Educational Ambition, Marital Sorting, and Inequality*

Frederik Almar† Benjamin Friedrich‡ Ana Reynoso§

Bastian Schulz¶ Rune Vejlin‖

January 19, 2024

Abstract

This paper revisits the link between education-based marriage market sorting and income inequality. Leveraging Danish administrative data, we develop novel marriage market types based on the starting wages and wage growth trajectories associated with educational programs: educational 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 or field of study. Hence, the mapping between education and marriage market types matters crucially for conclusions about the role of marital sorting in rising income inequality.

*This paper is a substantially revised version of IZA DP No. 15912 (January 2023), which was circulated under the title "Marital Sorting and Inequality: How Educational Categorization Matters". First draft: October 2022. We thank Edoardo Ciscato, Rasmus Landersø, Elena Mattana, Laura Pilossoph, Jakob Søgaard, Basit Zafar, and audiences at Aarhus University, IAB Nuremberg, the DGPE network, DAEiNA, University of Michigan, Yale University, and the 2024 ASSA meetings 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

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

¶Aarhus University, Department of Economics and Business Economics, 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 education-based assortative matching in the marriage market to rising household income inequality. Some studies find evidence that marital sorting on education has strengthened over the last decades and argue that this 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). Other papers argue against both findings (Kremer, 1997; Breen and Salazar, 2011; Breen and Andersen, 2012; Eika, Mogstad and Zafar, 2019; Gihleb and Lang, 2020). We contribute to this debate by showing that the way in which data on education are used to capture relevant traits for marriage market matching influences conclusions about the interplay of sorting and inequality.

Specifically, we leverage rich administrative data from Denmark and summarize individual educational attainment to construct novel *marriage market types* that are clearly distinct in the earnings potential and future work-life balance associated to the educational degree. Based on these types, we find that assortative matching in the marriage market increased and significantly affected between-household income inequality. We compare this finding to results obtained with commonly-used marriage market types constructed based on education levels (primary, secondary, bachelor's, and master's/PhD) and the post-secondary field of study. We find that these types conceal heterogeneity in both earnings potential and future work-life balance. In turn, sorting by levels and sorting by fields exhibit no trend and do not affect inequality growth.

Our novel definition of marital types uses detailed labor market outcomes that are associated with the most advanced educational program that individuals graduate from. Programs are defined based on four-digit codes from the education register, which identifies more than 1800 education programs in Denmark. Examples include the vocational training of *carpenters*, professional degrees held by *nurses* and *pre-school teachers*, and 5-year-or-more university degrees in *law* and *business*. As we show, these programs send an important signal in the marriage market: they capture variation in both earnings potential (Altonji, Kahn and Speer, 2014, 2016; Kirkeboen, Leuven and Mogstad, 2016) and future work-life balance (Wiswall and Zafar, 2021; Goldin, 2014). The idea is that individuals in the marriage market observe the educational program of potential partners and assess their attractiveness based on the typical career path of graduates from that program.

To capture heterogeneity across education programs in their earnings potential and future work-life balance, we merge the Danish education registers with the labor market histories of all program graduates and compute average *starting wages* and *wage growth* trajectories for each program. Using these two characteristics is a parsimonious, yet conceptually crucial step beyond using a summary measure such as average lifetime earnings. In particular, at the same level of lifetime earnings, different combinations of starting wages and wage growth can have opposite implications for the work-life balance. For example, individuals who graduate from programs with medium-level starting wages and medium-level wage growth may be expected to achieve the same lifetime earnings as individuals who start their careers at low wages and face a steep wage growth profile. However, individuals who chose a high-wage-growth career may be expected to invest much more time in their jobs than individuals who progress on a flatter career path. As a result, the two types of individuals are expected to allocate different amounts of time to the family and, hence, have different values in the marriage market.

We define four marriage market types by grouping programs and graduates based on similarity in the two dimensions—starting wages and wage growth—using k-means clustering, a well-established and popular partitioning method in machine learning and computer science (Steinley, 2006). This method has 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. An advantage of using k-means is that it allows us to consider multiple labor market characteristics and collapse them into a single, one-dimensional marriage market trait. This greatly simplifies the analysis of sorting.[1]

The first part of our analysis robustly shows that our novel ambition types convey important information about both earnings potential and work-life balance that marriage market types based on education levels or field of study fail to capture.

First, 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 signaling *ambition* in their later career and label our categorization *educational ambition*.[2] In contrast, groups based on *educational levels* and *educational fields* mask significant heterogeneity in the starting wages and wage growth trajectories of graduates.

Second, our four ambition types are also clearly distinct in the information they carry about career flexibility and expected time commitments to the family. Using regression analysis, we

[1] Multidimensional sorting has been explored, for example, by Lindenlaub (2017), Low (Forthcoming), Ciscato and Weber (2020), and Foerster, Obermeier, Schulz and Paul (2022).

[2] We interpret this as a noisy signal because it is based on having completed a program in which *peers* have high-starting-wages/high-wage-growth.

investigate the link between the program-level outcomes that we use to construct the ambition types and seven proxies for the graduates' work-life balance. These proxies include a flexibility-in-working-hours measure inspired by Goldin (2014), the probability of working full time, the probability of becoming a manager, and the age at which the first child is born. Higher starting wages or wage growth are associated with less work-life balance, even conditional on the level of education, the field of study, and life-time earnings.

In the second part of the analysis, we compare trends in marital sorting and their contribution to the rise in between-household inequality across the different definitions of marital types—ambition types, educational levels, and educational field.

Our first main finding is that marital 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 and fields has hardly changed. Thus, conclusions about sorting trends crucially depend on how the categorization of underlying educational programs into types is implemented. We follow Eika et al. (2019) and Chiappori, Costa Dias and Meghir (2020b) and flexibly control for changing marginal distributions of educational attainment over time by defining our sorting measures as the weighted sum of the matching frequencies of equally educated couples relative to the same frequencies 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 the different categorizations can be compared.

Our second main finding is that changes in marital sorting on 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 Denmark. In contrast, the small changes in sorting on educational level and field of study have negligible effects on inequality. Methodologically, we compare the observed between-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 first proposed by DiNardo, Fortin and Lemieux (1996), see also Fortin, Lemieux and Firpo (2011) and the application in Eika et al. (2019). 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, mainly because more females graduate from these programs in 2018 compared to 1980.

By proposing a new definition of marital types based on the labor market and marital value of *programs*, we extend the recent literature that emphasizes the value of college degrees in the marriage market (Kirkeboen, Leuven and Mogstad, 2022, 2016; Artmann, Ketel, Oosterbeek and van der Klaauw, 2021; Nielsen and Svarer, 2009; Wiswall and Zafar, 2021; Seiver and Sullivan, 2020; Han and Qian, 2022). Unlike these papers, our measure of educational ambition considers programs from all levels of education—from compulsory schooling to graduate school—allowing us to study trends in sorting and inequality considering the broad population including couples of mixed levels of education.

Moreover, we contribute to the debate on the relationship between trends in sorting and inequality by showing that the choice of how to map detailed educational programs into a small number of types affects the conclusions. The previous literature that considers matching on the level of education (Breen and Salazar, 2011; Breen and Andersen, 2012; Greenwood et al., 2014; Eika et al., 2019; Chiappori et al., 2020a) or on field of study (Seiver and Sullivan, 2020; Artmann et al., 2021; Han and Qian, 2022) arrive at conflicting results. Instead, we consider sorting on novel ambition types and reveal an upward trend in assortative matching and, in turn, a significant contribution to the increment in inequality. Interestingly, our results robustly remain with alternative ways of constructing marital types as long as they summarize the earnings potential and work-life balance of education.

Finally, our insight that marital traits matter for the relationship between sorting and inequality can potentially be applied in the very recent literature that studies the link between assortative matching and intergenerational mobility (Bailey and Lin, 2022; Binder, Walker, Eggleston and Murray-Close, 2022; Gayle, Golan and Soytas, 2015). Because these studies compare sorting measures across groups—defined by, for example, race or income—the definition of the trait on which people sort in the marriage market is potentially important for their conclusions as well.

Our analysis highlights the importance of considering the relevant aspects of educational attainment and expected career outcomes when defining marital types. All three definitions of marital types that we consider—ambition, educational levels, and educational fields—use information from the most advanced program an individual graduates from. Still, conclusions about whether sorting has changed and influenced inequality differ significantly.

The paper is organized as follows. Section 2 introduces our data. In Section 3, we derive our ambition types and document how they capture the heterogeneity in earnings potential and work-life balance that is masked by education levels and fields of study. In Section 4, we show that only marital sorting on ambition has increased over time, while sorting on educational level or field has been flat since the 1980s. Section 5 presents our analysis of the drivers of changes in inequality and Section 6 assesses the sensitivity of our findings. Section 7 concludes.

2 Data

We use Danish register data for the period 1980 to 2018. The data is organized on a yearly basis and provides information for all of Denmark's residents on their education, marital status, fertility, labor market outcomes, and the identity of their marriage or cohabiting partner (BEF, Statistics Denmark, 1980–2018). We complement this data with the Danish Labor Force Survey (LFS) (LFS, Statistics Denmark, 2000–2018), which contains more detailed information on hours of work for a sub-sample of residents. Unique person identifiers allow us to merge individuals across all datasets. Moreover, critical to our analysis, we match individuals to their partners using the partner's unique identifiers. We next describe the key variables we construct for our analysis and provide further details in Online Appendix A.

To construct education and labor market outcomes we include all individuals living in Denmark between 1980-2018 in the age range 19-60, *irrespective of marital status.* On average, we observe 3,031,511 individuals per year.

On this sample, we measure individual education as the most advanced *educational program* an individual graduates from. Educational programs are defined based on four-digit ISCED codes from the education register, which uniquely identifies around 1,800 educational programs in Denmark (UDDA, Statistics Denmark, 1980–2018). Some of the most popular programs include vocational degrees such as *bank advisor*, *carpenter* or *office clerk*, bachelor's degrees like *nurse* or *pre-school teacher*, and master's degrees in *business*, *law* or *medicine*.

We identify each program's graduates in this sample and use their *hourly wages* according to the employment register (IDAN, Statistics Denmark, 1980–2018) to calculate labor market outcomes at the program level. To abstract from the increasing mean and variance of the hourly wages, we use log hourly wages and regress them on year dummies with 2000 as the base year and use the residuals in the remainder of the analysis.[3] The income registers also give us access to individual and parental wealth, which we use to describe the properties of our novel marital types introduced in the next section.

