Marital Sorting by Education and Career Ambition*

Frederik Almar[†] Benjamin Friedrich[‡] Ana Reynoso[§]

Bastian Schulz[¶] Rune Vejlin[‖]

October 23, 2022

— Work in progress —

Abstract

We use detailed data from Danish education registers to measure the degree of educational assortative mating and scrutinize the effect of marriage market sorting on income inequality. We define marriage market types based education in various ways. Our main contribution is to categorize types based on "ambition", a novel classification of educational programs that considers starting wages and wage growth of peers, besides more traditional categorizations such as the level of education, years of schooling, and field of study. Based on a decomposition analysis and various counterfactual scenarios, we find that conclusions about which factors contribute to the distribution of household income and its trend greatly depend on the way we map education ("ambition", level, years of schooling or field) into marriage market types. Based on the ambition types, we find that increased sorting has amplified the increase in household income inequality. Moreover, increasing returns and selection into highly ambitious educational programs also contribute to the increasing income inequality.

*First draft: October 2022. 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

[¶]Aarhus University, Department of Economics and Business Economics, bastian.schulz@econ.au.dk

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

1 Introduction

There is an abundance of correlational evidence for homophily in marriage markets, i.e., the tendency that individuals marry their likes in terms of, e.g., age, education, physical attractiveness, or wealth (Chiappori et al., 2012; Lee and McKinnish, 2018; Fagereng et al., 2022). The analysis of marital sorting based on the educational attainment of (potential) spouses has emerged as the standard approach in the literature. For example, Persson (2020), Reynoso (2022), and Anderberg et al. (2022) study the effects of social policy, divorce law, and schooling reforms on the marriage market through the lens of education-based sorting.

The relationship between increasing marriage market sorting and household income inequality has received particular attention (Kremer, 1997; Fernández and Rogerson, 2001; Greenwood et al., 2014). Over the last couple of decades, both inequality and the correlation between spousal attributes in the data have increased. As income increases in education, more educational sorting may lead to increasing income differences across households with different levels of education. However, the marginal distributions of educational attainment for men and (especially) women have shifted over time, and this needs to be taken into account when sorting measures are compared across time periods.

Greenwood et al. (2016) find that changing marital decisions account for almost 20% of the rise in income inequality in the US between 1960 and 2005 (holding educational choices fixed). Consistently, Eika et al. (2019) argue that increasing returns to education, not marital sorting, are the primary source of increasing inequality. The effect of increased education-based sorting on inequality is limited because observing more couples in which both spouses have the same level of education primarily reflects shifting marginal distributions (see also Breen and Salazar, 2011; Breen and Andersen, 2012; Chiappori et al., 2020a). What existing studies have in common is that they rely on relatively coarse educational categories to measure sorting.[1] We argue that the categorization of types is not an innocuous choice when studying the link between sorting and inequality.

In this paper, we develop a novel education-based categorization of marriage-market types. We call them career-ambition types. These types are based on the average labor market outcomes that specific educational programs are associated with. To construct the types, we merge detailed Danish education registers with labor market outcomes for all program graduates and compute continuous starting wages and wage growth trajectories at the level of the educational program. For example, an individual is classified as ambitious if it graduates from

[1]For example, Greenwood et al. (2016) use two (non-college, college) and Eika et al. (2019) four categories (no high school, high school, some college, college).

an educational program that promises high starting wages and high wage growth. Based on these ambition types, we find increasing marital sorting and a significant contribution to rising income inequality.

We start out by assessing how different educational categorizations affect conclusions about trends in education-based sorting. We compare a number of common categorizations to our ambition types: the level (primary/secondary/tertiary) of education, the length (years of schooling), and the field of study within tertiary education (e.g., business, engineering, humanities, etc.). Based on a weighted sum of likelihood indices, which is a well-known measure of sorting (Eika et al., 2019; Chiappori et al., 2020b), we find that conclusions about the extent and trend of education-based sorting differ significantly across these categorizations. Specifically, the level of sorting is higher for length of education and even higher for fields of study compared to the level of education categorization, and although trends are mostly flat we find a slight increase in sorting by length of education and a slight decrease for the two other categorizations. In contrast the new proposed categorization based on ambition results in markedly increasing marital sorting patterns. The fact that findings differ across the different classifications can be explained by the vast heterogeneity that is hidden within broad educational categories (Kirkeboen et al., 2016; Eika et al., 2019).

The career ambition type takes this heterogeneity within and across educational categories into account. Based on starting wages and wage growth, we cluster educational programs into four categories using the k-means algorithm. The characteristics of the four types and their relation to common educational categories can be summarized as follows: i) medium-high starting wages and low growth comprising around 33% of the population with mostly secondary but also tertiary degrees (for example, nurses, teachers, and carpenters); ii) high starting wages and medium-high growth comprising around 11% of the population with mostly tertiary but also secondary degrees (for example, graduates in business, law, and medicine); iii) medium-low starting wages and high growth comprising around 6% of the population with mostly secondary education (for example, business high school and bank advisors); iv) low starting wages and medium-low growth comprising around 50% of the population with mostly primary but also secondary degrees (for example, compulsory schooling, high school graduates, and office clerks). While category (ii) could be interpreted as the top group, no clear ranking emerges among the other three.

The idea behind the career-ambition categorization is that the educational program an individual graduates from is a signal in the marriage market (Wiswall and Zafar, 2021). The flip side of high wage growth could be a high level of working hours and limited flexibility

(Goldin, 2014) or lower fertility in expectation.[2] Thus, the career ambition of the (potential) spouse can be thought of as a predictor for future time inputs to home production. This is a channel through which career ambition may affect marriage formation.

Using our novel categorization, we contribute to the discussion about the link between education-based marriage market sorting and increasing inequality. Based on a counterfactual re-weighting exercise (DiNardo et al., 1996; Eika et al., 2019), we find that the contribution of ambition sorting to income inequality has been growing over time. This stands in contrast to the conclusion of, e.g., Eika et al. (2019), who use standard education categories and find that the contribution of increased education-based sorting to growing inequality is limited.

The remainder of the paper is organized as follows: Section 2 introduces our data and the strategies we use to measure education-based types and marital sorting. Section 3 discusses differences between common educational categories and ambition types and how these differences influence the measurement of marital sorting. Section 4 presents the counterfactual analysis. Section 5 concludes.

2 Measurement

2.1 Data

We use Danish register data provided by Statistics Denmark for our analysis. Our data contains detailed yearly observations for all residents between 1980 and 2018. Unique person IDs identify individuals across registers. The data cover demographic variables, the person ID of the (married or cohabiting) partner, education variables, earnings, wages, and labor market experience. 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.[3] We build our analysis samples based on the full population, which corresponds to 118,228,932 person-years for all residents in the age range 19–60 in the years 1980-2018. The population of couples with both partners in the age range 19–60 consists of 70,233,782 observations and an average of 1,800,866 individuals per year. There is an upward (downward) trend in cohabitation (legal marriage) but the combined stock of couples is relatively stable over time.[4] We

[2] Historically, cross-sectional correlations between income and fertility have been negative in many countries (Jones et al., 2008), although this relationship has become weaker in recent years (Doepke et al., 2022).

[3] Cohabitation is not a legal status. 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

[4] Figure A.1a in the Appendix depicts the evolution of the stocks of different couple types and their age composition during our period of interest.

calculate yearly household income by adding the income of spouses, which includes both wages and income from self-employment.

2.2 Measuring Sorting

From an empirical perspective, positive assortative matching (PAM) is reflected by a positive association of spousal types in the cross section. This association can be measured based on correlation coefficients, distances, or the frequency distribution of couples' types (Fernández and Rogerson, 2001; Greenwood et al., 2014, 2016; Abbott et al., 2019). Determining whether sorting changes over time has proven elusive because the marginal distributions of types in the marriage market change over time as well (Eika et al., 2019; Chiappori et al., 2020b, 2021).

