are standardized starting wages and standardized wage growth in the first ten years after graduation. Section 2 explains how the sample and the underlying labor market outcome (residualized log hourly wages) are constructed. Section 3.1 explains how we calculate starting wages and growth.

Panel (b) shows how the k-means algorithm groups educational programs based on starting wages w_0 and growth g. By construction, programs within a cluster are similar in terms of (w_0, g). While the subsequent analysis of sorting does not require a rank-order of groups, the red cluster (circles), which includes programs with both high starting wages and high wage growth, can be seen as the top category. Programs in the gray (diamonds) cluster have the lowest starting wages and wage growth. In between, the orange cluster (triangles) includes programs with low starting wages but relatively high wage growth. The blue cluster (squares) has relatively low wage growth but high starting wages. Graduates from programs in the blue cluster have on average higher labor income, life-time income, and net wealth than graduates from the orange programs, but the differences are relatively small (see Table A.1). Moreover, individuals in the orange group become parents late and have relatively few children, which could be interpreted as a sign of career ambition.

To obtain a sense of which educational programs are included in the respective clusters, we locate 14 large[8] 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[9]—is,

[8]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 this sample, the examples cover 21% of all graduates.

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

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.[10] 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.

Additionally, we describe the four educational ambition categories in terms of population shares, gender, and education levels (see Table A.1). 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 to the orange category with similar low starting wages but higher wage growth. Here, 56% are females. Programs in the blue category with high starting wages but relatively low wage growth generate 22.7% of all graduates. The red top group with high starting wages and high wage growth is the smallest with 9.7% of graduates. About a third of graduates are female in the categories with high starting wages. In the low-starting-wage categories, the majority of individuals have primary or secondary education. The share of graduates with a tertiary degree is highest in the high-starting-wage categories. However, even in the top group, we see a sizable minority (10.3%) with secondary degrees.

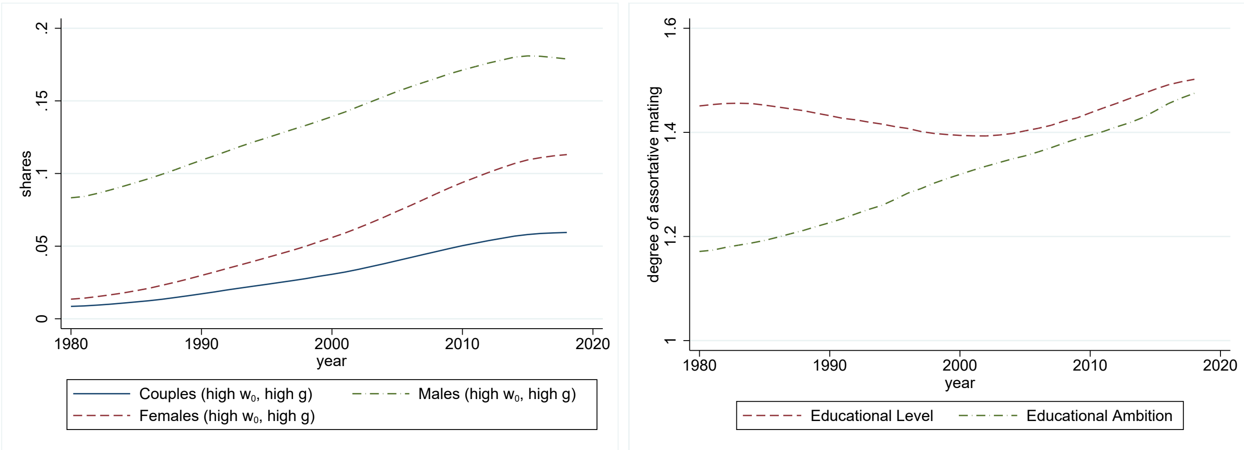
3.2 Measurement of Sorting

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 (a contingency table). Determining whether sorting changes over time has proven elusive because the marginal distributions of types in the marriage market change over time as well. Panel (a) of Figure 2 shows that the population share (solid blue line) of couples in which both spouses are highly ambitious has more than quintupled between 1980 and 2018, from around 1% to more than 6%. This observation alone, however, is not sufficient to conclude that marital sorting based on educational ambition has increased. The reason is that the marginal type distributions have changed as well. The share of highly-ambitious men increased from around 8% to just below 20% (green dash-dotted line). The share of highly-ambitious

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

Figure 2: The Measurement of Marital Sorting

(a) High-type couples and marginal distributions (b) Increasing sorting based on educational ambition



Note: Panel (a) shows shares of males and females in the top category of the educational-ambitions categorization, i.e., who graduated from educational programs associated with with high starting wages and high wage growth, along with the share of couples in which this is true for both spouses. Panel (b) shows the sorting measure $\mathcal{S}$ derived in Section 3.2 (equation 2) for educational-level types (red dashed line) and educational-ambition types (green dash-dotted line). Both categorizations use 4 types. The underlying educational-ambition categories are constructed using the k-means algorithm with standardized starting wages and standardized wage growth in the first ten years after graduation as inputs, see Section 3.1. Section 2 explains how the sample and the underlying labor market outcome (residualized log hourly wages) are constructed.

women very low in 1980 (around 2%) and increased to more than 10% in 2018. The increasing "supply" of highly-ambitious individuals in the marriage market mechanically increases the probability that two high-type spouses form a couple.