In order to compute *household* income, we follow the literature and exclude one-person households because our focus lies on the link between marital sorting and between-household inequality (as in Eika et al., 2019). Our sample of couples thus consists of all individuals between 19 and 60 years old who are either married to or cohabiting with another individual in this age range of the opposite sex.[4] On average, we observe 1,800,866 individuals in couples per

[3]See Online Appendix A.2 for details on the detrending procedure.

[4]Legal institutions in Denmark guarantee almost equal treatment of married and cohabiting couples. In the data, cohabiting couples are identified based on a number of criteria: two opposite-sex individuals who have a

year. There is an upward (downward) trend in cohabitation (legal marriage), but the combined stock of couples is stable over time.[5] Household income is the sum of each spouse's yearly labor income from both regular employment and self-employment, based on the income register (IND, Statistics Denmark, 1980–2018).

As mentioned above, we use the representative Danish LFS to investigate how career choices manifest themselves in labor market outcomes beyond wages and participation. For example, the LFS contains questions about the exact number of hours worked, while register data only contains information on contractual hours. The LFS also delivers information on whether the individual is working in the evenings, on weekends, and from home, which helps us to assess the work-life balance across educational programs and ambition types. The LFS data are available from 2000 onward and can be merged with our main data set at the individual level.[6]

3 Educational ambition Marriage Market Types

The literature has established that education is a valuable trait in the marriage market because it predicts both earnings potential and future time investments into career and family (Chiappori et al., 2018; Gayle and Shephard, 2019; Calvo et al., 2022; Reynoso, 2023; Calvo, 2023). First, education is associated with labor market trajectories. Spouses pool earnings that evolve over time, and the parameters of their earnings processes depend on initial education. Second, spouses jointly produce a public good that requires time inputs. The cost and productivity of these inputs likewise depends on the education of the spouses.

It is therefore desirable for the analysis of sorting in the marriage market to define *marriage market types*—the variable on which individuals are assumed to match in the marriage market—such that those marital types capture the earnings potential and work-life balance expectations of the individual. We turn to this next.

3.1 Conceptual Framework to Define Marriage Market Types

To fix ideas on how we use information on education to construct marriage market types, consider a marriage market in which men and women are distinguished by their program of education, $i \in \mathcal{P} = \{Program_1, Program_2, ..., Program_I\}$, as defined in Section 2. I is

joint child and/or share an address without other adults, exhibit an age difference of less than 15 years, and have no family relationship.

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

[6] Approximately 72,000 participants are surveyed on an annual basis. The sample is weighted to ensure that it is representative of the entire population of Denmark.

the total number of programs. Each program i is characterized by an N-dimensional vector of observable characteristics, $x_i = (x_{1i}, x_{2i}, ..., x_{Ni})$, and a dataset at the program level is defined as $x: \{x_i\}_{i \in \mathcal{P}}$. Examples of such characteristics include the length of the educational program, the field of study, the level of education, and the labor market outcomes of graduates.

For both tractability and ease of interpretation in marriage and labor market research, the programs of education are commonly grouped into a small number of marital types based on their similarity in *selected* characteristics, $\tilde{x} \subset x$.

Formally, let $\mathcal{T}_{\tilde{x}} : \tilde{x} \rightarrow t = \{Type_1, Type_2, ..., Type_T\}$ be a mapping that defines T marriage market *types* by grouping the I programs of education based on their similarity in the sub-vector of observable characteristics $\tilde{x}$. Importantly, $T << I$.

Many papers in the literature use the mapping $\mathcal{T}_{Levels}$, which maps programs based on similarity in one characteristic, namely, their level of education. Typically, four types are chosen, $t_{Levels} = \{Primary, Secondary, Bachelor, Master\&PhD\}$. Another commonly used one-dimensional mapping in the literature focuses on post-secondary education only (Kirkeboen et al., 2022; Seiver and Sullivan, 2020; Han and Qian, 2022; Artmann et al., 2021). It groups educational programs by field of post-secondary study, which maps the fields of the individual programs, $\tilde{x}_i = field_i$, into larger groups of fields, $t_{Fields} = \{Field_1, Field_2, ..., Field_T\}$.

3.2 Construction of Ambition Types

We take advantage of our rich data and construct marital types using two averages of labor market outcomes at the program level. As we show below, these outcomes capture well both the earnings potential and the expected future work-life balance of graduates for each educational program. We use the *average starting wage*—denoted by w_0—and *average wage growth over the early career*—denoted by g.

Formally, for each of the more than 1800 programs of education i we observe $\tilde{x}_i = (w_{0,i}, g_i)$, calculated using information on all individuals who completed their education after 1980.[7] As explained in more detail in Online Appendix A.2, we first residualize log hourly wages and then compute w_0 as the average hourly wage of program graduates during the first five years in the labor force.[8] 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 program graduates in years 9-11 in the labor

[7] 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.

[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.

force.[9] We average over years for both w_0 and w_1 to smooth out year-to-year variation that is unrelated to worker productivity.

In our benchmark analysis, we construct $T = 4$ marriage market types using the mapping $\mathcal{T}(w_0, g)$. It clusters programs based on standardized starting wage and growth using the k-means algorithm (Steinley, 2006). This method minimizes the within-cluster variation in the two dimensions and thus produces homogeneous groups in terms of starting wages and growth. We denote this mapping $\mathcal{T}_{Ambition}$, and it creates the following four types

$$t_{Ambition} = \{(low\ w_0, low\ g), (high\ w_0, low\ g), (low\ w_0, high\ g), (high\ w_0, high\ g)\}.$$

Panel (a) in Figure 1 plots the mapping $\mathcal{T}_{Ambition}$. It shows how the program-specific (w_0, g) tuples map into our ambition types. The plot locates all of the 1800+ programs in the space of standardized starting wages (horizontal axis) and standardized wage growth (vertical axis) and distinguishes the resulting four types with different colors and markers. The mapping delivers four groups that are clearly (and by construction) distinguished by whether starting wages and wage growth are low or high. Our interpretation is that graduates from high-starting-wages/high-wage-growth programs signal *career-ambition* in the marriage market. Thus, we label our categorization *educational ambition*. We confirm below that graduates from high-starting-wages/high-wage-growth programs indeed have high earnings potential but also a work-life balance that is likely to constrain their time investments into the family.

For comparison with previous studies, we also construct the mappings $\mathcal{T}_{Levels}$ and $\mathcal{T}_{Fields}$, closely following their definitions in the literature. Panels (c) and (e) repeat the placement of educational programs in the (w_0, g) plane from Panel (a). The positions of all programs are identical across panels, but (c) and (e) assign those programs to different color-marker groups depending on the level or post-secondary field, respectively. $\mathcal{T}_{Levels}$ groups programs based on educational level, i.e., Primary (compulsory schooling), Secondary (high school and vocational degrees), Bachelor (tertiary degrees of four years or less of duration), and Master & PhD (tertiary degrees with study times of five years or more). To construct $\mathcal{T}_{Fields}$ without losing non-tertiary programs, we keep primary and secondary from $\mathcal{T}_{Levels}$ but subdivide the tertiary category based on the field of study. We consider six post-secondary fields of study that closely resemble the fields used in Kirkeboen et al. (2022) and Eika et al. (2019): Humanities, Health & Welfare, Social Science, STEM, Business, and Other.

[9] We focus on the first 9-11 years because wage profiles stabilize later in one's career (Bhuller et al., 2017). Moreover, extending the time window during which we assess wage growth would reduce the number of individual observations that we can use to calculate the average.

Online Appendix Table A.1 describes the four educational ambition types in terms of population shares, gender composition, income moments, and parental wealth at graduation. Roughly 10% of the individuals in our sample are in the high-starting-wages/high-wage-growth group. Two thirds of these individuals are male. The group with high starting wages but low wage growth is predominantly male. Moreover, individuals who graduate from programs with high starting wages tend to have wealthy parents.

Additionally, Table A.2 shows the cross-tabulation of ambition types and levels or fields. Our ambition types go across educational levels and fields. For example, 10% of graduates who can expect high starting wages and growth do not have a tertiary but a secondary degree. At the same time, more than 40% of individuals in the high-starting-wages/low-wage-growth group have a tertiary degree. This clearly shows that graduating from a university does not guarantee high wage growth. Moreover, Social Science, Business, and STEM fields dominate in the group with high starting wages and growth. Graduates with degrees from the Humanities or Health & Welfare, however, are most commonly found in the groups with low wage growth and either high or low starting wages.

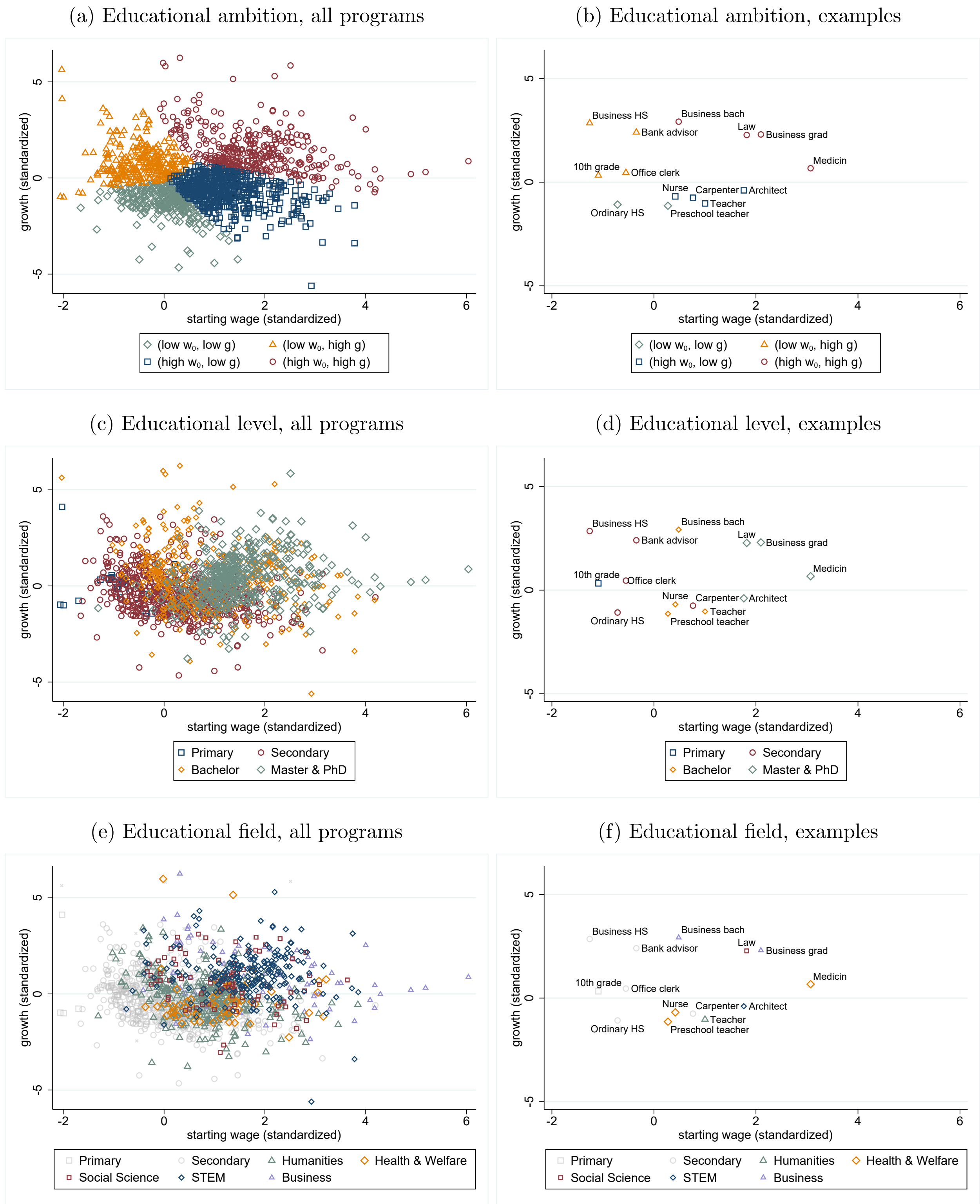
3.3 The Labor Market Value of Education Types