Based on educational types, Greenwood et al. (2016) study the frequency distribution of type-combinations among couples in 1960 and 2005 and argue that sorting has increased. In 2005, there are more couples in which both spouses have the same type. To quantify this change, they divide the sum of the diagonal elements (trace) of the matrix formed by the observed contingency table by the trace of a (counterfactual) matrix under random matching.[5] This ratio is above one, which is indicative of PAM, and it has increased over time.

Eika et al. (2019) highlight the importance of the changing marginal distributions of (educational) types. They propose a weighted version of the frequency-based measure in Greenwood et al. (2016). Specifically, they weight the sum along the diagonal of the contingency table by the share of couples in the cells along the diagonal. These weights change over time as the marginal distributions of male and female types change. We further discuss the link between the marginal distributions and the weights below. The resulting measure of sorting is a weighted sum of ratios (likelihood indices). These ratios capture the likelihood relative to random matching of observing a "sorted" couple, i.e., a couple in which both spouses are of the same type, and the weighting makes the yearly sorting measure comparable over time.[6]

We follow the Eika et al. (2019) approach. Assume every individual has a one-dimensional type t_g, where $g \in \{m, f\}$ indexes males and females. Let the type be a categorical variable T_g^I with $I \in \{1, \ldots, N\}$ and the number of categories N. For example, these categories may represent levels of education (primary, secondary, tertiary, i.e., $N = 3$). For each category, the likelihood index relates the observed frequency of couples in which both spouses have the same type (the numerator) to the expected frequency under random matching (the denominator),

[5] In the latter case, each matrix entry is simply the product of the marginal probabilities of observing the respective male/female types in the data.

[6] Chiappori et al. (2020b, 2021) discuss formal criteria that measures of marital sorting should fulfill. The measure of Eika et al. (2019), which we describe formally in the following, meets these criteria.

which is given by the product of the marginals:

$$s\left(t_f, t_m\right) = \frac{P\left(T_f = t_f, T_m = t_m\right)}{P\left(T_f = t_f\right) P\left(T_m = t_m\right)}. \tag{1}$$

By summing across categories with the appropriate weights, we get the aggregate sorting measure

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

where, following Chiappori et al. (2020b), the weights are defined as a convex combination of the male/female marginal distribution for category I:

$$w_i = \lambda P\left(T_m = i\right) + (1 - \lambda) P\left(T_f = i\right), \tag{3}$$

where the first (second) term is the contribution of the male (female) marginal distribution to the weight and λ is in the unit interval. In the following, we discuss different education-based categorizations, for which we compute and compare the aggregate sorting measure $\mathcal{S}$.

The Eika et al. (2019) weights are a special case of this where

$$w_i = \frac{P\left(T_m = i, T_f = i\right)}{\sum_{k=1}^{N} P\left(T_m = k, T_f = k\right)}. \tag{4}$$

We show in Appendix B for which λ the weights as defined by equations (3) and (4) coincide and discuss results for different weights in Section 3.

2.3 Educational Categorizations

We consider three common categorizations of individual educational attainment. First, the highest level of completed education, i.e., primary (compulsory schooling), secondary (vocational degrees), and tertiary (higher) education. Second, the length of educational programs. This amounts to dividing the tertiary education category into two subcategories: Bachelor degree programs (four years or less, vocational colleges and universities) and Master/PhD programs (five years or more, graduate-level education at universities). Third, we subdivide the tertiary education category into six fields of study: (i) social and health; (ii) teaching; (iii) business, admin, and law; (iv) social sciences; (v) engineering, natural sciences, and technology; (vi) humanities.

In additional to the three commonly used categorizations, we propose a "career ambition" categorization. This is also based on education, but it distinguishes programs by the associated labor market outcomes of graduates. We use four-digit educational program codes (ISCED codes) that uniquely identify all educational programs in Denmark. We assume that individuals select educational programs based on the observed career trajectories of peers. Peers are defined as individuals who graduated from the same educational program as oneself. Hence, one's peer group exclude oneself. We measure the career trajectories of peers along two dimensions: starting wage, w_0, and wage growth, g. We do not condition on the timing of labor market entry.[7] Throughout, we use hourly wages to abstract from the intensive margin. We deflate hourly wages by regressing wages on year effects with 2000 as the base year. The starting wage w_0 is the average hourly wage of peers in the first five years in the labor force.[8] To calculate the average growth rate g, we measure the percentage change between w_0 and w_1 with w_1 being the average hourly wage of peers in years 9–11 in the labor force.[9] We average over multiple years for both w_0 and w_1 to smooth out year to year variation that is unrelated to worker productivity. The (expected) wage growth associated with an educational program is thus the percentage change between w_0 and w_1, averaged over all peers in the sample. To cluster individuals based on these continuous labor market outcomes (starting wage and growth), we use the k-means algorithm.[10]

3 Results

In Figure 1, we compare two classifications of educational programs in terms of starting wages and wage growth: in Panels (a) and (c), we use a standard classification based on the program length (primary, secondary , shorter tertiary (Bachelor), and longer tertiary (Master & PhD)). In Panels (b) and (d), we use the career-ambition classification that groups educational programs into four k-means clusters using starting wages and wage growth. The upper two Panels show scatter plots of all programs in the space of standardized starting wages and wage growth rates with their respective classification. The first important observation is the vast

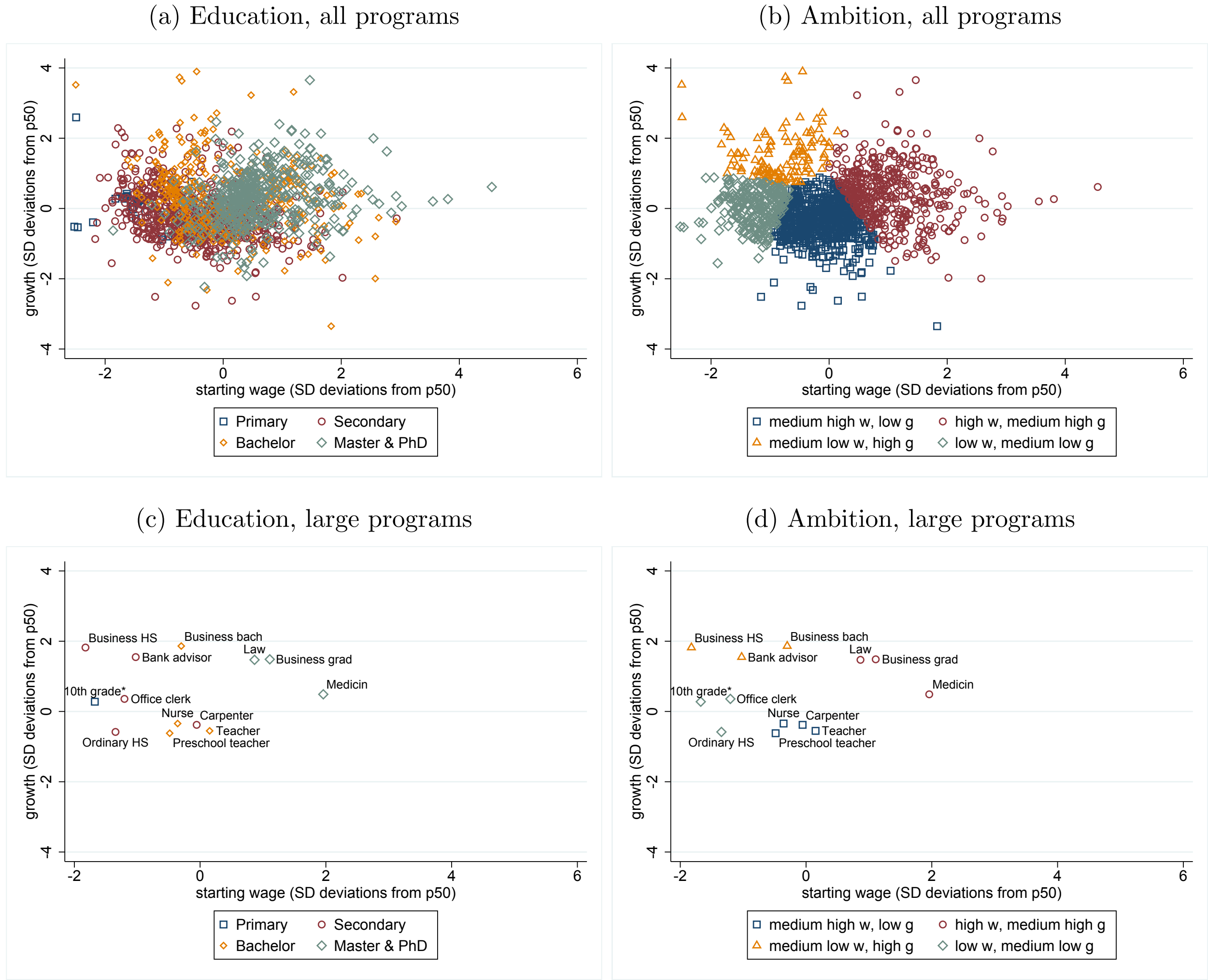
[7]That is, expected wage growth g_i of an individual who enters in 1990 is based on the observed wage growth trajectories of peers that enter later, e.g., in 2005. We could have chosen to only use peers that already graduated or only graduated around the time of starting the education. However, this would imply less peers in the group. Which of the two strategies is best depend on where the uncertainly lies. If we assume that individuals have perfect knowledge about the labor market trajectories of a given educational program and that programs are stable over time then including future peers is optimal.