We follow Eika et al. (2019) and Chiappori et al. (2020), who take the changing marginal distributions into account by constructing weights for couples of 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:

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

Panel (b) of 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 hardly increased since 1980. Eika et al. (2019) find the same 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 finding that sorting is flat for educational-level types but increasing for educational-ambition types does not depend on the number of categories.[11]

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) 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 2b. In addition, the tendency to match assortatively decreases in the growing tertiary categories and increases in the shrinking primary education category. 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.

[11] Educational-level sorting is flat with three categories, see Panels (a)/(b) of Figure A.4. Educational-ambition sorting is increasing with either three or five categories, see Panels (c)–(f).

4 Marriage Market Sorting and Inequality

To study the link between marriage market sorting and inequality, we follow Eika et al. (2019) and construct a stochastic matching algorithm to re-match married individuals under different counterfactual scenarios. The algorithm samples pairs of potential spouses from the male and female marginal type distributions and forms new couples based on type-dependent matching probabilities p.[12] We keep either sorting, the income distribution, or the marginal type distributions fixed at their 1980 value to study how changes in these dimensions contributed to rising inequality. To measure between-household inequality, we use the Gini coefficient as an overall measure and zoom in on the upper and lower halves of the income distribution using percentile ratios.

We discuss three counterfactual scenarios: (i) fixed marriage market sorting at the 1980 level; (ii) fixed labor market returns to educational type at the 1980 level; and (iii) fixed educational-type composition at the 1980 level. 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.

(i) Fixed marriage market sorting

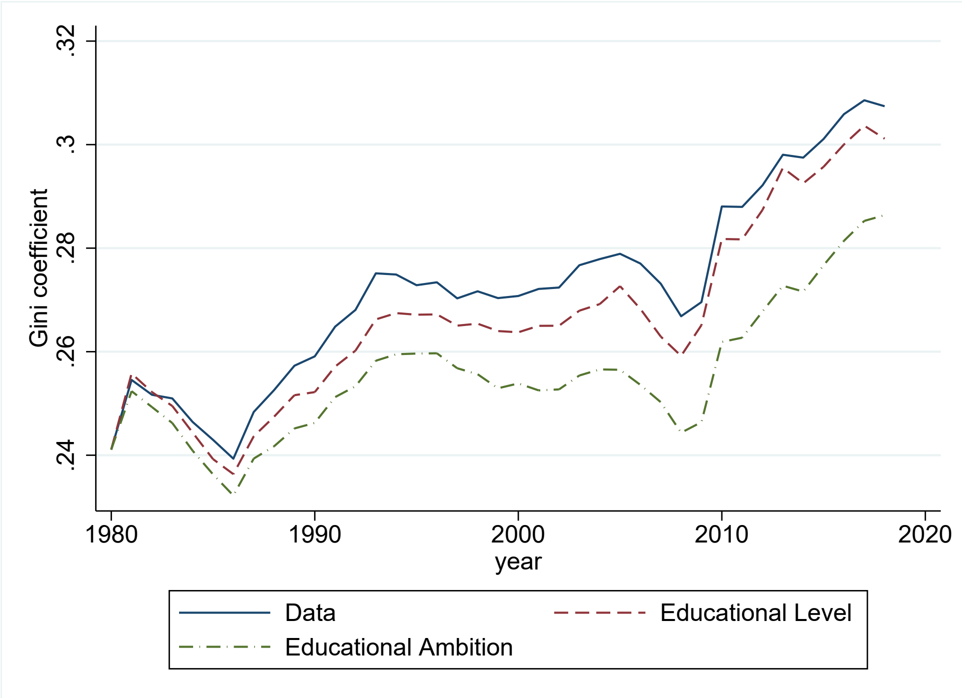
We construct scenario (i) by rematching couples based on the matching probabilities and marginal distributions from 1980. That is, we fix both the likelihood indices (1) and the marginal type distributions.[13] We construct counterfactual inequality measures for all years $\tau \in (1981, 1982, \ldots, 2018)$ 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. 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

[12] To calculate $p \in [0, 1]$, we divide the likelihood index $s(j, j')$ (equation 1) by the sum of indices across all potential partner types for both genders. This number may differ for men and women, so we take the average to compute the matching probability of a $(t_{i,m} = j, t_{i,f} = j')$ couple. To determine whether a match is formed, we draw from a binomial distribution with parameter p. We repeat the process until all individuals are matched with a new partner.

[13] Our goal is to keep the aggregate sorting measure fixed, and, as we have discussed in Section 3.2, changing marginal distributions contribute to aggregate sorting. We isolate the effect of changing marginal distributions on inequality in scenario (iii).