Our clustering of programs based on the average labor market outcomes w_0 and g implies that the four ambition types reflect the labor market value of education significantly better than the commonly used educational level or educational field types.

The four ambition types in Panel (a) are clearly distinct in terms of labor market starting conditions and wage progression. 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 interpreted as the high-ambition 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 while the blue cluster (squares) has relatively low wage growth but high starting wages.

In contrast, the relationship between marital types and the two labor market outcomes becomes more blurry when the types are constructed based on level or field of education. Panel (c) shows that marital types based on educational level exhibit a significant overlap in terms of both w_0 and g. For example, even though starting wages are on average low for Primary (blue squares) and high for Master & PhD (gray diamonds), many Secondary (red circles) and Bachelor (small orange diamonds) programs have higher starting wages than many Master &

Figure 1: Education-Based Marriage Market Types and their Starting Wages and Wage Growth



Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The horizontal axis corresponds to the standardized w_0 and the vertical to the standardized g. Points in the panels locate all educational programs with at least ten graduates in 2018—described in Section 2—along these two dimensions. Colors and markers uniquely assign each program to a marriage market type, depending on the panel's definition.

PhD programs. Moreover, there is no clear pattern in the growth dimension because programs in all four educational level types can be found in the same range of g. In Panel (e), types defined based on post-secondary fields of study show a similar overlap. For example, graduates from programs categorized as STEM (blue diamonds), Business (lilac small triangles), and Social Science (red squares) are relatively similar in terms of average wage growth but the variation in starting wages within each group is vast.

To obtain a sense of which specific educational programs are included in the respective clusters, in Panels (b), (d), and (f) we locate the 14 largest programs in the (w_0, g) plane for the ambition, levels, and fields mappings, respectively.[10] While graduates from 5-year business and architecture programs face very different labor market prospects—with architects expecting a flatter wage growth than business graduates—and are therefore assigned different ambition types, they are grouped together according to their level of education. Similarly, medical doctors and nurses have the same field of study but their assigned ambition types are different because doctors face a steeper wage profile than nurses.

3.4 The Marital Value of Ambition Types

Our definition of marital types based on (w_0, g) reflects expectations about the future work-life balance of graduates beyond the signaling value that the level of education or the field of study possess. To show this, we build on the literature and construct seven proxies for the work-life balance of graduates of specific educational programs. These proxies emphasize the trade-off between career investments and time commitments to the family (Wiswall and Zafar, 2021; Goldin, 2014; Calvo et al., 2022). Each row in Panel A of Table 1 corresponds to one of the proxies while the columns label the ambition types. The table shows the mean (and standard deviation in parenthesis) of the proxy conditional on each ambition type—in the first four columns—and for the pooled sample—in the last column.

Inflexibility is constructed as the ratio of the w_1 of full-time workers to that of part-time workers. It reflects the additional return to working long hours, a measure inspired by Claudia Goldin's analysis of differences across occupations in the US (Goldin, 2014). *Ever manager* is the fraction of graduates who reach a managerial position (hold a corresponding occupational code for at least two consecutive years). *Participation* captures the average share of time across the life-cycle during which program graduates are active in the labor market (work at

[10]These are the 4 four-digit educational programs with the largest number of graduates within each cluster. We count graduates in the 2018 sample of couples (defined in Section 2). In this sample, the examples cover 21% of all graduates.

Table 1: The Work-Life Balance Profile of Educational Ambition Types

Ambition type, (w_0, g)	(low, low)	(high, low)	(low, high)	(high, high)	All
Panel A: Main sample					
Inflexibility	1.052	1.066	1.123	1.119	1.095
	(0.059)	(0.076)	(0.034)	(0.069)	(0.063)
Ever manager	0.029	0.052	0.043	0.125	0.051
	(0.167)	(0.223)	(0.202)	(0.330)	(0.219)
Participation	0.728	0.843	0.729	0.847	0.766
	(0.345)	(0.270)	(0.362)	(0.253)	(0.335)
Full-time	0.770	0.889	0.807	0.850	0.824
	(0.324)	(0.223)	(0.326)	(0.256)	(0.300)
Age at first child	29.58	31.37	30.55	31.68	30.70
	(6.163)	(6.321)	(6.878)	(4.832)	(6.342)
Wealth at age 50	0.198M	0.326M	0.190M	0.679M	0.260M
	(2.004)	(1.670)	(1.497)	(5.036)	(2.105)
Life-time earnings	4.772M	6.315M	5.240M	11.39M	6.124M
	(12.33)	(3.227)	(3.307)	(9.454)	(7.336)
Panel B: Labor Force Survey sample					
Weekly hours > 37	0.172	0.229	0.219	0.332	0.228
	(0.377)	(0.420)	(0.414)	(0.471)	(0.420)
Evening work	0.375	0.428	0.328	0.580	0.402
	(0.484)	(0.495)	(0.470)	(0.493)	(0.490)
Works from home	0.254	0.365	0.298	0.600	0.350
	(0.435)	(0.481)	(0.457)	(0.490)	(0.477)
Works overtime	0.0777	0.108	0.0962	0.158	0.104
	(0.268)	(0.310)	(0.295)	(0.365)	(0.306)

Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. Columns' labels indicate the ambition type, as defined in Section 3.2. Row labels indicate the work-life balance proxy to be considered, defined in the text of this section. The first four columns report averages of individual-level proxies conditioning on each of the four ambition types. The final column reports the same statistics for the sample of couples (defined in Section 2).

least part-time). *Full-time* captures the fraction working at least 32 hours per week.[11] *Age at first child* is the average age among graduates at which the first child is born. *Wealth at age 50* is the average net wealth accumulated at age 50 in Danish Crowns (henceforth DKK).[12] Finally, *Lifetime earnings* is the sum of deflated annual earnings over 30 years after graduation.[13]

Overall, Table 1 documents two key patterns. First, graduates in the most ambitious $(high, high)$ category are more career-focused than graduates in the other categories with lower w_0 or lower g. For example, graduates of marital type $(high, high)$ are penalized in terms of hourly wages for part-time work (inflexibility) and are more likely to participate in the labor market and to work full time, relative to graduates from less ambitious programs. Moreover,

[11] Based on the RAS register. Before 2008 the threshold between part-time and full-time is at 30 hours (1980 to 1992) or 27 hours (1993 to 2007), see (Lund and Vejlin, 2016).

[12] Net wealth excludes assets in pension funds and is deflated with 2000 as the base year.

[13] To compute lifetime earnings, we deflate earnings by running a regression of log annual earnings, for each program separately, on year dummies (base year 2000) and dummies for years since graduation to account for compositional differences by programs in the share of graduates at different life-cycle stages.

Table 2: The Work-Life Balance of Ambition beyond Levels, Fields, and Lifetime Earnings

FE model:	None	Levels	Fields		None	Levels	Fields	
Controls:	*None*	*None*	*None*	*Earnings*	*None*	*None*	*None*	*Earnings*
	(a) Inflexibility				(b) Ever manager			
w_0	-0.009	0.007	0.003	-0.001	0.023	0.025	0.013	0.008
	(0.006)	(0.007)	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.006)
g	0.023	0.021	0.016	0.011	0.023	0.027	0.020	0.020
	(0.006)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)
Mean		1.098		1.081		0.050		0.065
Obs.		985		438		1,837		491
Adj. R^2	0.155	0.316	0.480	0.405	0.409	0.467	0.450	0.529
	(c) Participation				(d) Full time work			
w_0	0.054	0.040	0.031	0.016	0.036	0.098	0.087	0.064
	(0.014)	(0.009)	(0.015)	(0.018)	(0.006)	(0.012)	(0.017)	(0.013)
g	0.025	0.037	0.037	0.038	0.008	0.023	0.022	0.013
	(0.009)	(0.007)	(0.009)	(0.012)	(0.008)	(0.008)	(0.010)	(0.008)
Mean		0.766		0.806		0.820		0.853
Obs.		1,837		491		1,837		491
Adj. R^2	0.235	0.459	0.477	0.335	0.110	0.333	0.309	0.403
	(e) Age at first child				(f) Wealth at age 50			
w_0	0.305	0.603	0.631	0.106	0.134	0.143	0.133	0.121
	(0.266)	(0.366)	(0.363)	(0.400)	(0.012)	(0.016)	(0.016)	(0.013)
g	0.316	0.368	0.435	-0.022	0.095	0.095	0.087	0.073
	(0.173)	(0.186)	(0.218)	(0.200)	(0.012)	(0.013)	(0.015)	(0.014)
Mean		31.51		31.88		0.241M		0.291M
Obs.		1,824		491		1,309		491
Adj. R^2	0.013	0.018	0.025	0.180	0.454	0.456	0.463	0.555

Notes: *FE* stands for *fixed effects*, *Obs.* for *observations*, and *Adj.* for *adjusted.* w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. *Earnings* stands for *Life-time earnings* as defined in the text of this section. Each panel (a) to (f) shows the coefficients on w_0 and g in a regression of the work-life balance proxy (defined in the text of this section), which includes the fixed effects and controls as indicated in the columns' labels. Robust Standard errors in parentheses.

graduates from ambitious programs have their first child on average more than two years later than (low, low) graduates. The strong career-focus of graduates from the most ambitious programs is associated with substantially higher life-time earnings and levels of wealth. Using survey data from the LFS introduced in Section 2, Panel B further shows that graduates from the most ambitious programs are more likely to work long hours, that is, more than 37 hours per week (the union-bargained standard work week in the Danish context), and they report working overtime more frequently. Moreover, they work irregular hours, e.g., in the evenings (conditional on not working in shifts), and they are more likely to work from home.