[8]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.

[9]The focus on early career wage profiles is motivated by studies finding that wage profiles stabilize later in one's career (Bhuller et al., 2017).

[10]For an overview, see Steinley (2006).

Figure 1: Starting wages and growth of educational programs by type



Source: The population of couples see section 2
Note: Each point in the figures represents the average wage growth and starting wages in educational programs with at least ten graduates (panels a and b) observed by 2018. In panels c and d we show 13 exemplary programs with many graduates. The symbols of points refer to the educational level or ambition type associated with the program. Axes are deviations measured in standard deviations from the median.

heterogeneity in both dimensions, which a standard education-based classification hardly takes into account. In Panel (a), 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 (large gray diamonds). But 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. All four categories include programs with low and high wage growth.

Panel (b) shows how the k-means algorithm partitions the educational programs into four categories based on starting wages and growth. By construction, these clusters are internally homogeneous in terms of starting wages and growth. Here, the red category (circles) includes the highest starting wages. Tertiary programs with low starting wages are not part of it.

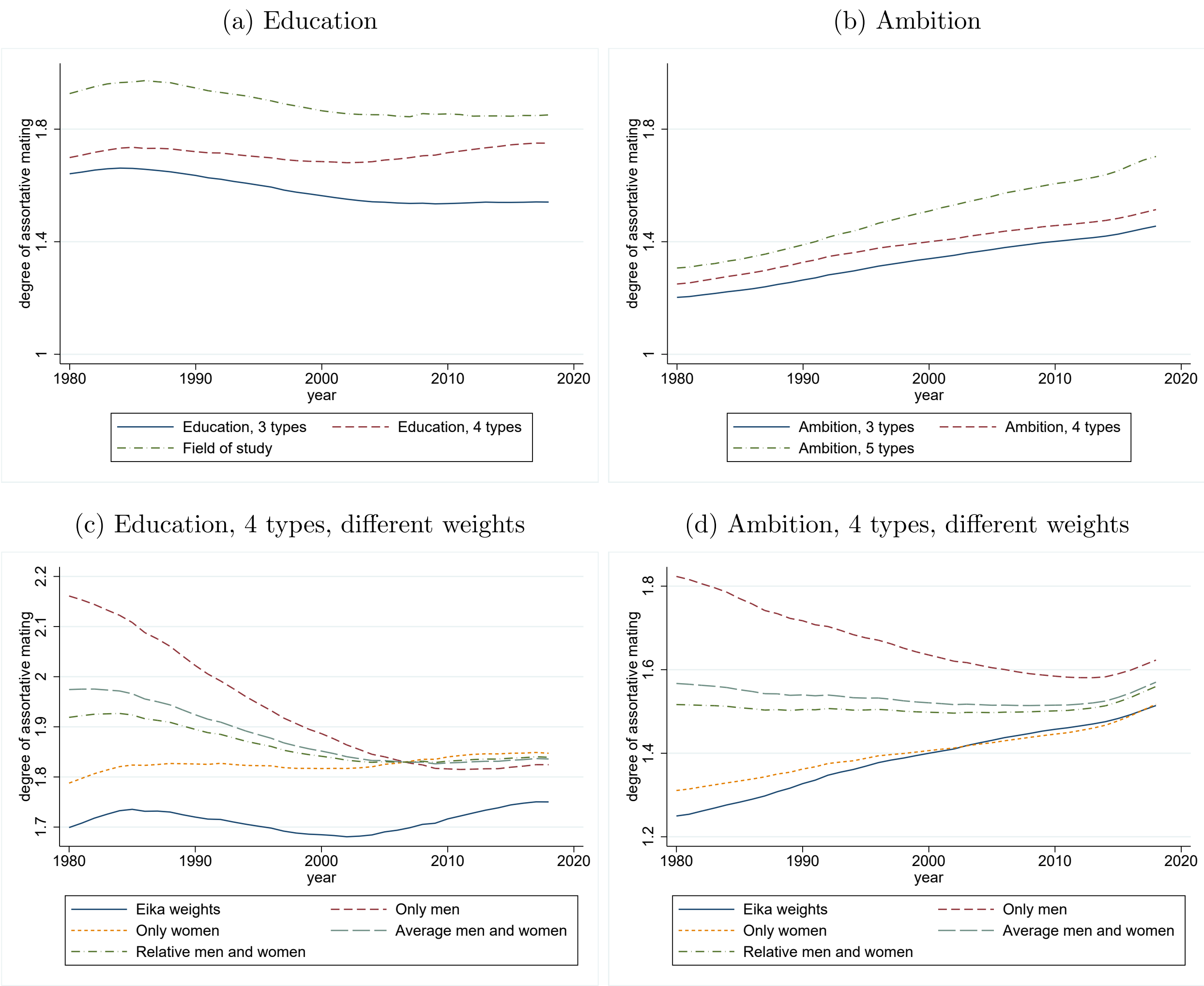
The dispersion in wage growth is relatively large, it ranges from -2 to 4 standard deviations relative to the median. The blue category (squares) includes starting wages within $\pm$ 1 standard deviations from the median. It is clearly delimited from above and does not include programs with high wage growth. The orange category (triangles) has relatively low starting wages but some programs with the highest wage growth overall. Finally, the gray category (diamonds) has the lowest starting wages and is relatively homogeneous in terms of wage growth, which is delimited within $\pm$ 1 standard deviations.

Panels (c) and (d) zoom in on 13 exemplary educational programs with a large number of graduates. In total these program constitute xx percent of the population. The three 5-year tertiary programs included—law, business, and medicine—end up in the career-ambition category with very high starting wages and relatively high wage growth. For shorter programs, however, we observe that the k-means algorithm combines programs that are similar in terms of labor market outcomes but separate according to the program length categorization. For example, the degrees of (preschool) teachers, nurses, and carpenters are quite similar in terms of labor market outcomes, but teachers and nurses are Bachelor degree programs while carpenters have a secondary, vocational degree. Similarly, individuals who finished compulsory schooling, high school, or a vocational degree as office clerks, are grouped together in the career-ambition categorization (due to comparable starting wages) but are separated in the educational-length classification. Finally, three programs with relatively low starting wages but high growth—degrees from high schools that specialize in business, bank advisors, and short tertiary degrees in business—are in the same ambition category but separate in terms of educational length.