Figure 3: Growth in Educational-Ambition Sorting Amplified Inequality



Note: The plot shows the development of the Gini coefficient for the joint labor income of spouses in two-person households between 1980–2018 in the data (blue solid line), and in the counterfactual scenario (i) (fixed marriage market sorting) for both educational-level (red dashed line) and educational-ambition (green dash-dotted line) types. Section 4 explains how the counterfactuals are constructed. Section 3.1 explains how the underlying educational-level and educational-ambition categories are constructed. Section 2 explains how the sample and the underlying labor market outcome (joint annual labor income) are constructed.

by educational ambition amplifies the inequality trend, while flat sorting by educational level contributes relatively little.

In Table 1, we zoom in on the end points of Figure 3 and decompose the total change in household-level income inequality between 1980 and 2018. 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 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%.

Panel (i) shows our main result. With fixed sorting, the counterfactual 2018 Gini coefficient amounts to 57% of the true coefficient with educational-ambition types. That is, more than 40% of the increase since 1980 can be explained with more pronounced sorting based on educational ambition. Using the common educational-level categorization, we do not arrive at the same conclusion. The corresponding counterfactual inequality measure is 91%, which suggests that sorting had a limited impact on the inequality trend (consistent with Eika et al., 2019).

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%
	Δ_{Gini}	$\frac{\Delta_{Gini}}{\Delta_{Data}}$	$\Delta_{P90/P50}$	$\frac{\Delta_{P90/P50}}{\Delta_{Data}}$	$\Delta_{P50/P10}$	$\frac{\Delta_{P50/P10}}{\Delta_{Data}}$
(i) Fixed sorting						
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: The table shows changes in inequality between 1980 and 2018 in the data and for each of the counterfactual scenarios constructed and discussed in Section 4. 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., $\Delta_{Gini}/\Delta_{Data}$. Section 3.1 explains how the underlying educational-level and educational-ambition categories are constructed. Section 2 explains how the sample and the underlying labor market outcome (joint annual labor income) are constructed.

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 54% of the true change, which suggests that increasing positive sorting by educational ambition did amplify 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 suggest an amplification of the inequality trend, but the amplification is more than twice as strong with educational-ambition types.

(ii) Fixed labor market returns to educational type

In this scenario, we analyze how income inequality would have developed had the income distribution across types remained unchanged. To this end, we introduce a household reweighting factor to construct the counterfactual 2018 household income distribution with 1980 returns:

$$\widehat{F}(y|\tau_y = 1980, \tau_x = 2018, \tau_p = 2018) = \int F_{Y|X}(y|x, \tau_y = 1980)\psi_y dF(x|\tau_x = 1980), \tag{4}$$

where subscript y denotes household income, subscript x the couple-type combination ($t_{i,m} = j, t_{i,f} = j'$), and subscript p the matching probabilities in the respective calendar year.[14] The reweighting factor is

$$\widehat{\psi_y} = \frac{P(\tau_x = 2018|x, \tau_p = 2018)}{P(\tau_x = 1980|x, \tau_p = 2018)} \frac{P(\tau_x = 1980)}{P(\tau_x = 2018)}. \tag{5}$$

Building on Fortin et al. (2011), we calculate $\widehat{\psi_y}$ by first rematching 1980 couples based on 2018 matching probabilities. Next, we compute the conditional probabilities in the numerator and denominator of the first term on the RHS. Intuitively, couple combinations that are more common in 2018 than in 1980 get a weight greater than one in the counterfactual income distribution (and vice versa).

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 premia 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 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 have decreased with both categorizations.

[14] The (non-counterfactual) 2018 income distribution is defined as $F(y|\tau_y = 2018, \tau_x = 2018, \tau_p = 2018) = \int F_{Y|X}(y|x, \tau_y = 2018) dF(x|\tau_x = 2018)$. The counterfactual income distribution is defined as $\widehat{F}(y|\tau_y = 1980, \tau_x = 2018, \tau_p = 2018) = \int F_{Y|X}(y|x, \tau_y = 1980) dF(x|\tau_x = 2018)$. This is not observable. However, if we insert $\frac{dF(x|\tau_x=1980)}{dF(x|\tau_x=1980)}$ and rearrange, we obtain (4) with $\psi_y = \frac{dF(x|\tau_x=2018)}{dF(x|\tau_x=1980)}$, which can be estimated.

(iii) Fixed composition in terms of educational types

In the last counterfactual scenario, we fix the marginal distributions. The approach is similar to (ii). We reweight households in the 2018 income distribution based on changes in the marginal distributions of $t_{i,m}$ and $t_{i,f}$. In this case, the reweighting factor is $\widehat{\psi_x} = (\widehat{\psi_y})^{-1}$.