Second, conditioning on one dimension—i.e, either on w_0 or on g—graduates from programs with a higher value of the other dimension are more career focused. For example, graduates of ambition type ($high, low$) are more likely to work, become managers, work full-time, and delay fertility than graduates of ambition type (low, low). This pattern is verified for any pair-wise comparison of ambition types and for all proxies.

To analyze these patterns further, we use regression analysis and show in Table 2 that the two building blocks of our ambition types—the labor market outcomes w_0 and g—jointly explain each work-life balance proxy, even when comparing graduates within the same education level, within the same field of study, and conditional on lifetime earnings. The table shows the coefficients from regressions of the first six proxies (we control for the seventh in some specifications) from Table 1 on w_0 and g. *FE model* labels specifications according to whether they include fixed effects (FE) and, if so, at what level. *Controls* identifies specifications in which we additionally control for lifetime earnings. FE models in columns labeled *None* compare the mean signaling factors across programs of different starting wages and wage growth. FE models in columns labeled *Levels* and *Fields* compare programs within the same level and field of education, respectively. Finally, models in columns with controls labeled *Earnings* include lifetime earnings as defined in Table 1 as a control.

For all proxies, we find that higher starting wages and higher wage growth are associated with more career focus, which affects the work-life balance negatively. This is true even within the same levels or fields of study. For example, a one standard deviation change of wage growth (recall that w_0 and g are standardized) implies that the inflexibility measure (part time penalty) increases by 2.3% relative to the mean inflexibility across programs. Within the same level or field of education, the effects become somewhat smaller (2.1% and 1.6%, respectively) but remain highly significant.

Similarly, we find that higher starting wages or wage growth are associated with higher labor force participation and full-time work, as well as higher chances of manager promotion and higher net wealth, but also higher age at first child. Interestingly, the results often become stronger within level and/or field, for example for full-time work, manager promotions, and age at first child. These patterns emphasize important heterogeneity in labor market and family outcomes across programs that our ambition types can capture within educational levels, fields, and conditional on life-time earnings.

Moreover, the fourth specification in each model shows that starting wages and wage growth correlate with our measures of work-life balance even conditional on discounted lifetime earnings. The main takeaway from this result is that collapsing the multi-dimensional labor market trajectories of individuals into a one-dimensional measure of earnings misses the interplay between labor market starting conditions and growth trajectories. Different combinations of starting wages and wage growth can have opposite implications for work-life balance. Consider that the same level of lifetime earnings can be reached through a high starting wage and relatively low growth or a low starting wage and relatively high growth. Indistinguishable by means of lifetime outcomes, these different combinations imply, e.g., different working hours (recall Table 1). Thus, they have different values from the perspective of marriage and family and should be separate marital types. Capturing this heterogeneity in the family value of different career types is the main advantage of our ambition types.

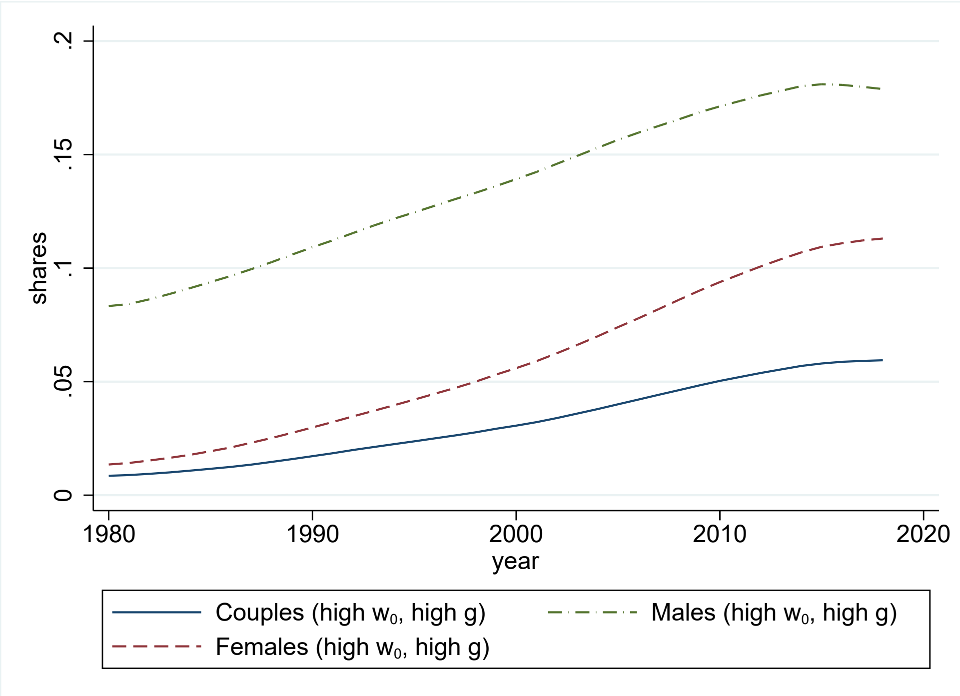
4 The Measurement of Sorting

We now turn our attention to measuring marital sorting based on the new ambition types. 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 spousal types among couples (the contingency table). Determining whether sorting patterns have changed over time has proven elusive because the marginal distributions of types in the marriage market have changed as well.

Figure 2 shows that the population share (solid blue line) of couples in which both spouses are in the top ambition category increased between 1980 and 2018, from around 1% to more than 6%. However, this observation alone 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 men who graduated from ambitious programs increased from around 8% to just below 20% (green dash-dotted line); the share of women who graduated from ambitious programs was very low in 1980 (around 2%) and more than quintupled until 2018 (around 11%). The increasing "supply" of highly-ambitious individuals, and particularly women, mechanically increases the probability that two high-type spouses meet and form a couple. Therefore, a more refined measurement strategy is necessary.

We use a sorting measure that directly takes changing marginal type distributions into account: the likelihood ratio (see also Eika et al., 2019; Chiappori et al., 2020b). This measure captures marital sorting by comparing the observed probability that a man of a given type is married to a woman of the same type to that probability under random matching. Assume that every couple consists of two individuals that are characterized by $j = t_m$ and $j' = t_f$, where m and f indicate gender, so j (j') represents the male (female) characteristics. The one-dimensional types $t \in \{1, \ldots, j, \ldots, T\}$ are the categories defined in Section 3.1. T is the

Figure 2: High-Type Couples and Marginal Distributions



Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. Dashed lines plot the shares of males and females who graduated from educational programs in the high-w_0 and high-g marital type as defined Section 3.2. The solid line plots the share of couples in which this is true for both spouses.

total number of categories. For each combination of male and female types, the likelihood ratio is defined as follows:

$$s(j, j') = \frac{P(t_m = j, t_f = j')}{P(t_m = j)\, P(t_f = j')}. \tag{1}$$

This likelihood ratio relates the observed frequency of a couple type (the numerator) to the expected frequency under random matching (the denominator), which is the product of the shares of men and women in the respective category. Thus, the denominator reflects the marginal distributions. For couples with two spouses of the same type ($j = j'$), a likelihood ratio above one implies PAM.

To compute an aggregate sorting measure, we focus on the likelihood ratios for same-type couples. We aggregate by summing across all categories in which the male and female types are identical ($j = j'$), and use the weights $\{\pi_1 \dots \pi_T\}$ for the respective categories. This yields the following sorting measure:

$$\mathcal{S} = s(1,1) \times \pi_1 + s(2,2) \times \pi_2 + \cdots + s(T,T) \times \pi_T. \tag{2}$$

This weighted sum of likelihood ratios fulfills the formal criteria for sorting measures outlined by Chiappori et al. (2020b, 2021). The weights are meant to capture the relative importance of sorting within different couple-type combinations for aggregate sorting. They depend on the shares of males and females of the respective type and, thus, can be used to compensate for changing type distributions.[14] We follow Eika et al. (2019) and construct the weights based on the expected frequencies under random matching:

$$\pi_j = \frac{P\left(t_m = j\right) P\left(t_f = j\right)}{\sum_{k=1}^{T} P\left(t_m = k\right) P\left(t_f = k\right)}. \tag{3}$$

These weights sum to one, so the aggregate sorting measure $\mathcal{S}$ can be understood as a weighted average of the likelihood ratios for all couples in which male and female types coincide.