Figure 2 shows differences between the aggregate sorting measure, $\mathcal{S}$, from Equation (2) which one obtains based on the two different categorizations. The figure also shows results using different ways of taking into account changes of the marginal type distributions through the weights. Consistent with Eika et al. (2019), Panel (a) shows that the weighted sum of likelihood indices does not increase after 1980. This is true for both three and four educational types, as well as the classification that subdivides tertiary education by field of study. However, based on the career-ambition categorization, we see a steady increase in sorting, regardless of the number of categories (three, four, or five). We show below that this increasing trend has implications for the link between ambition-based sorting and income inequality.

Panels (c) and (d) of Figure 2 illuminate the role that the weights play for these conclusions. The results in Panel (a) and (b) use the weighting of Eika et al. (2019). That is, the blue solid line in Panel (c) corresponds to the red dashed line for four educational types in Panel (a). Similarly, the blue solid line in Panel (d) corresponds to the red dashed line for four ambition

Figure 2: Aggregate Sorting Measures



Source: The population of couples see section 2
Note: Panels a and b report aggregated sorting $\mathcal{S}$ (equation (2)) by various education and ambition types from 1980-2018 using the weights from Eika et al. (2019) defined in equation (4). The detailed likelihood indices and weights can be found in Figures A.4 and A.5. Panels c and d report $\mathcal{S}$ using the same "Eika weights" as in panels a and b (Equation (4)) but also weights defined in Equation (3) for different values of λ. For only men $\lambda = 1$, only women $\lambda = 0$, average men and women $\lambda = 0.5$, and relative men and women $\lambda \approx 0.375$

types in Panel (b). The Eika et al. (2019) weights take into account changes to the underlying marginal distributions of couple types through their effect on the frequencies of couples observed "along the diagonal" of the contingency table. Specifically, they divide the frequency for each diagonal element by the total number of couples along the diagonal.

In practice, we find in Panels (c) and (d) that the diagonal weights used in Panels (a) and (b) based on Eika et al. (2019) most closely correspond to a weighting that puts all the emphasis on changes to the marginal type distribution of females. Referring back to the definition of weights in Equation (3), this means that the implied λ is relatively close to zero. To understand why diagonal weights implicitly put the weight on the marginal distribution of female types, consider that historically the educational attainment of women lagged behind men but caught up over the last couple of decades (Goldin, 2006). If there are more educated men than educated females, and educated men have a preference to marry their like, some of them have to stay single or marry down because highly educated women are in short supply. Figure A.2 in the Appendix shows the development of the marginal distributions for Denmark since 1980. The share of men with long-cycle tertiary education has more than tripled between 1980 and 2018. But for women, this share has increased by factor of 13. Similarly, for our ambition types, the share of men with high starting wages and high wage growth has almost doubled, while there are nearly eight times as many highly ambitious women in 2018 compared to 1980. If high-type women are relatively scarce, it is reasonable to expect that an increase in their supply will have a big impact on the allocation in the marriage market. A constraint on marriage market matching is relaxed and, hence, more sorting occurs. For this reason, it is unsurprising that the Eika et al. (2019) diagonal weights correspond closely to the case in which most weight lies on the marginal distribution of female types. Other weights, e.g., the average of the male and female marginal distributions ($\lambda = 0.5$) or the ratio of the male changes in the marginal distribution to the total change for male and female marginal distributions ($\lambda \approx 0.375$) imply a relatively flat sorting trend also for the ambition types. However, for all weightings, an increase occurs for ambition types after 2013, while sorting by educational types remains stable.

4 Counterfactual Analysis

We rely on the counterfactual re-weighting method introduced by DiNardo et al. (1996) and follow the implementation of Eika et al. (2019). Our goal is to show how sensitive conclusions about the influence of marriage market sorting on trends in income inequality across households are to the alternative categorizations of the marriage-market types that we have discussed.

Specifically, we analyze counterfactual household-level income inequality in 2018 under four different scenarios: (i) couples match randomly in the marriage market; (ii) couples match according to matching probabilities fixed in 1980; (iii) couples have their return to education or career ambition fixed at the 1980 level; (iv) the marginal distributions of education or career ambition types are fixed at the 1980.

We construct (i) with a rematching algorithm. First, the rematching algorithm treats all 2018 individuals as singles. Second, it draws potential couples by sampling males and females according to the respective marginal distributions. Third, it is decided by a draw from a binomial distribution with $p = 0.5$, due to the random matching scenario, if the potential couple becomes a household in this counterfactual setting. Finally, all the remaining nonmatched individuals are treated as singles and the process is reinitiated. The rematching algorithm stops when all individuals have been assigned to a counterfactual household.

Scenario (ii) is constructed similarly to (i). The only difference is that we use different matching probabilities, p. These probabilities correspond to the matching probabilities implied by the 1980 allocation of couple types. For each possible couple combination of (T_f, T_m), we get the implied 1980 matching probability by taking the average between the conditional probability that a male (female) of type t_m (t_f) is matched with a female (male) of type t_f (t_m).

To analyze the counterfactual income inequality under scenario (iii), we use a household re-weighting factor, $\widehat{\psi_y}$, to construct the counterfactual household income distribution, where y denotes household income, x couple combination of (T_f, T_m), s matching probabilities, and τ time:

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

With

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

We obtain $\widehat{\psi_y}$ by following the approach suggested by Fortin et al. (2011). To get 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 which are relatively more present in 2018 are weighted greater than one in the counterfactual income distribution and vice versa.

We obtain the last scenario (iv) with a similar approach as for (iii) by re-weighting households in the 2018 income distribution relative to changes in the marginal distributions of T_m and T_f. In this case the re-weighting factor is $\widehat{\psi_x} = (\widehat{\psi_y})^{-1}$.

Importantly, these counterfactual results rest on an independence assumption between matching probabilities s, the income distribution y, and marginal distributions of T_m and T_f. For example, we assume that changing matching probabilities in the marriage market do not change returns to types or the selection into education or career ambition types. In this sense the exercise is similar to a standard extended Oxcaca-Blinder decomposition, where we can one set of parameters at a time.

We summarize the counterfactual household-level income inequality in 2018 under the four different scenarios in Table 1. Income inequality is reported by both the P90 to P50 ratio

Table 1: Counterfactual Changes in Income Inequality

	(a) P90/P50			(b) Gini		
	1980	2018	Δ	1980	2018	Δ
Data	1.52	1.69	100%	0.241	0.307	100%
(i) Random matching	1.54	1.64	–	0.239	0.281	–
(ii) Fixed matching probabilities	1980	2018	Δ	1980	2018	Δ
Education, 4 types	1.52	1.66	84%	0.241	0.286	68%
Ambition, 4 types	1.52	1.67	89%	0.241	0.288	70%
(iii) Fixed returns to type	1980	2018	Δ	1980	2018	Δ
Education, 4 types	1.52	1.65	75%	0.241	0.250	13%
Ambition, 4 types	1.52	1.59	41%	0.241	0.245	5%
(iv) Fixed marginal distributions	1980	2018	Δ	1980	2018	Δ
Education, 4 types	1.52	1.72	121%	0.241	0.335	142%
Ambition, 4 types	1.52	1.64	69%	0.241	0.313	108%

Source: The population of couples see section 2
Note: Panel a reports the ratio of the 90th and 50th percentile in the income distribution, while panel b reports the Gini coefficient. We do this for 1980 and 2018 in the data and for four counterfactual scenarios using the re-weighting method introduced by DiNardo et al. (1996). For scenarios (ii)-(iv), we distinguish between categorizing educational programs by four education- or ambition types. Changes between 1980-2018 are denoted by Δ and scaled according to the change in the data being 100%.

reflecting the inequality in the upper half of the income distribution, and the Gini coefficient which summarizes inequality in the entire distribution. As a first observation, we see that household-level income inequality for couples has increased from 1980 to 2018 both measured by the P90 to P50 ratio (Panel (a)) and the Gini coefficient (Panel (b)). The true change of the respective inequality measure in the data is denoted as 100%, and we compare the arising inequality in the different counterfactual scenarios to this baseline.