First, we keep the type distributions for both genders fixed at the 1980 level, see Panel (iii) in Table 1. We then repeat the exercise keeping either the male or female marginal type distribution fixed, see Panels (iiia) and (iiib). The marginal distributions shifted such that the numbers of individuals in the top categories increased, and this change is more pronounced for women.[15] That is, we have more men and women who graduate with tertiary degrees and/or from ambitious educational programs in 2018 compared to 1980.

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 actual 2018 value. The mitigating effect manifests itself mainly in the lower half of the income distribution. The 50/10 percentile ratio in column (c) 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. 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 educational-level categories are not. 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.

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 drive the mitigating effects we find. If we instead keep the male marginal distributions fixed (iiia), the mitigating effects on inequality are considerably smaller.

In summary, the importance of distinguishing between the two categorizations becomes evident from the distinct effects that changing marginal distributions had on between-household

[15] The share of men (women) with long-cycle tertiary education increased by a factor of 3 (13) between 1980 and 2018. For educational-ambition types, the share of men (women) in the top category doubled (increased eight-fold), see Figure A.5.

inequality. While the move of females into tertiary education overall had a considerable mitigating effect, their entry into (ambitious) high-wage/high-growth programs amplified inequality in the upper half of the distribution, offsetting the mitigating effect in the lower half. For men, the shifts in both marginal distributions amplified inequality overall, and the difference between the two classifications is smaller.

5 Sensitivity Analysis

In this section, we assess the sensitivity of our main results. First, with respect to the properties of the matching algorithm that we use to produce the counterfactuals. Second, with respect to the granularity of the data and the labor market outcomes that we use to categorize educational programs and their graduates.

(i) Matching algorithm

The matching algorithm is one-dimensional, i.e., it takes only the education-based types into account. Thus, we essentially assume random matching conditional on type. If other dimensions correlate with the labor market outcomes that we use to categorize programs, sorting within cells could arise and bias the counterfactual inequality measures. To investigate this, we use the algorithm to rematch couples randomly ($p = 0.5$) in 2018 within couple-type-combination cells and check how well the empirical inequality measures are reproduced. Table A.2 shows that the level of inequality implied by the algorithmic re-matching fits the data well, irrespective of whether we use educational-level or educational-ambition types. The fit is nearly perfect for the 90/50 income percentile ratio and slightly worse for the 50/10 ratio. Overall, the reproduced Gini coefficients correspond to 95–96% of the values in the data. We conclude that the one-dimensional matching algorithm produces reliable counterfactual marriage market allocations.

(ii) Alternative categorizations

To show that our main results do not depend on a specific way to summarize heterogeneity across educational programs, we redo the fixed-sorting scenario (i) with alternative educational-ambition categorizations. Table 2 shows the resulting sorting trends and inequality contributions. The first row repeats results for the benchmark (starting wages and wage growth, program level). Sorting increased by about 26% and explains more than 40% of increasing inequality.

In the second row, we group fields of study by level instead of using educational programs. Specifically, we consider a classification in which we subdivide educational levels by field of

Table 2: Scenario (i)—Fixed Sorting—with Alternative Categorizations

	N (1,000s)		Sorting			Gini, data		Gini, (i)	$\frac{\Delta_{Gini,(i)}}{\Delta_{Gini,data}}$
	1980	2018	1980	2018	Change	1980	2018	2018	
Benchmark (w_0, g)	1,758	1,653	1.17	1.48	25.9%	0.241	0.307	0.279	57%
Field of study (levels × fields)	1,854	1,630	1.19	1.45	21.8%	0.243	0.304	0.279	60%
Life-time earnings (hourly wages)	700	774	1.28	1.53	19.6%	0.224	0.295	0.263	56%
Life-time earnings (income)	700	774	1.25	1.52	21.4%	0.224	0.295	0.262	54%

Note: Each row shows our analysis of sorting and inequality for different definitions of marriage market types. *Benchmark* refers to our definition based on starting wages (w_0) and wage growth (g) defined in Section 3.1. *Field of study* refers to an alternative definition using a subdivision of educational levels into fields of studies to group education based also on w_0 and g, as explained in this section. *Life-time earnings (hourly wages)* and *Life-time earnings (income)* use a definition of marriage market types that is based on grouping educational programs by life-time hourly wages or individual labor income, respectively, as explained in footnote 17. Columns N display the number of observations in each case in thousands of individuals. Columns *Sorting* show our measure of marriage market sorting $\mathcal{S}$ derived in Section 3.2. *Gini, data* corresponds to the observed Gini coefficient in each case. *Gini, (i)* refers to the counterfactual Gini coefficient in scenario (i), i.e., had sorting had stayed fixed at its 1980 level in each case. $\Delta_{Gini,(i)}/\Delta_{Gini,data}$ shows the fraction of the observed change in Gini that is captured by the change in the counterfactual scenario. Section 2 explains how the sample and the underlying labor market outcome (joint annual labor income) are constructed.

study and cluster the resulting 48 categories using k-means based on average starting wages and wage growth (as in the benchmark). The results hardly change. Sorting increased by about 22% and explains 40% of the increasing inequality between households.