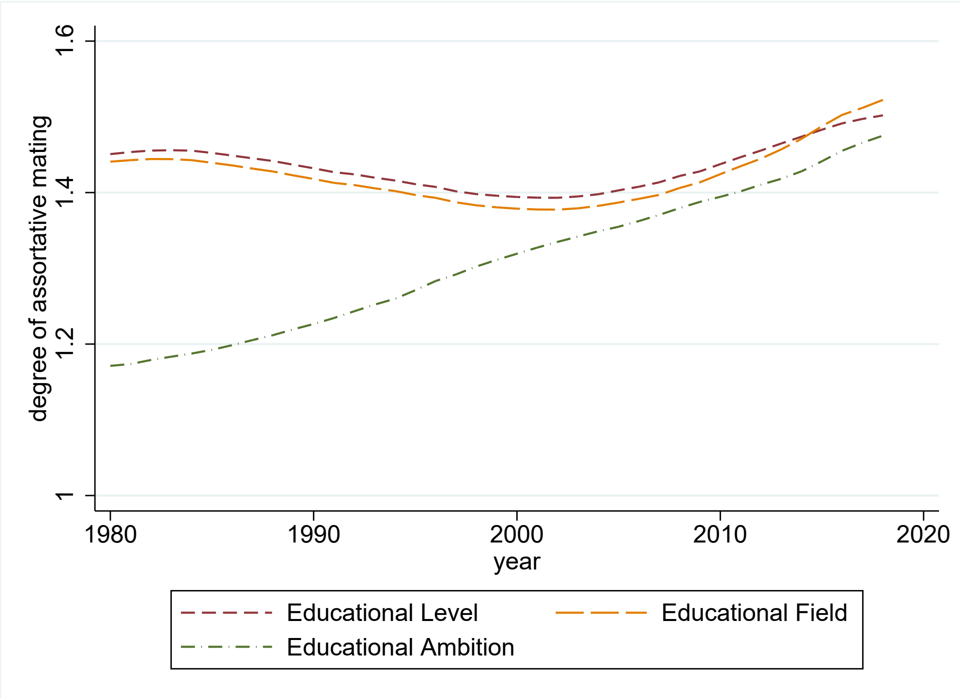
Figure 3 shows how the aggregate sorting measure $\mathcal{S}$ evolved between 1980 and 2018 for educational level types (short-dashed red), educational field types (long-dashed orange), and educational ambition types (dashed-dotted green). Overall, sorting is positive, as all three measures are consistently greater than one. Based on educational level and educational field types, we find that sorting has hardly increased since 1980, consistent with Eika et al. (2019) for the US. Based on the educational ambition types, however, we find a strong increase in sorting. Relative to random matching, the likelihood of observing couples with the same educational ambition type has increased from below 1.2 to approximately 1.5.

The striking difference in the sorting trends for educational field and level on the one hand and educational ambition on the other is explained by the different evolution of within-category likelihood ratios and marginal type distributions (plotted in Figures C.1 and C.2 in the Online Appendix, respectively). In the educational level categorization, the "secondary" category is large and relatively stable for both men and women (Figure C.2a and b). The likelihood ratio for 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 short-dashed line in Figure 3. PAM 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 field, the picture is quite similar. Recall that the primary and secondary categories are identical to the educational level categorization. For men, the composition of

[14] Chiappori et al. (2020b) suggest that the weights can in general be thought of as a convex combination of the shares of males and females with the same level of education. Almar and Schulz (2023) provide a discussion of different weighting strategies employed in the literature and show how the measurement distortion due to changing type distributions can be minimized.

Figure 3: Increasing Marital Sorting based on Ambition Types



Notes: The figure shows the sorting measure $\mathcal{S}$ derived in Section 4, equation (2) for educational level types (red short-dashed line), educational field types (orange long-dashed line) and educational ambition types (green dash-dotted line). Types are constructed as explained in Section 3.2.

graduates across fields is relatively stable. For women, the fields "Business", "STEM", and "Social Sciences" became more important over time but "Health" and "Humanities" still dominate. Sorting within fields is positive, albeit decreasing over time (Figure C.1c), which is consistent with Eika et al. (2019). The share of individuals with post-secondary degrees increases while the share with primary education falls, so the aggregate sorting measure for educational fields remains stable as well.

For educational ambition sorting, the picture is different. In terms of likelihood ratios, we obtain a clear distinction between the group with high starting wages and wage growth, in which PAM is pronounced (but falling over time), and the other three educational ambition categories with very little PAM, see Figure C.1b. 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 3.

5 Marriage Market Sorting and Inequality

To study the link between changes in marriage market sorting and changes in inequality, we apply a semi-parametric decomposition technique first proposed by DiNardo, Fortin and Lemieux

(1996) and recently implemented to study household income inequality in Eika et al. (2019). We consider three counterfactual scenarios: (i) fixed marriage market sorting; (ii) fixed labor market returns to educational type; and (iii) fixed educational-type composition. For each scenario, we study how changes in the respective dimension contribute to rising inequality between 1980 and 2018. We then compare these contributions under alternative categorizations of marriage market types, based on levels, fields, or ambition types. To measure between-household inequality, we use the Gini coefficient as an overall measure. Additionally, we zoom in on the upper and lower halves of the income distribution using percentile ratios.

We implement the counterfactual scenarios as in Eika et al. (2019) by constructing a stochastic matching algorithm that re-matches married individuals. The algorithm samples pairs of potential spouses from the male and female type distributions and forms new couples based on type-dependent matching probabilities p, which are derived from the likelihood ratios $s(j, j')$ defined in Equation (1).[15] It is important to note that our re-matching algorithm abstracts from sorting within couple-type cells. Yet, randomly re-matching couples within cells closely matches the observed overall inequality across households for all three alternative categorizations of marriage market types (we provide evidence of our algorithm performance in Online Appendix B and Table B.1). This suggests that the vast majority of sorting relevant for income inequality occurs across couple types.

(i) Fixed marriage market sorting

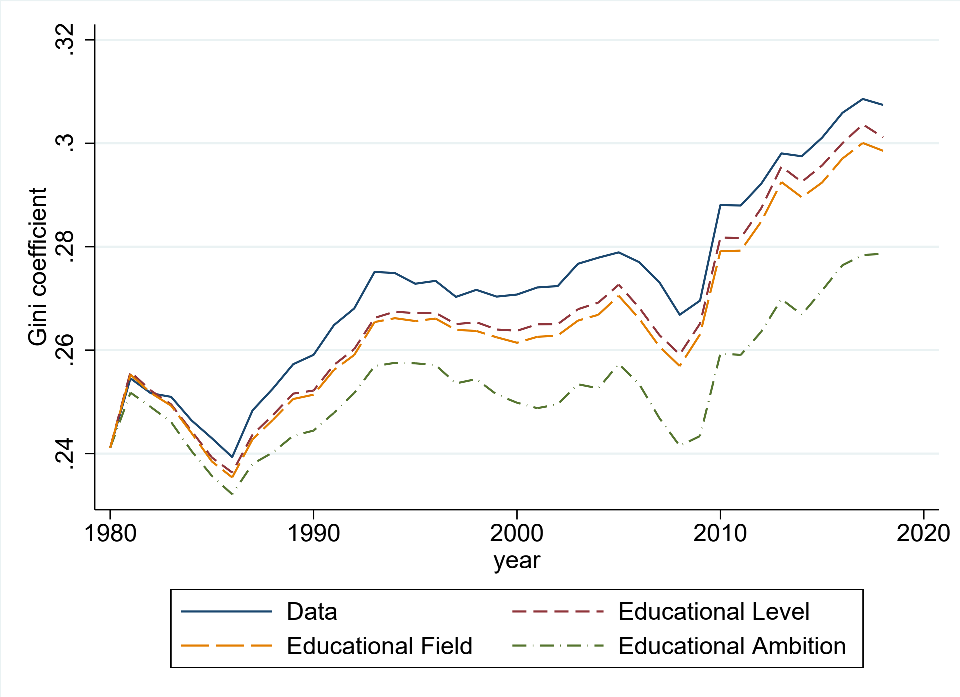
We construct scenario (i) by re-matching couples based on the matching probabilities and marginal distributions from 1980. That is, we fix both the likelihood indices (1) and the marginal type distributions.[16] We construct counterfactual inequality measures for all years $\tau \in (1981, 1982, \dots, 2018)$ 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 4 shows how much of the increase in between-household inequality can be explained by increasing sorting for all years τ based on the three categorizations that we compare: educational level, educational field, and educational ambition. Overall, we see that inequality would

[15] To calculate $p \in [0, 1]$, we divide the likelihood ratio $s(j, j')$ (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 $(t_m = j, t_f = j')$ 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.

[16] Our goal is to keep the aggregate sorting measure fixed, and, as we have discussed in Section 4, changing marginal distributions contribute to aggregate sorting. We isolate the effect of changing marginal distributions on inequality in scenario (iii).

Figure 4: Growth in Educational Ambition Sorting Amplified Inequality



Notes: 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 explained in this section) for educational level (red short-dashed line), educational field (orange long-dashed line), and educational ambition (green dash-dotted line) types.

have increased less with fixed sorting because all counterfactual trends are below the trend in the data (solid blue line). However, the red short-dashed and the orange long-dashed trends that hold educational level and educational field sorting fixed, respectively, are much closer to the data than the green dashed line for educational ambition sorting. That is, increasing positive sorting by educational ambition amplifies between-household inequality, while educational level and educational field sorting contribute relatively little. The reason is that sorting has not increased according to these categorizations, recall Figure 3.

Table 3 shows the same finding numerically and links the implied change of the Gini coefficient relative to the data to the sorting trend for each categorization. For educational ambition types, the gap between the actual and the counterfactual Gini coefficient in 2018 amounts to 43%. That is, had the sorting patterns remained the same as in 1980, between-household inequality would have increased by only 57% of the actual increase in the data. Over the same period, the extent of educational ambition sorting increased by 26%. Conversely, for types based on educational levels (fields), sorting increased by only 4% (6%). Therefore, holding sorting fixed reveals a much smaller contribution to increasing between-household inequality. For educational level (field) types, the gap between the actual and the counterfactual Gini

Table 3: Linking the Contribution to Rising Inequality to the Sorting Trend

	N (1,000s)		Sorting			Gini, data		Gini, (i)	$\frac{\Delta_{Gini,(i)}}{\Delta_{Gini,data}}$
	1980	2018	1980	2018	Change	1980	2018	2018	
Educational Level	1,758	1,653	1.45	1.50	4%	0.241	0.307	0.301	91%
Educational Field	1,758	1,653	1.44	1.52	6%	0.240	0.307	0.299	87%
Educational Ambition	1,758	1,653	1.17	1.48	26%	0.241	0.307	0.279	57%

Notes: Columns *N (1000s)* 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 4. *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 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. Each row shows the columns' statistic for one of the three definitions of marriage market types we construct (as explained in Section 3): educational level, educational field, and educational ambition types.

coefficient in 2018 amounts to 9% (13%). This finding is qualitatively similar to Eika et al. (2019), who find even smaller gaps for the US.

In Table 4, we decompose the total change in between-household 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}). Between-household income inequality has increased according to all three measures. The Gini coefficient has increased by 0.066 (from 0.241 to 0.307, recall Table 3), the 90/50 percentile ratio has increased by 0.165 (from 1.523 to 1.688), and the 50/10 percentile ratio has even increased by 0.573 (from 1.944 to 2.518). These changes correspond to 100%.

For the 90/50 ratio in Panel (i)-column (b), fixed educational level (and field) sorting leads to counterfactual inequality measures that are even slightly above the benchmark (110% and 103%). That is, sorting based on educational levels and fields 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), sorting on all three categories suggests an amplification of the inequality trend, but the amplification is more than twice as strong for educational ambition types than for educational level or fields.