The results for random matching (i) show that the level of income inequality is lower in 2018 if couples form with probabilities that resemble random matching, i.e., do not sort. This is in line with findings from the US (Eika et al., 2019) showing that sorting increases the *level* of household income inequality. However, the 1980 level of inequality implied random matching is not far from the truth.

Scenarios (ii)-(iv) decompose the increasing income inequality *trend*. Had the matching probabilities been kept fixed at the 1980 level (ii), then the increase in inequality would have been less pronounced. In other words, changes in matching probabilities between 1980 and 2018 have amplified the increasing income inequality trend holding all other variables fixed. The impact of changing matching probabilities on the inequality trend is similar for education and ambition types. The increase in inequality would have been 16% lower than the baseline for the education categorization and 11% lower for the ambition categorization (P90/P50). For the Gini coefficient, the effect of holding matching probabilities fixed is even larger. Interestingly, our findings differ from others in the literature (Breen and Salazar, 2011; Breen and Andersen, 2012; Eika et al., 2019; Chiappori et al., 2020a). We find that changing matching probabilities can explain a substantial part of the increasing trend in inequality in the Danish case. This indicates that sorting patterns are not only influenced mechanically by changing marginal distributions of T_m and T_f but also by changing preferences over partner types or changes to the matching process, e.g., reduced search frictions.

We confirm the conclusion made by Eika et al. (2019) that changes in returns to types (iii) explain a larger part of the increasing inequality than changing matching probabilities (ii). Focusing on the counterfactual Gini coefficients in 2018 when keeping returns to types fixed at their 1980 level, we see that Gini coefficients would barely have changed. This indicates that changes in returns to types have substantially amplified the increasing income inequality. The difference is less pronounced when looking at the P90/P50 ratios reflecting how changes in returns have mostly affected the difference between the top and bottom of the income distribution. Interestingly, the counterfactual P90P50 ratio for ambition types is lower than that of education types. Thus, changing returns to career ambition types explains more of the increasing inequality in the upper half of the distribution. This might reflect heterogeneity in increasing returns to educational programs within tertiary education.

Finally, we find that conclusions about the impact of changing marginal distributions on increasing income inequality (iv) differs for the education and ambition categorizations. In line with previous findings, we conclude that a shift in marginal educational distributions towards secondary and tertiary education has had a dampening effect on the inequality trend. The changes in the P90/P50 ratio and the Gini coefficient would have been 21% and 42% higher than the baseline, respectively, had the educational marginal distributions not changed since 1980. We reach the opposite conclusion for the P90/P50 ratio and ambition types. Hence, the shift towards higher starting-wage-and-growth ambition types has amplified the increasing inequality in the upper half of the income distribution. This highlights the importance of disentangling heterogeneity within broad educational categories, especially for tertiary education, which includes educational programs that are associated with very different labor market outcomes and, thus, career ambition types.

5 Conclusion

During the last couple of decades, both household income inequality and the correlation between spousal attributes have increased. However, previous studies (Breen and Salazar, 2011; Breen and Andersen, 2012; Eika et al., 2019; Chiappori et al., 2020a) have shown that the effect of increased education-based sorting on inequality is limited because observing more couples in which both spouses have the same level of education primarily reflects shifting marginal distributions. These studies have in common that they rely on relatively coarse educational categories to measure sorting. In this paper, we argue that the categorization of types is not an innocuous choice when studying the link between sorting and inequality.

Using detailed Danish register data for legally married and cohabiting couples aged 19-60 for 1980-2018, we develop a novel education-based categorization of marriage-market types. We call them career-ambition types. These types are based on labor market outcomes of specific educational programs. We cluster programs by starting wages and early-career wage growth with the k-means algorithm.

We find that different educational categorizations affect conclusions about trends in education-based sorting. As for the American case (Eika et al., 2019), trends are mostly stable in the Danish case too. However, sorting is higher for the four-group length of education categorization and even higher for the eight-group fields of study categorization compared to the three-group level of education categorization. We find a slight increase in sorting by length of education and a slight decrease for the two other categorizations. The fact that findings differ across the different classifications can be explained by the vast heterogeneity hidden within broad educational categories.

The career ambition type takes this heterogeneity within and across educational categories into account by, e.g., splitting high-starting-wage-and-growth tertiary programs (law, business, and medicine) and medium-starting-wage-and-low-growth tertiary programs (teachers, nurses, and preschool teachers) into different groups. The idea behind the career-ambition categorization is that the educational program an individual graduates from is a signal in the marriage market. In contrast to educational categories, we see a steady increase in sorting, regardless of the number of categories (three, four, or five) based on the career ambition categorization.

We also contribute to the ongoing debate in the literature (Eika et al., 2019; Chiappori et al., 2020b) about how to use weights on the sum of likelihood indices in order to take changing marginal distributions into account when measuring aggregate marriage market sorting over time. We find that conclusions about sorting prior to 2013 depend on the specification of weights. After 2013, the conclusions about stable sorting by educational categories and increasing sorting by career ambition types are robust to different weightings. Based on this weighting exercise, we find indications that the rapid changes in the marginal ambition type distribution for women towards high career ambitions explain a substantial part of increased sorting.

Based on a counterfactual re-weighting exercise (DiNardo et al., 1996), we find that sorting increases the level of household income inequality in line with findings from the US (Eika et al., 2019). We further show how increased sorting by ambition types can explain a substantial part of the increasing trend in inequality in the Danish case. This indicates that sorting is not only influenced mechanically by changing marginal distributions but also by changing preferences over partner types or changes to the matching process, e.g., reduced search frictions.

We confirm the conclusion made by Eika et al. (2019) that changes in returns to types explain a larger part of the increasing inequality than changing matching probabilities. Interestingly, changing returns to career ambition types explains more of the increasing inequality in the upper half of the income distribution. This might reflect heterogeneity in increasing returns to educational programs within tertiary education.

Finally, we find that conclusions about the impact of changing marginal distributions on increasing income inequality differ for the education and ambition categorizations. While a shift in marginal educational distributions towards secondary and tertiary education has had a dampening effect on the inequality trend, the shift towards higher starting-wage-and-growth ambition types has amplified the increasing inequality in the upper half of the income distribution. We argue that this highlights the importance of disentangling heterogeneity within broad educational categories. This is especially important for tertiary education which includes educational programs associated with very different labor market outcomes and, thus, career ambition types.

References

Abbott, Brant, Giovanni Gallipoli, Costas Meghir, and Giovanni L. Violante (2019) "Education Policy and Intergenerational Transfers in Equilibrium," *Journal of Political Economy*, Vol. 127, pp. 2569–2624, URL: https://doi.org/10.1086/702241, DOI: http://dx.doi.org/10.1086/702241.

Anderberg, Dan, Jesper Bagger, V. Bhaskar, and Tanya Willson (2022) "Marriage Market Equilibrium with Matching on Latent Ability: Identification using a Compulsory Schooling Expansion," unpublished.

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, URL: https://doi.org/10.1086/692509, DOI: http://dx.doi.org/10.1086/692509.

Breen, Richard and Signe Hald Andersen (2012) "Educational Assortative Mating and Income Inequality in Denmark," *Demography*, Vol. 49, pp. 867–887, URL: http://www.jstor.org/stable/23252675.

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, URL: https://doi.org/10.1086/661778, DOI: http://dx.doi.org/10.1086/661778.