In the last two rows, we use life-time earnings of graduates at the program-level instead of starting wages and growth. In this case the k-means clustering is based on a single standardized variable (instead of two in the benchmark).[16] We compute life-time earnings in two different ways: based on hourly wages (abstracts from the intensive margin, row 3) and annual income (takes intensive-margin choices into account, row 4).[17] Again, we see consistent results. Based on both versions of life-time earnings, we find that sorting has increased by 20–21% and explains more than 40% of the increasing inequality between households.

Taken together, the alternative categorizations show that our main results do not depend on specific labor market outcomes or the granularity of educational-program data. Education-based categorizations of marriage market types expose trends in sorting and its role for between-household inequality whenever heterogeneity in labor market outcomes *within* educational levels

[16] A potential concern of comparing categorizations based on two vs. one clustering variable is that groups may differ in size. Standardizing the variables alleviates this concern to some extent. In our case, we end up with similar group sizes across categories.

[17] Specifically, we compute life-time earnings as the sum of deflated hourly wages or income over 30 years after graduation. The deflation is done at the program level in a regression of log hourly wages or annual income on year dummies (base year 2000) and dummies for years since graduation. This ensures the comparability of program averages. Note that the number of observations falls because we only include individuals that we observe for at least 30 years.

is taken into account. However, it is important to consider that categorizations based on life-time outcomes may confound endogenous decisions post marriage. Therefore, our preferred categorization relies on average early-career labor market outcomes of graduates.[18]

6 Conclusions

We provide new insights into the relationship between education-based marital sorting and between-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 between-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.

Our baseline results use starting wages and wage growth of graduates to categorize educational programs. In a sensitivity analysis, we show that our conclusions hold with alternative labor market outcomes (life-time earnings) and do not depend on using granular program-level data. Specifically, the results do not change if we interact educational levels with fields of study. This suggests that most of the heterogeneity in labor market outcomes within educational levels

[18] Another variable considered in the marital sorting literature is parental wealth (Fagereng et al., 2022). We find that parental wealth sorting in Denmark hardly increases over time (in line with Wagner et al., 2020).

is captured by the field of study.[19] This is promising for future research, because combining the level of education, the field of study, and labor market outcomes—variables commonly available in, e.g., the US Panel Study of Income Dynamics and the American Community Survey—can allow researchers in countries without detailed educational data to create similar classifications to ours and gain valuable new insights into the link between marital sorting and inequality.

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.

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

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.

Fagereng, Andreas, Luigi Guiso, and Luigi Pistaferri (2022) "Assortative Mating and Wealth Inequality," Working Paper 29903, National Bureau of Economic Research.

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.

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.

——— (2022) "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.

Lund, Christian Giødesen and Rune Vejlin (2016) "Documenting and Improving the Hourly Wage Measure in the Danish IDA Database," *Nationaløkonomisk tidsskrift*, Vol. 2016, pp. 1–35.

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.

Wagner, Sander, Diederik Boertien, and Mette Gørtz (2020) "The Wealth of Parents: Trends Over Time in Assortative Mating Based on Parental Wealth," *Demography*, Vol. 57, pp. 1809–1831.

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

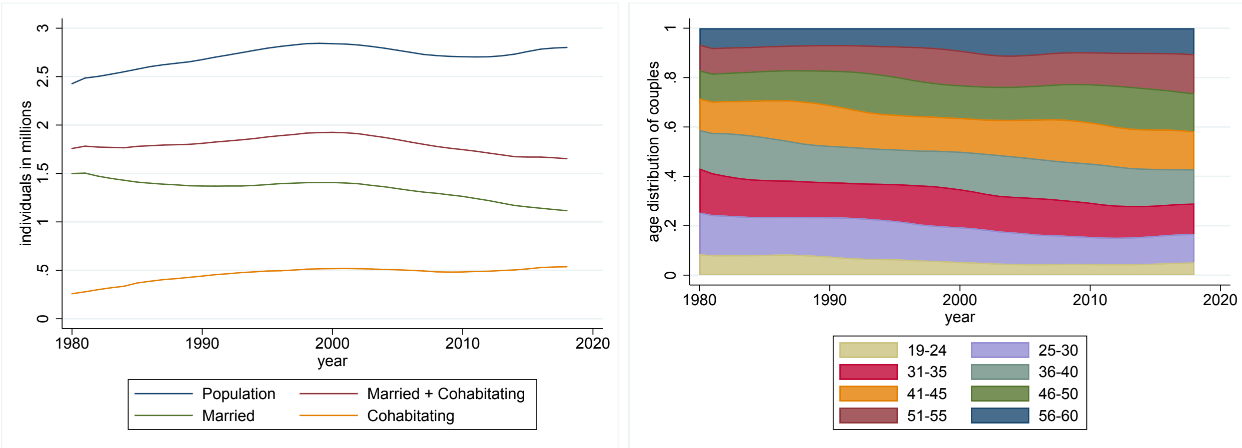
Marital Sorting and Inequality: How Educational Categorization Matters