(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 labor mar-

Table 4: Decomposing 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 Field	0.057	87%	0.170	103%	0.383	67%
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 Field	0.003	5%	0.092	56%	-0.059	-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 Field	0.091	137%	0.184	112%	1.711	298%
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 Field	0.058	88%	0.099	60%	0.726	127%
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 Field	0.092	138%	0.213	129%	1.102	192%
Educational Ambition	0.067	101%	0.146	89%	0.633	110%

Notes: The table shows changes in inequality between 1980 and 2018 in the data and for each of the counterfactual scenarios constructed and discussed in this section. Column (a) reports the Gini coefficient, while Columns (b) and (c) report the ratio of the 90th and 50th percentile ($P90/P50$) and the ratio of the 50th and 10th percentile ($P50/P10$) in the income distribution. The first row labeled Δ_{Data} shows the inequality changes in the 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., $\Delta_{Gini}/\Delta_{Data}$. Each row withing each counterfactual case shows the columns' statistic for one of the three definitions of marriage market types we construct (as explained in Section 3): educational level, educational field, and educational ambition types.

ket returns:

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

where subscript y denotes household income, subscript x the couple-type combination ($t_m = j, t_f = j'$), and subscript p the matching probabilities in the respective calendar year.[17] The

[17] 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.

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 re-matching 1980 couples based on 2018 matching probabilities. This allows us to compute the conditional probabilities in the numerator and denominator of the first factor on the right-hand-side. 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 labor market returns to educational types are the major source of increasing income inequality; see Panel (ii) of Table 4. Without changing returns, the increase in the Gini is only 15% of the true increase for educational level types, 11% for educational ambition types, and 5% for educational field 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 construct marital 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 field) types, the 90/50 ratio still increases by 77% (56%) relative to the data. That is, increasing returns amplified inequality less in the upper half of the distribution than overall, but changing returns across fields of study are more important than changing returns to broadly defined tertiary education. While changing returns to "ambitious" education programs can explain an even larger share—about half—of increasing inequality, the results for educational fields and educational ambition types are relatively similar in the upper part of the distribution. In the lower part of the income distribution (50/10 ratio), we find that absent increasing returns to education, inequality would have decreased by 5% (educational ambition types) to 10% (educational field and -level types).

(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_m and t_f. In this case, the reweighting factor is $\widehat{\psi_x} = (\widehat{\psi_y})^{-1}$ using equation (4) above.

First, we keep the type distributions for both genders fixed at the 1980 level, see Panel (iii) in Table 4. We then repeat the exercise keeping either the male or the 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.[18] That is, there are 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 (or fields), we find that increasing educational attainment had a mitigating effect on inequality. Without the shift, inequality would have increased to 142% (137%) 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. For the 90/50 ratio in column (b), we only see a modest mitigating effect for educational level and field types (119% and 112%). 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 that top and bottom educational ambition categories are distinct in terms of wage growth while the top and bottom educational level and field categories are not. With ambition types, inequality rises as the number of individuals in the top category increases, but this is not true for educational level and field types due to the large heterogeneity in wage growth within the tertiary categories.

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 4 are similar for all categorizations. Only the mitigating effect in the lower half of the distribution is less pronounced, especially for educational levels and fields. This implies that changes to the female marginal distributions drive the mitigating effects. 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 three definitions of marital types becomes evident from the distinct effects that changing marginal distributions had on between-household inequality. For example, while women's move 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. Differences across categorizations for the male type distributions are much smaller.

[18] 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), as shown in Figure C.2 in the Online Appendix.

6 Robustness

In this section, we assess the sensitivity of our main results to alternative ways of constructing marriage market types based on educational ambition. Using the notation from the conceptual framework in Section 3.1, recall that our benchmark categorization creates four types using information at the level of the educational program i, the sub-vector of characteristics $\tilde{x}_i = (w_{0i}, g_i)$, and k-means mapping algorithm $\mathcal{T}_{Ambition}(\tilde{x})$.

Table 5—which has the same structure as Table 3—shows the sorting trends and inequality contributions in our fixed-sorting counterfactual (analyzed in Section 5) and the data based on different ways of constructing ambition types. For convenience, the first row repeats results for our benchmark categorization of ambition types from the third row of Table 3.

In the rows labeled *Types by gender*, we consider the possibility that programs of education may send different signals depending on the gender of the graduate. We construct four ambition types for women and four for men by using information at the program level i and sub-vector of characteristics (w_0, g) but mapping programs to types separately by gender. Formally, we consider the sub-vector of characteristics $\tilde{x}_i^f = (w_{0i}^f, g_i^f)$ for women and $\tilde{x}_i^m = (w_{0i}^m, g_i^m)$ for men and create female and male ambition types through k-means mappings $\mathcal{T}_{Ambition}(\tilde{x}^f)$ and $\mathcal{T}_{Ambition}(\tilde{x}^m)$, respectively. Online Appendix Figure C.3 (which has the same structure as Figure 1) shows that our method successfully generates four types clearly distinct in terms of labor market prospects (as is the case for our benchmark). Even though most of the biggest programs are assigned to the same ambition type for men and women, there are exceptions. For example, architecture is a program associated with a high wage growth type for women but a low wage growth type for men. Table 5 shows that sorting based on gendered ambition types increased slightly less than in our benchmark (but still significantly more than sorting by levels or fields) and explains slightly less of the changes in inequality than our benchmark (but still significantly more than when considering levels or fields).

Similarly, in the row labeled *Types by cohort*, we construct ambition types by cohorts of graduates defined by decade (individuals who graduated before 1990, between 1990 and 2000, and after 2000). In this robustness exercise, we account for the possibility that the signaling value of degrees may change over time (similar to the Goldin (2014) argument that occupations have evolved over time). We define three sub-vectors of characteristics by graduation cohort, $\tilde{x}_i^{80} = (w_{0i}^{80}, g_i^{80})$, $\tilde{x}_i^{90} = (w_{0i}^{90}, g_i^{90})$, and $\tilde{x}_i^{00} = (w_{0i}^{00}, g_i^{00})$, and map programs to types by cohort using the k-means algorithm. While Online Appendix Figure C.4 shows that many large programs are remarkably stable in their characteristics over time, we detect some changes. For example, while an ordinary high school diploma is categorized as a type with low starting wage and high growth early in the sample, these growth opportunities decline over time and the degree moves into the low-low category. Other changes are based on shifts in relative pay levels, which can also revert back. For example, preschool teachers are classified as low-low in the first and last part of the sample, but fall into the high starting wage and low growth category in the 1990s. Our main conclusions regarding the relationship between trends in sorting and inequality are unchanged when constructing the ambition types by cohort.

The row labeled *Sub-field level* uses a level of observation that is more aggregate than the granular education programs used for our benchmark categorization. We aggregate programs by levels and sub-fields of study and define $i' =$ levels $\times$ sub-fields as our unit of observation. Specifically, we consider 48 observation units that we obtain by subdividing each of the four educational levels by sub-field of study. Essentially, sub-fields are a more detailed version of the fields of study used above and in the literature that go across educational levels.[19] As our clustering variables, we use the same sub-vector as in our benchmark, $\tilde{x}_{i'} = (w_{0,i'}, g_{i'})$. The results only slightly change. Sorting increased by about 22% and explains 40% of the increasing inequality between households.

Finally, in the last two rows, we show our main analysis when we define three or five ambition types instead of four, using again information at the level of educational program i, the sub-vector of characteristics $\tilde{x}_i = (w_{0i}, g_i)$, and the k-means clustering algorithm. Once again, our main conclusions remain the same. While more categories allow us to detect an even stronger increase in marriage market sorting over time, these differences are inconsequential for the contribution to rising inequality. This suggests that the sorting changes that contribute to inequality mostly occur between broad groups at the top and bottom.

What all these alternative categorizations have in common is that they use a sub-vector of characteristics that correlates with both the earnings potential and the work-life balance attached to the unit of observation, as shown in Table 2. Our conclusion that sorting on career ambition has increased over time and that it contributes significantly to the rise in between-household income inequality is robust to constructing ambition types differently.

[19] For example, we consider the STEM sub-fields "Construction" and "Mechanics & Metal" and further distinguish programs within this sub-field by the required level of schooling for entry into the programs, i.e., high school (secondary programs) and college/university (tertiary programs).

Table 5: Scenario (i)—Fixed Sorting—with Alternative Ambition Types

	N (1,000s)		Sorting			Gini, data		Gini, (i)	$\frac{\Delta_{Gini,(i)}}{\Delta_{Gini,data}}$
	1980	2018	1980	2018	Change	1980	2018	2018	
Benchmark	1,758	1,653	1.17	1.48	25.9%	0.241	0.307	0.279	57%
Types by gender	1,757	1,651	1.05	1.27	21.0%	0.241	0.307	0.286	68%
Types by cohort	1,742	1,651	1.16	1.50	29.4%	0.240	0.307	0.284	65%
Sub-field level	1,854	1,630	1.19	1.45	21.8%	0.243	0.304	0.279	60%
Three types	1,756	1,653	1.16	1.31	12.7%	0.241	0.307	0.281	60%
Five types	1,756	1,653	1.20	1.58	32.1%	0.241	0.307	0.281	60%

Notes: Columns *N (1000s)* 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 4. *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 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. Each row shows the columns' statistic for one of the alternative definitions of marriage market types we construct for this robustness analysis and explain in this section.

7 Conclusion

We provide new insights into the relationship between education-based marriage market sorting and between-household inequality. We show that the way in which data on education are used to capture the traits that are relevant for marriage market matching influence conclusions about this relationship.

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 both the earnings potential and future time commitments to career and family (work-life balance) of individuals on the marriage market, ambition types are better suited to study marriage market sorting and its effect on inequality than categorizations based on educational levels and fields of study.

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 and field of education remains 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 a crucial choice.

Our second main result reveals that changes in who marries whom in terms of educational ambition had a large 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, between-household inequality growth would have been mitigated by approximately 40%. In contrast, marriage market sorting trends based on education level and field of study 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 main findings are robust to alternative categorizations of ambition types, in particular grouping programs separately by gender or by decades. Furthermore, the results are similar when defining educational ambition based on more aggregate levels of observation such as subfield of education. What is crucial for the robustness is that the variables used to group units of observations capture both earnings potential and work-life balance.