Chiappori, Pierre-Andre, Monica Costa Dias, and Costas Meghir (2021) "The measuring of assortativeness in marriage: a comment."

Chiappori, Pierre-André, Sonia Oreffice, and Climent Quintana-Domeque (2012) "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market," *Journal of Political Economy*, Vol. 120, pp. 659–695, URL: https://doi.org/10.1086/667941, DOI: http://dx.doi.org/10.1086/667941.

Chiappori, Pierre-André, Monica Costa-Dias, Sam Crossman, and Costas Meghir (2020a) "Changes in Assortative Matching and Inequality in Income: Evidence for the UK*," *Fiscal Studies*, Vol. 41, pp. 39–63, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-5890.12217, DOI: http://dx.doi.org/https://doi.org/10.1111/1475-5890.12217.

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

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, URL: http://www.jstor.org/stable/2171954.

Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt (2022) "The Economics of Fertility: A New Era," Working Paper 29948, National Bureau of Economic Research.

Eika, Lasse, Magne Mogstad, and Basit Zafar (2019) "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, Vol. 127, pp. 2795–2835, URL: https://doi.org/10.1086/702018, DOI: http://dx.doi.org/10.1086/702018.

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, URL: https://doi.org/10.1162/003355301753265589, DOI: http://dx.doi.org/10.1162/003355301753265589.

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

Goldin, Claudia (2006) "The Quiet Revolution That Transformed Women's Employment, Education, and Family," *American Economic Review*, Vol. 96, pp. 1–21.

——— (2014) "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, Vol. 104, pp. 1091–1119, URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1091, DOI: http://dx.doi.org/10.1257/aer.104.4.1091.

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, URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.5.348, DOI: http://dx.doi.org/10.1257/aer.104.5.348.

——— (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.

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, URL: https://doi.org/10.1093/qje/qjw019, DOI: http://dx.doi.org/10.1093/qje/qjw019.

Kremer, Michael (1997) "How Much does Sorting Increase Inequality?*," *The Quarterly Journal of Economics*, Vol. 112, pp. 115–139, URL: https://doi.org/10.1162/003355397555145, DOI: http://dx.doi.org/10.1162/003355397555145.

Lee, Wang-Sheng and Terra McKinnish (2018) "The marital satisfaction of differently aged couples," *Journal of population economics*, Vol. 31, pp. 337–362.

Persson, Petra (2020) "Social Insurance and the Marriage Market," *Journal of Political Economy*, Vol. 128, pp. 252–300.

Reynoso, Ana (2022) "The impact of divorce laws on the equilibrium in the marriage market.," June, unpublished.

Steinley, Douglas. (2006) "K-means clustering: A half-century synthesis," *British Journal of Mathematical and Statistical Psychology*, Vol. 59, pp. 1–34, URL: https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/000711005X48266, DOI: http://dx.doi.org/https://doi.org/10.1348/000711005X48266.

Wiswall, Matthew and Basit Zafar (2021) "Human Capital Investments and Expectations about Career and Family," *Journal of Political Economy*, Vol. 129, pp. 1361–1424, URL: https://doi.org/10.1086/713100, DOI: http://dx.doi.org/10.1086/713100.

Online Appendix

(not for publication)

A Data Appendix

Here provide all the details.

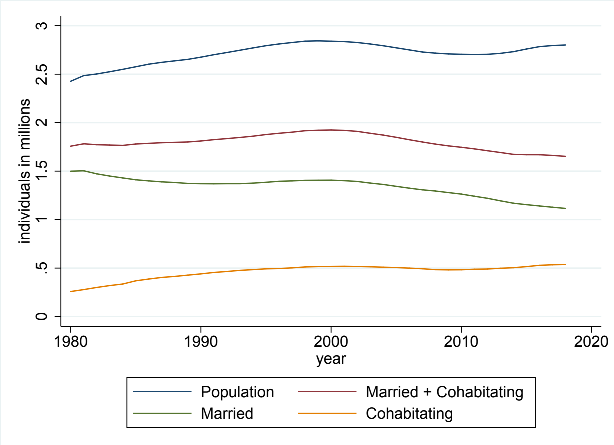
B Comparison of Weights

Here we show for which λ the Chiappori et al. (2020b) and Eika et al. (2019) weights coincide.

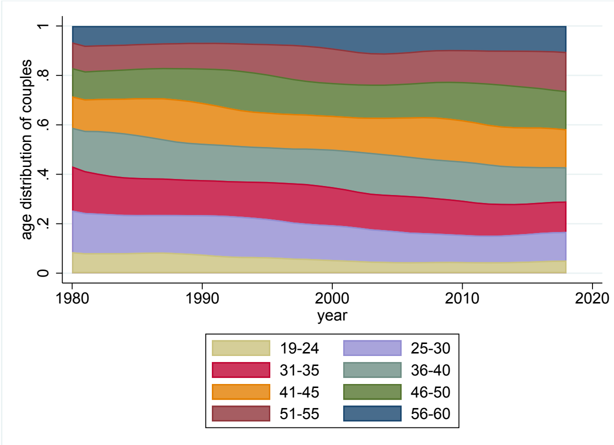
C Additional Results

Figure A.1: Marriage, Cohabitation, Age Composition

(a) Stocks of Couples



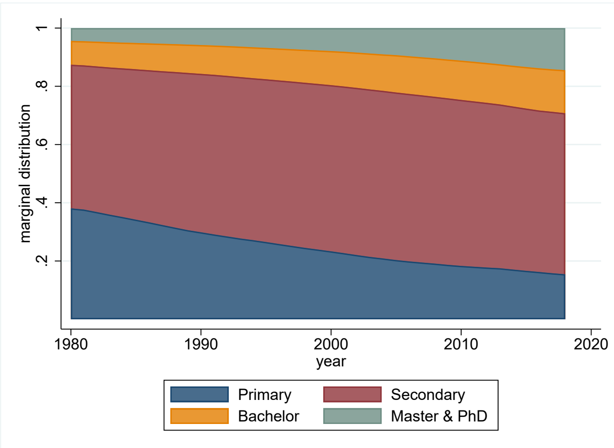
(b) Age Composition of Couples



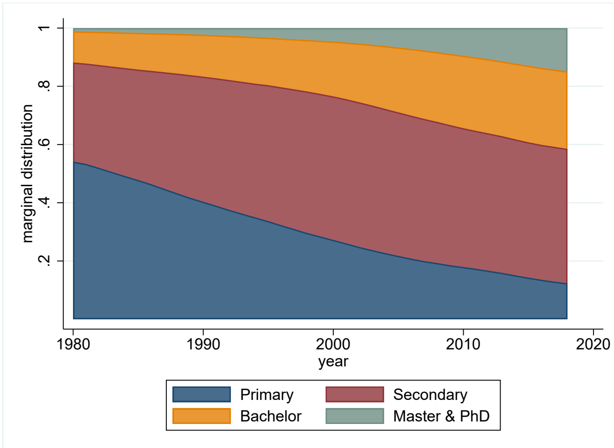
Source: Panel a is based on the full population while panel b is based on the population of couples see section 2
Note: Panel a reports the development in numbers of individuals by marital status. Panel b plots the age distribution of individuals who are either legally married or cohabiting.