Frederik Almar, Benjamin Friedrich, Ana Reynoso, Bastian Schulz and Rune Vejlin

A Complementary Exhibits

Figure A.1: Marriage, Cohabitation, Age Composition

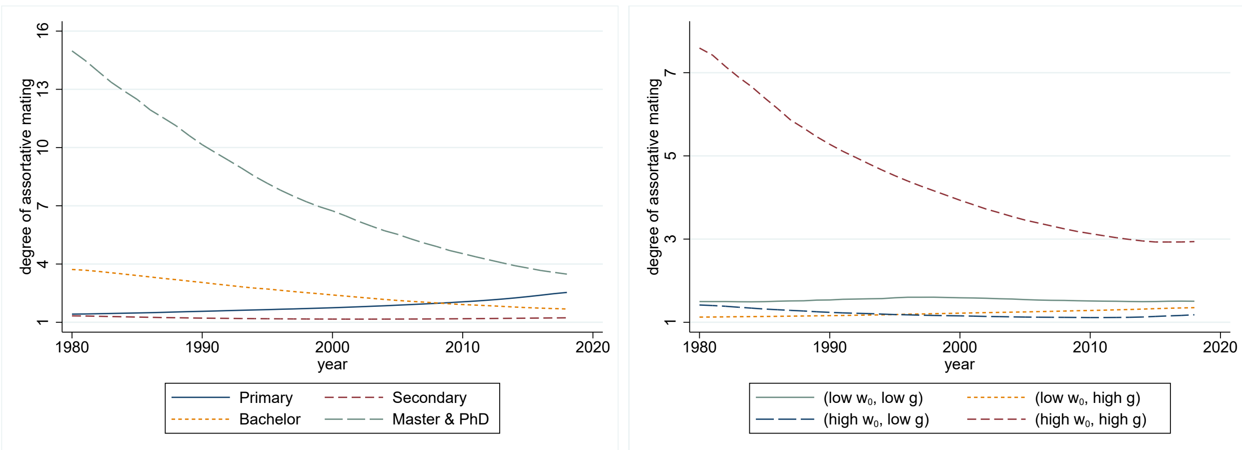
(a) Stocks of Couples (b) Age Composition of Couples



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. Section 2 explains how the sample, educational levels, and the labor market outcome (residualized log hourly wages) underlying the educational-ambition types are constructed.

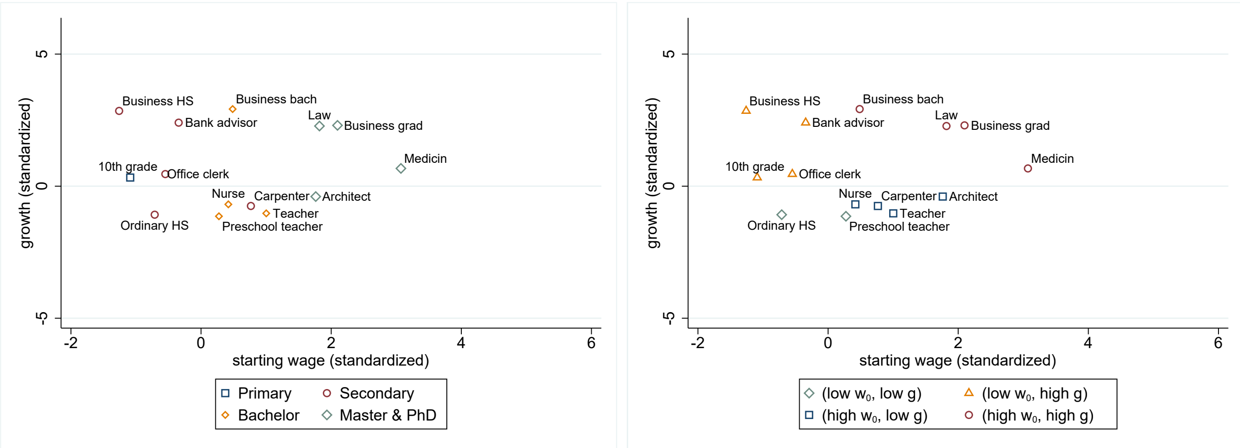
Table A.1: Descriptive Statistics for Educational Ambition Types

Category (w_0, g)	(low, low)	(low, high)	(high, low)	(high, high)	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)
Labor 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)
Life-time Income	4,771,608	5,239,822	6,315,139	11,389,772	6,124,398
	(1,233,0216)	(3,226,927)	(9,454,225)	(3,306,838)	(7,336,409)
Net wealth	51,364	117,017	147,635	250,292	123,600
	(925,522)	(1,205,725)	(1,118,021)	(2,963,253)	(1,413,112)
Age at first child	29.63	31.90	31.37	31.77	31.21
	(6.226)	(8.355)	(6.325)	(5.002)	(7.016)
Number of children	2.034	1.680	1.894	2.072	1.843
	(5.697)	(4.460)	(4.159)	(6.956)	(4.979)

Note: The four first columns report averages of individual-level descriptive statistics for each of the four educational-ambition categories identified in Section 3.1. The final column reports the same 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. Labor income is measured in DKK and deflated with base year 2000. Life-time income is measured in DKK and computed as the sum of exponentiated deflated annual income over 30 years after graduation. The deflation is done at the program level in a regression of log annual income on year dummies (base year 2000) and dummies for years since graduation. Net wealth is computed as the average of the individual-level averages over all years that a given individual is in the sample. Deflated with base year 2000. Age at first child is computed based on the Danish birth register. The number of children refers to children living in the household. It is the average of the maximum at individual-level taken over all years that a given individual is in the sample. Standard deviations in parentheses.