Overall, our analysis suggests that considering richer type classifications than the level or field of education can be a promising direction for future research on marriage markets. A number of administrative data sources across countries provide long panel data with program-level information that allows researchers to implement our baseline approach to define marital types. On top of this, our robustness analysis with more aggregated units of observation also suggests a path for applying our insights with survey data sources that provide coarser information about educational attainment, so long as those data include detailed labor market information from graduates.

References

Almar, Frederik and Bastian Schulz (2023) "Optimal Weights for Marital Sorting Measures," Discussion Paper 16368.

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.

Artmann, Elisabeth, Nadine Ketel, Hessel Oosterbeek, and Bas van der Klaauw (2021) "Field of study and partner choice," *Economics of Education Review*, Vol. 84, p. 102149.

Bailey, Martha J. and Peter Z. Lin (2022) "Marital Matching and Women's Intergenerational Mobility in the Late 19th and Early 20th Century U.S," unpublished.

BEF, Statistics Denmark (1980–2018) "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.

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.

Binder, Ariel J, Caroline Walker, Jonathan Eggleston, and Marta Murray-Close (2022) "Race, Class, and Mobility in US Marriage Markets."

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 (2023) "The Effects of Institutional Gaps between Cohabitation and Marriage," unpublished.

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

Chiappori, Pierre-Andre, Monica Costa Dias, and Costas Meghir (2021) "The Measuring of Assortativeness in Marriage: a Comment."

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir (2018) "The Marriage Market, Labor Supply, and Education Choice," *Journal of Political Economy*, Vol. 126, pp. S26–S72.

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.

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.

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

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

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

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

Foerster, Hanno, Tim Obermeier, Bastian Schulz, and Alexander Paul (2022) "Job Displacement, Remarriage, and Marital Sorting," unpublished.

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

Gayle, George-Levi, Limor Golan, and Mehmet Soytas (2015) "What Accounts for the Racial Gap in Time Allocation and Intergenerational Transmission of Human Capital?", unpublished.

Gayle, George-Levi and Andrew Shephard (2019) "Optimal Taxation, Marriage, Home Production, and Family Labor Supply," *Econometrica*, Vol. 87, pp. 291–326.

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.

Han, Siqi and Yue Qian (2022) "Assortative Mating Among College Graduates: Heterogeneity Across Fields of Study," *Marriage & Family Review*, Vol. 58, pp. 647–673.

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.

IDAN, Statistics Denmark (1980–2018) "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.

IND, Statistics Denmark (1980–2018) "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.

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.

LFS, Statistics Denmark (2000–2018) "AKU - Arbejdskraftundersøgelsen, Statistics Denmark: Labour and Income, Social statistics.," URL: https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/labour-force-survey--lfs-, Data accessed through the ECONAU project database.

Lindenlaub, Ilse (2017) "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, Vol. 84, pp. 718–789.

Low, Corinne (Forthcoming) "The Human Capital - Reproductive Capital Tradeoff in Marriage Market Matching," *Journal of Political Economy*.

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 Inter-generational 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.

Reynoso, Ana (2023) "The impact of divorce laws on the equilibrium in the marriage market.," June, unpublished.

Seiver, Daniel Alan and Dennis H Sullivan (2020) "A study of major impact: Assortative mating and earnings inequality among US college graduates," *Atlantic Economic Journal*, Vol. 48, pp. 237–252.

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

UDDA, Statistics Denmark (1980–2018) "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.

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

Educational Ambition, Marital Sorting, and Inequality

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

A Additional details on our data

A.1 Data sources

All registers used are yearly population-wide data sets. We use all the persons living in Denmark at the end of a year from 1980-2018, whom are observed in the data sets PERSONER and BEF. These data sets contain yearly information on age, partner ID, municipality, gender, civil status, and number of children. We merge these data sets with additional information as follows. We measure the highest achieved education from the education register (UDDA). From the income register (IND), we measure income and wealth information, and the hourly wage earned in the primary job held in the last week of November each year.[20]. The datasets EXPYEAR and IDAP provide information on real labor market experience. Finally, we use information from the registers RAS and AKM in order to get occupational information and a part-time/full-time indicator. When available we also merge information from the Labor Force Survey in order to get more details on flexibility and hours worked than what is available from the registers. The Labor Force Survey covers the years 2000 to 2018. We keep information on individuals aged 19 to 60.

A.2 Definition of key variables

Income measure Our income variable, ERHVERVSINDK_13, measures all earned income during a year, where earned income is defined as income from employment and self-employment.

Hourly wage 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).

We run a regression of log hourly wages on year dummies and educational specific experience

[20] We rank job types and keep the highest rank available in JOB_TYPE: H, 3, A, S, M. The variable for the hourly wage is TIMELON prior to 2008 and SMAL_TIMELOEN in 2008 and after.

profiles in order to take into account an aging population and differences in educations over time. We then subtract the coefficients on the year dummies (2000 base) from the log hourly wages. This gives us our residualized log hourly wages, which is what we use for the analysis.

Education We find the highest completed education of the individuals when they are the oldest (if not reaching 35 in the data) or when they are age 35. This is the program and year of graduation we use as their (final) educational program.

Educational programs As a point of departure each educational program is an ISCED code. However, in a few places we change the definition slightly. In the start of the sample we have a group of individuals who have only 7th or 8th grade, because compulsory schooling ended in 7th grade until around 1960. We pool 7th-9th grade into one group called 9th grade. We split up both 9th grade (1109) and 10th grade (1110) up into 5 sub programs each based on region of graduation.[21] Finally, some individuals have an older code for high school than those graduating in 1980 and later. We assume that the high school education did not change much and we use starting wages and wage growth for the new high school education for those who graduated with the old code prior to 1980.

Starting wages and growth Starting wages are the average of the log hourly residualized wages in years 1-5 after graduation. The growth is calculated based on the difference between the average in years 9-11 and the starting wages. This is done for each individual in the sample (both singles and couples). In order to get information on the program level we average across individuals, but condition on individuals who graduated in 1980 or later, for whom we observe both for starting wages and having wages in some of the years 9-11 after graduation. We also only use information from individuals, whose wage growth is below the 99th percentile (extreme values are likely due to measurement error).

With this in place, we can standardize starting wage and growth. All individuals in the data have been assigned the average values from their final program. We generate the standardized variables by subtracting the mean and dividing by the standard deviation.

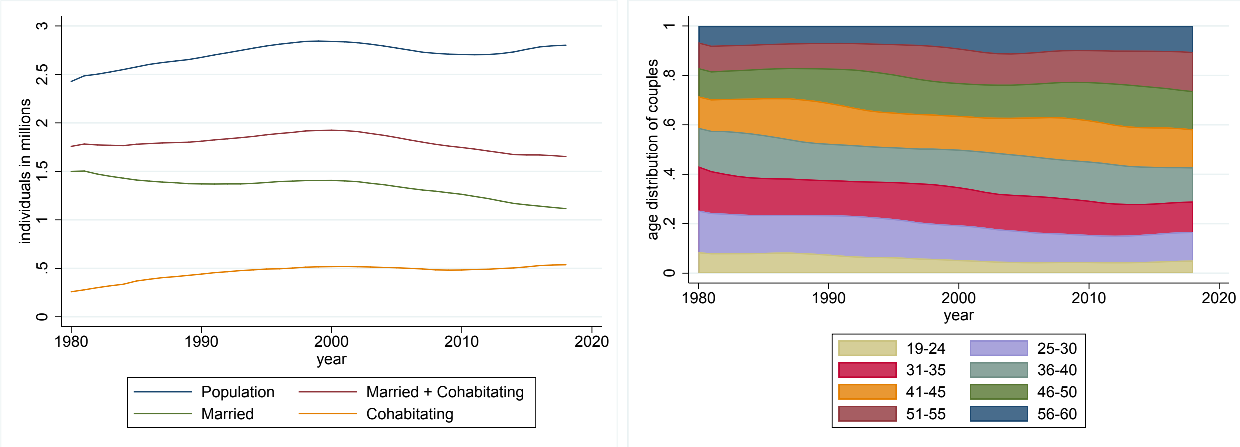
Next, we construct our four ambition types by using k-means clustering on the standardized starting wages and growth. All individuals are still in the data set, but because everybody from the same program has the same value, we are grouping at the program level.

[21] In particular we do the following. Use 9th grade (1109)(split by region) starting wages and growth for the following codes (all also split by region depending on where the individual lives in the first year we see them in the data, e.g., 1980): 1007,1008,1023,1123,1009,1022. Use 10th grade (1110) (split by region) for: 1010.

A.3 Descriptive statistics

Figure A.1: Marriage, Cohabitation, Age Composition

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



Notes: 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.

Table A.1: Basic Descriptive Statistics for Educational Ambition Types

Category (w_0, g)	(low, low)	(high, low)	(low, high)	(high, high)	Population
Population share	20.2%	22.7%	47.5%	9.7%	100.0%
Female share	64.8%	31.0%	56.0%	33.4%	50.0%
Starting wage	4.841	5.015	4.728	5.181	4.860
	(0.0613)	(0.0775)	(0.0488)	(0.134)	(0.170)
Wage growth	0.0807	0.118	0.211	0.301	0.172
	(0.0339)	(0.0436)	(0.0574)	(0.0756)	(0.0862)
Parental wealth at graduation	401347.0	664844.4	269760.8	1189937.8	474762.7
	(259668.7)	(1609532.9)	(307755.7)	(353775.9)	(858804.7)
Wage growth SD	0.323	0.298	0.430	0.365	0.359
	(0.0682)	(0.0536)	(0.0946)	(0.0731)	(0.0945)

Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The four first columns report averages of individual-level descriptive statistics for each of the four educational ambition types identified in Section 3. 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. Parental wealth at graduation is computed as the sum of both parents' net wealth in the year in which the individual graduates from the most advanced educational program. Deflated with base year 2000. Standard deviations in parentheses.