Figure A.2: Marginals

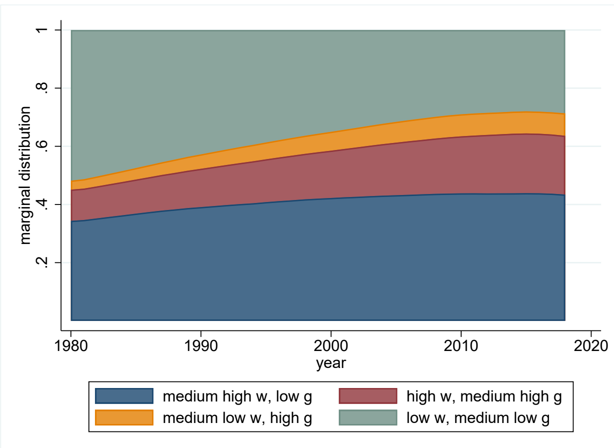
(a) Education, 4 types, Men



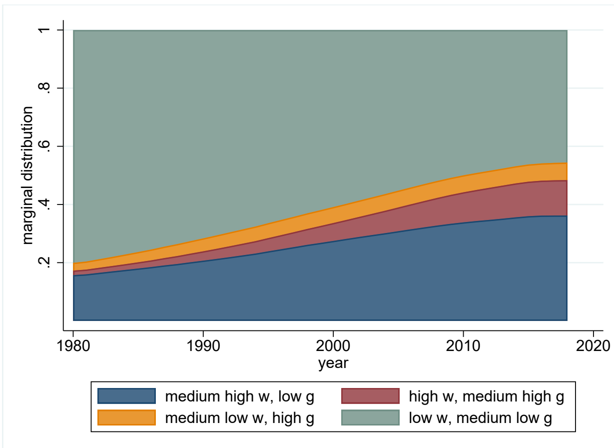
(b) Education, 4 types, Women



(c) Ambition, 4 types, Men



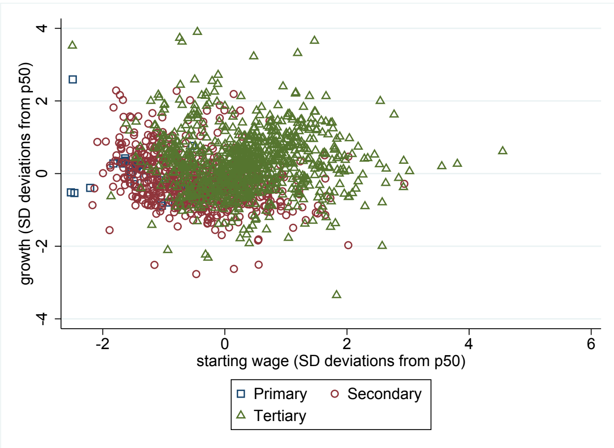
(d) Ambition, 4 types, Women



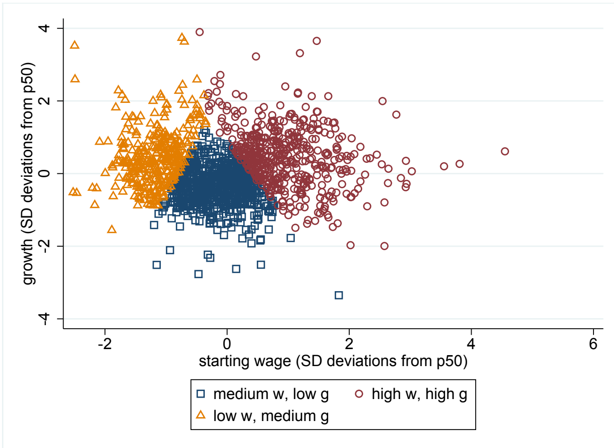
Source: The population of couples see section 2
Note: Marginal distributions for men and women over time by four education- or ambition types.

Figure A.3: Other classifications by starting wages and growth

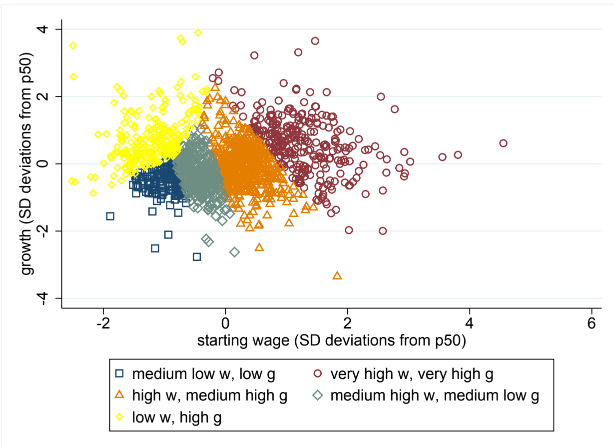
(a) Education, 3 types



(b) Ambition, 3 types



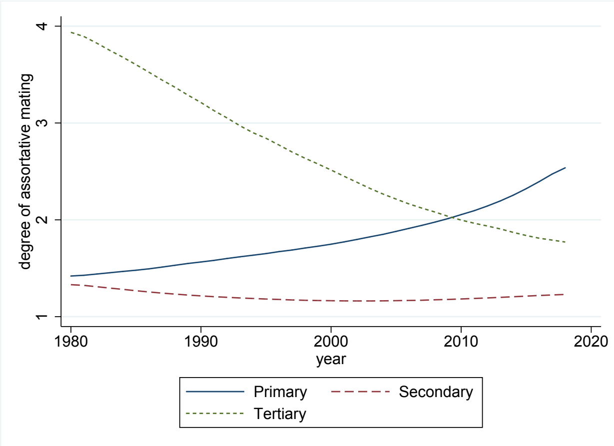
(c) Ambition, 5 types



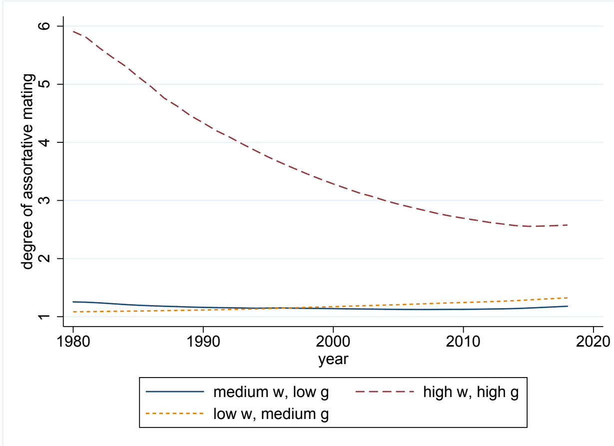
Source: The population of couples see section 2
Note: Each point in the figures represents the average wage growth and starting wages in educational programs with at least ten graduates observed by 2018. The symbols of points refer to the educational level or ambition type associated with the program. Axes are deviations measured in standard deviations from the median.