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. Section 2 explains how the sample, educational levels, and the labor market outcome (residualized log hourly wages) underlying the educational-ambition types are constructed.

Table A.2: Matching Algorithm Performance

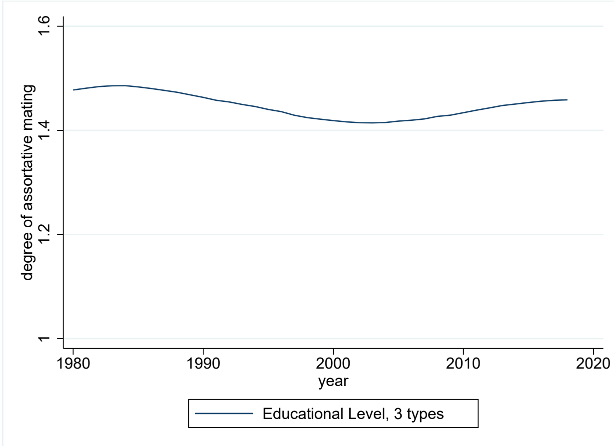
	(a) Gini		(b) $P90/P50$		(c) $P50/P10$	
Data (2018)	0.307	100%	1.688	100%	2.518	100%
Within-cell reshuffling						
Educational Level	0.291	95%	1.675	99%	2.178	87%
Educational Ambition	0.295	96%	1.690	100%	2.189	87%

Note: The table reports results of the sensitivity analysis for the matching algorithm discussed in Section 5. We rematch couples randomly in 2018 within couple-type-combination cell and check how well the empirical inequality measures are reproduced. 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 measures in the data for 2018. Section 2 explains how the underlying sample of couples is constructed.

Figure A.4: Alternative Numbers of Categories and Sorting Trends

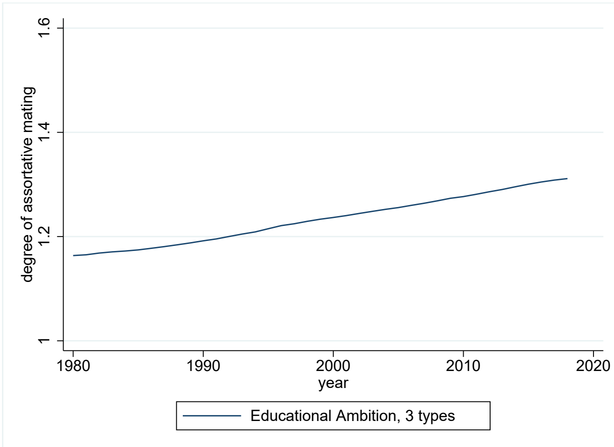
(a) Educational Level, Categorization with 3 types

(b) Educational Level, Sorting Trend with 3 types

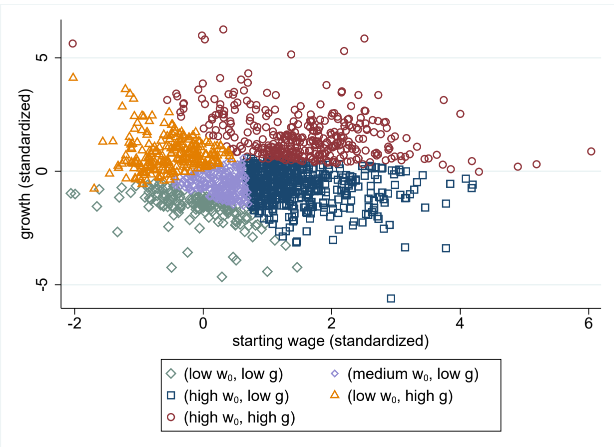


(c) Educational Ambition, Categorization with 3 types

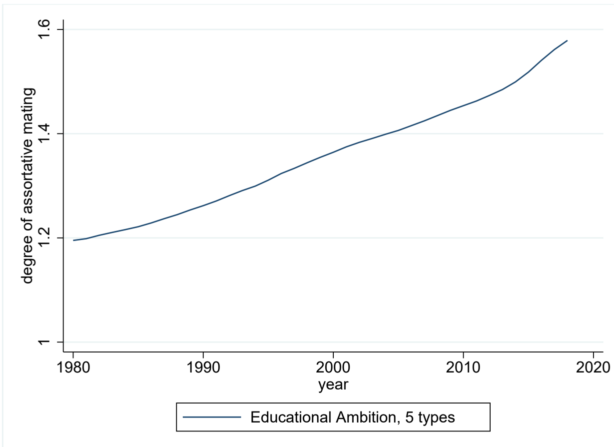
(d) Educational Ambition, Sorting Trend with 3 types



(e) Educational Ambition, Categorization with 5 types

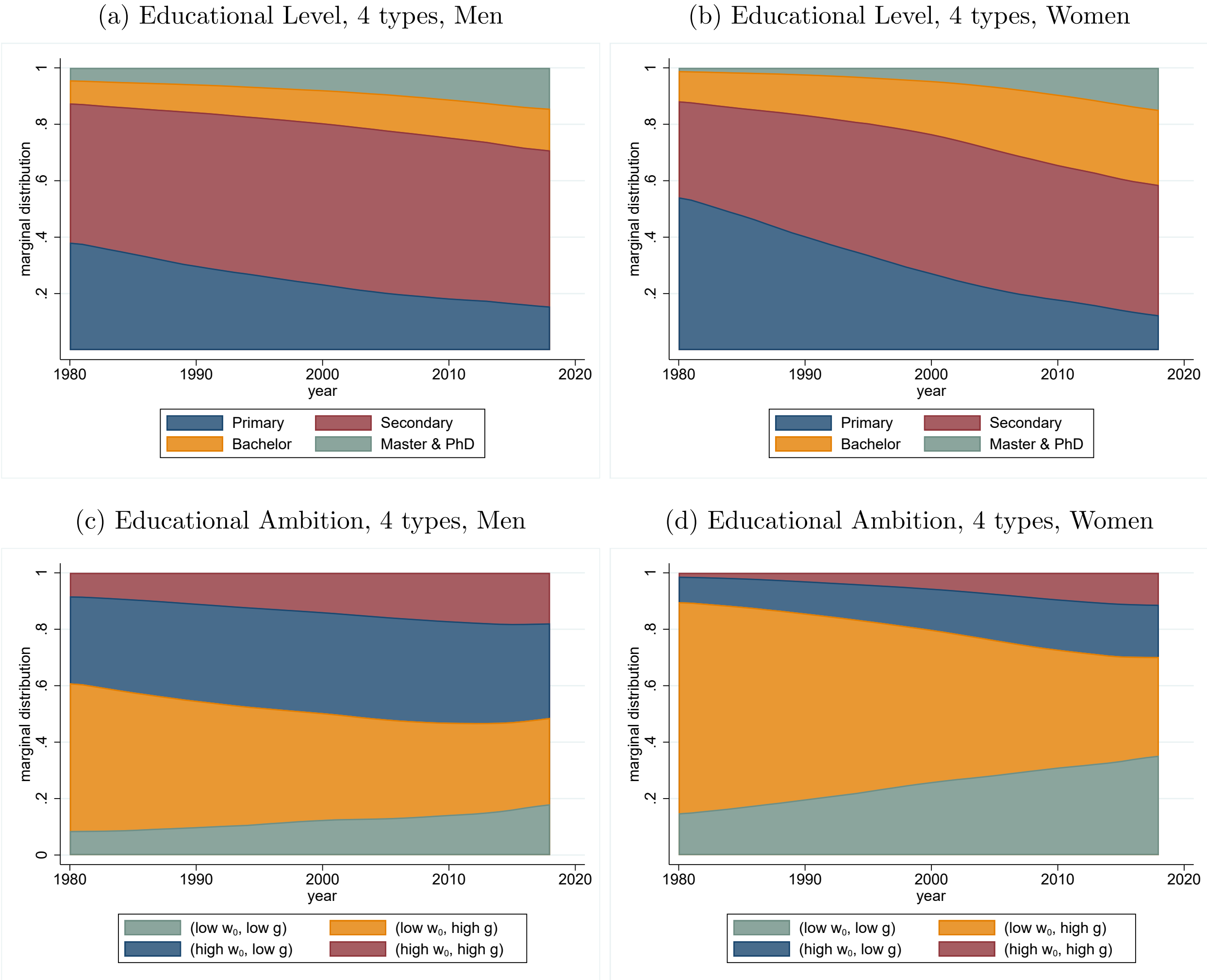


(f) Educational Ambition, Sorting Trend with 5 types



Note: Panels (a), (c), and (e) are alternative versions of Figure 1 in which we vary the number of categories. 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. Panels (b), (d), and (f) show the sorting measure $\mathcal{S}$ derived in Section 3.2 (equation 2) for different numbers of categories. Section 3.1 explains how educational-level and educational-ambition categories are constructed. Section 2 explains how the sample, educational levels, and the labor market outcome (residualized log hourly wages) underlying the educational-ambition types are constructed.

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. Section 2 explains how the sample, educational levels, and the labor market outcome (residualized log hourly wages) underlying the educational-ambition types are constructed.