Figure A.4: Disaggregated Sorting Measures

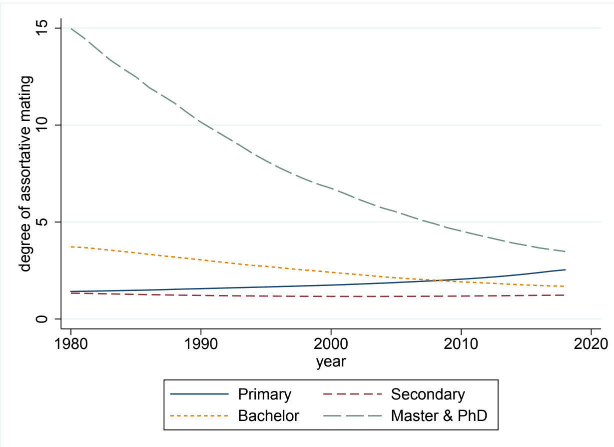
(a) Education, 3 types



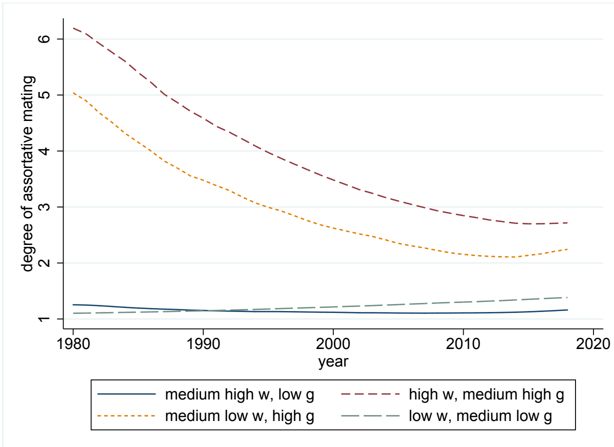
(b) Ambition, 3 types



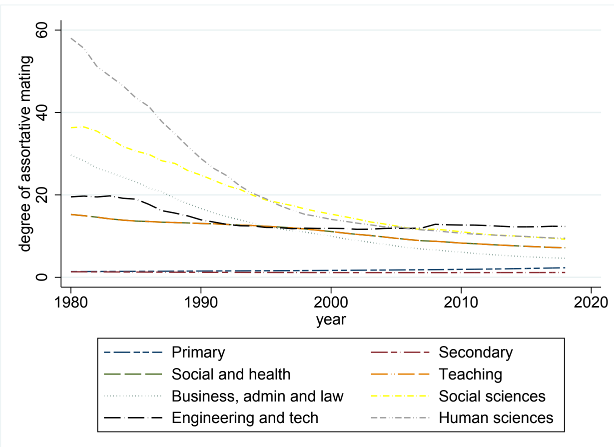
(c) Education, 4 types



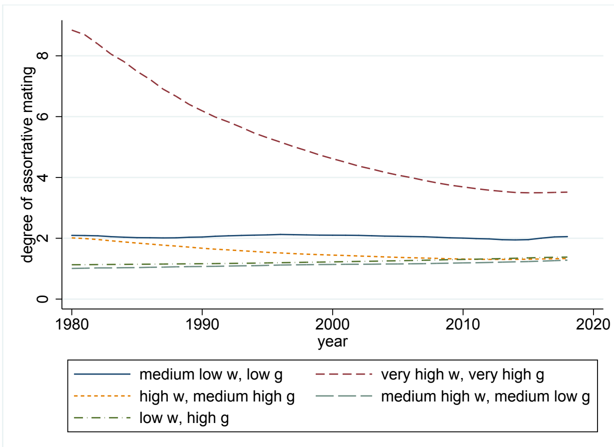
(d) Ambition, 4 types



(e) Field of study



(f) Ambition, 5 types

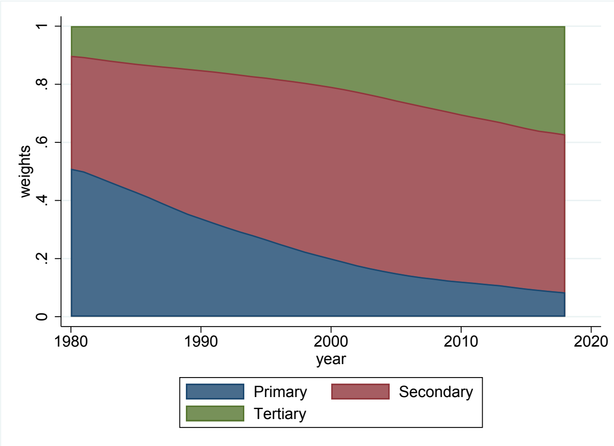


Source: The population of couples see section 2

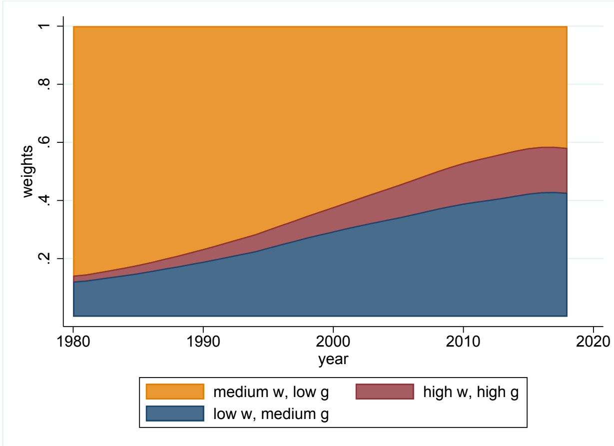
Note: Likelihood indices for assortatively matched couples cf. equation (1) for six different education- and career ambition type categorizations.

Figure A.5: Weights

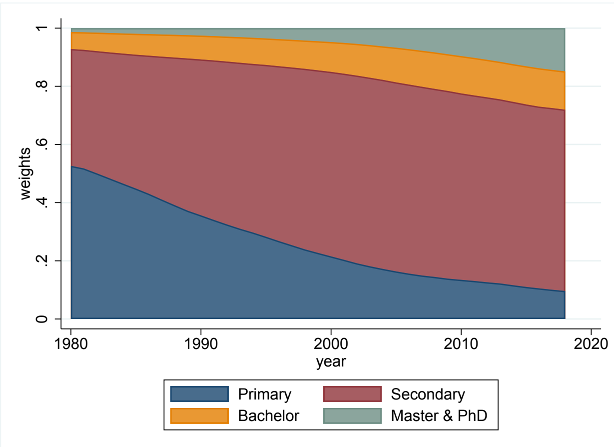
(a) Education, 3 types



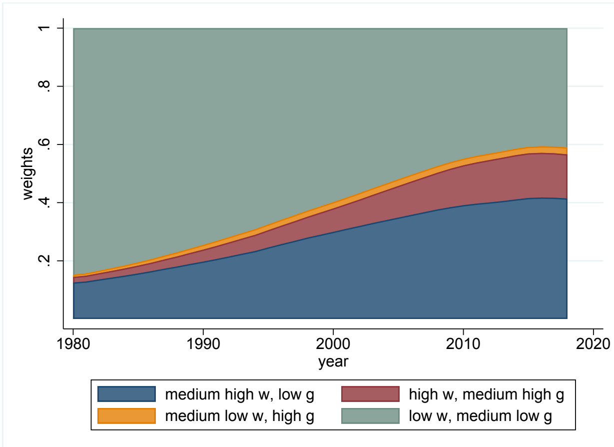
(b) Ambition, 3 types



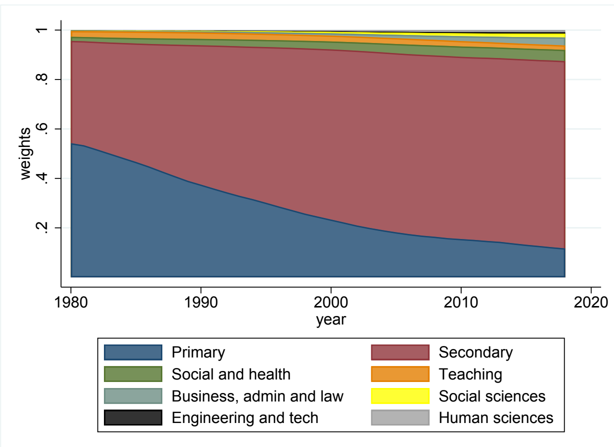
(c) Education, 4 types



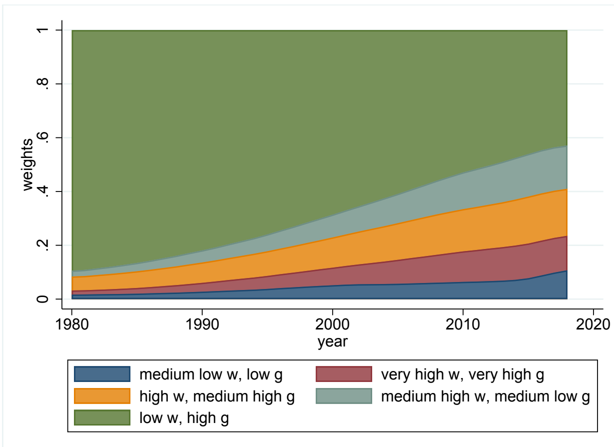
(d) Ambition, 4 types



(e) Field of study



(f) Ambition, 5 types



Source: The population of couples see section 2
Note: Weights for assortatively matched couples cf. equation (4) for six different education- and career ambition type categorizations.