Table A.2: Educational Levels, Fields, and Ambition Types

Category (w_0, g)	(low, low)	(high, low)	(low, high)	(high, high)	Population
Educational Level					
Primary	8.3%	0.5%	56.2%	0.2%	28.5%
Secondary	66.2%	57.3%	40.1%	10.3%	46.4%
Tertiary	24.9%	42.1%	8.2%	89.3%	24.9%
Educational Level within Tertiary					
Bachelor	24.1%	29.4%	3.1%	30.3%	15.9%
Master & PhD	0.8%	12.7%	0.5%	59.0%	9.0%
Educational Field within Tertiary					
Humanities	2.2%	18.0%	1.2%	2.7%	5.4%
Social Science	0.1%	3.0%	0.5%	16.4%	2.5%
Business	0.3%	0.5%	0.3%	21.4%	2.4%
STEM	0.2%	3.9%	0.2%	34.3%	4.4%
Health & Welfare	18.5%	12.3%	1.1%	11.6%	8.2%
Other	3.7%	4.4%	0.3%	3.0%	2.2%

Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The four first columns and the first panel report population shares for each of the four educational ambition types identified in Section 3 across educational levels. We further subdivide the tertiary shares into Bachelor and Master/PhD as well as post-secondary fields of study. The final column reports the all shares for the entire population of couples as defined in Section 2.

B Matching Algorithm Performance

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 B.1 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-

Table B.1: 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 Field	0.293	95%	1.678	99%	2.203	87%
Educational Ambition	0.295	96%	1.690	100%	2.189	87%

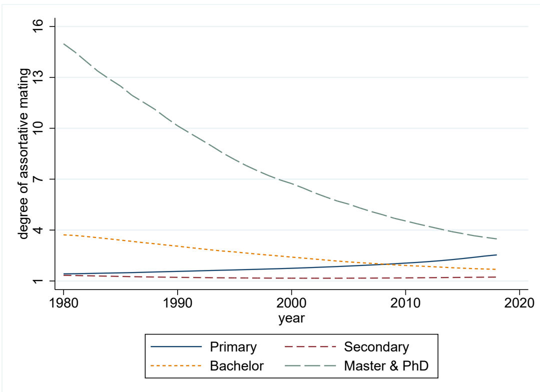
Notes: The table reports results of the sensitivity analysis of the matching algorithm discussed in Online Appendix Section B. We rematch couples randomly in 2018 within couple-type-combination cell and check how well the empirical inequality measures are reproduced. Panel (a) reports the Gini coefficient, 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.

dimensional matching algorithm produces reliable counterfactual marriage market allocations.

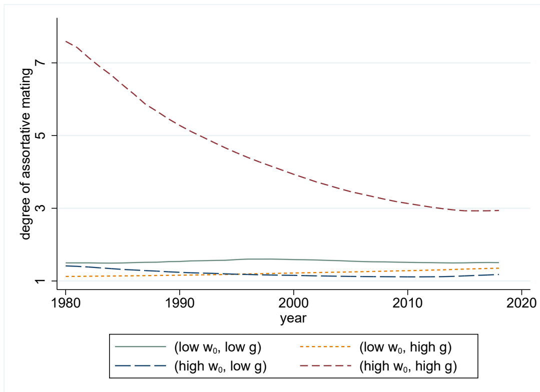
C Additional Results

Figure C.1: Likelihood Indices

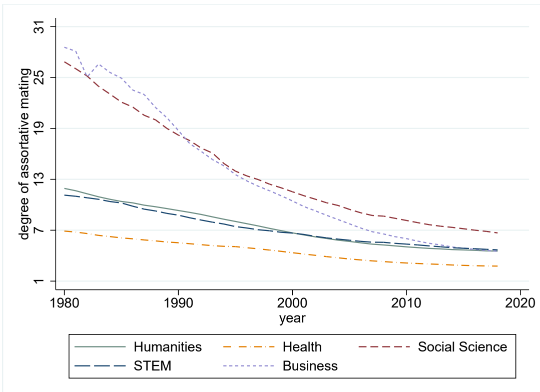
(a) Educational Level



(b) Educational Ambition



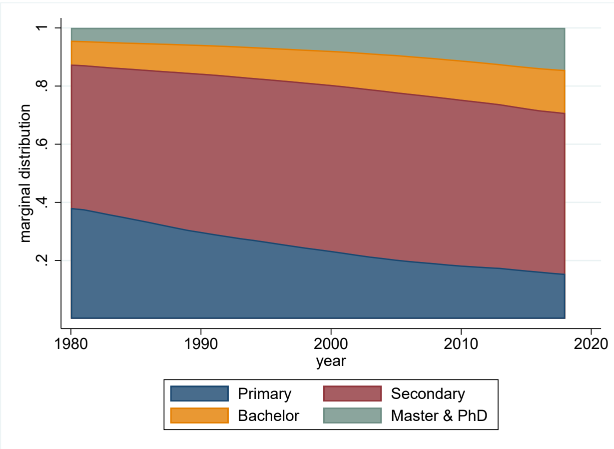
(c) Educational Field



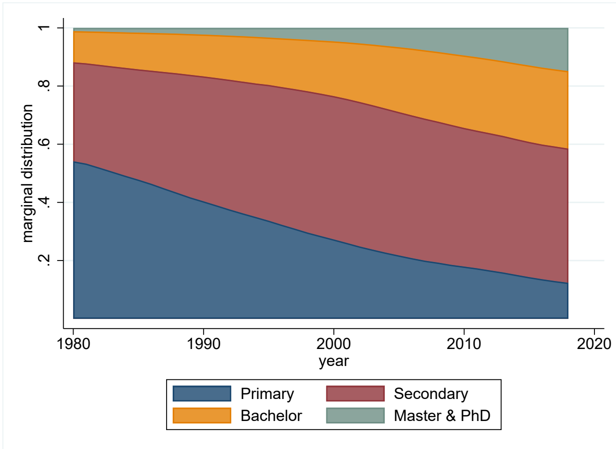
Notes: Likelihood indices for assortatively matched couples (identical type $j = j'$) cf. equation (1) for educational level and educational ambition, and educational field categorizations. Section 2 explains how the sample, educational outcomes, and the labor market outcome (residualized log hourly wages) underlying the educational ambition types are constructed.

Figure C.2: Marginal Type Distributions

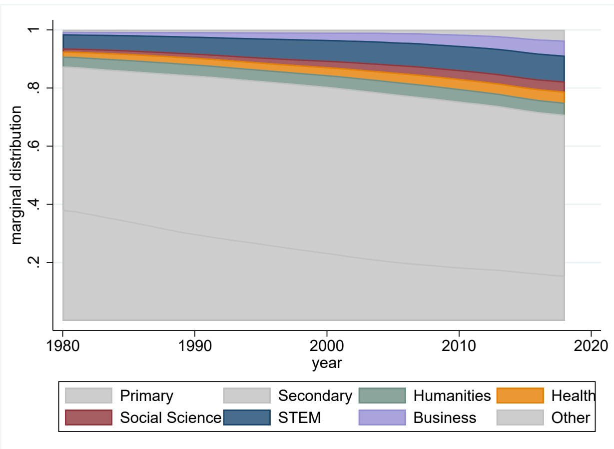
(a) Educational Level, Men



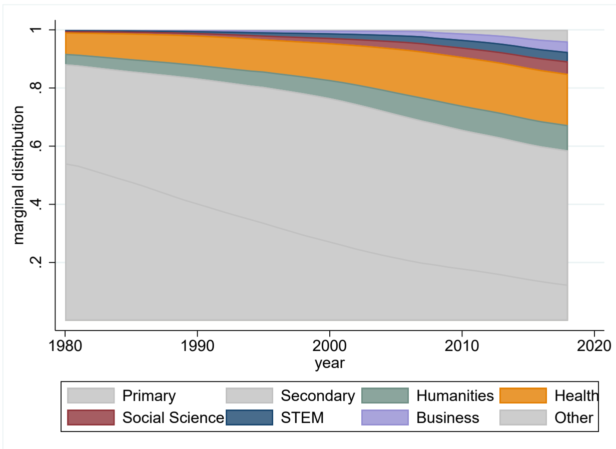
(b) Educational Level, Women



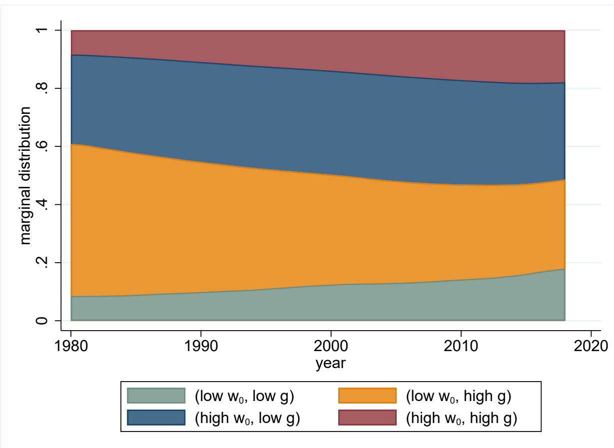
(c) Educational Fields, Men



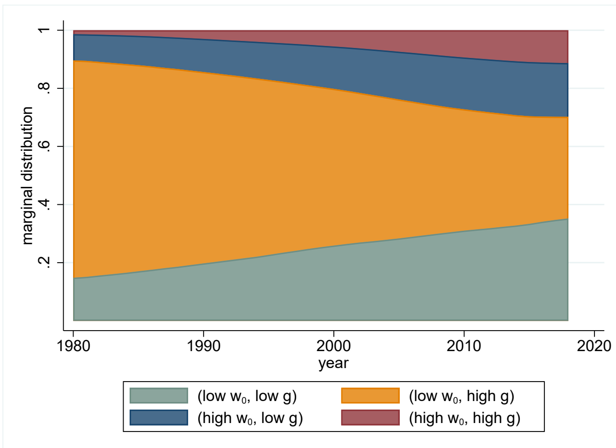
(d) Educational Fields, Women



(e) Educational Ambition, Men



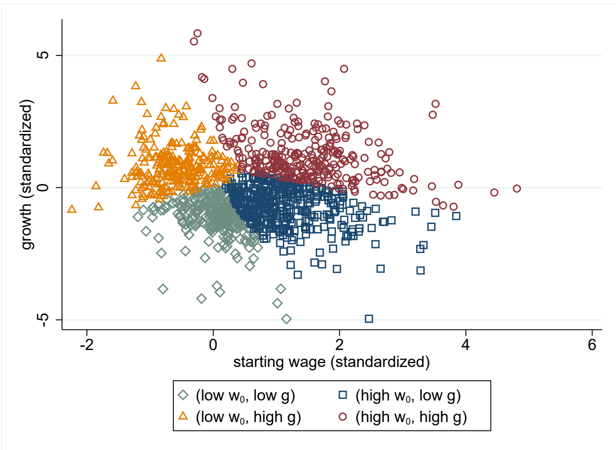
(f) Educational Ambition, Women



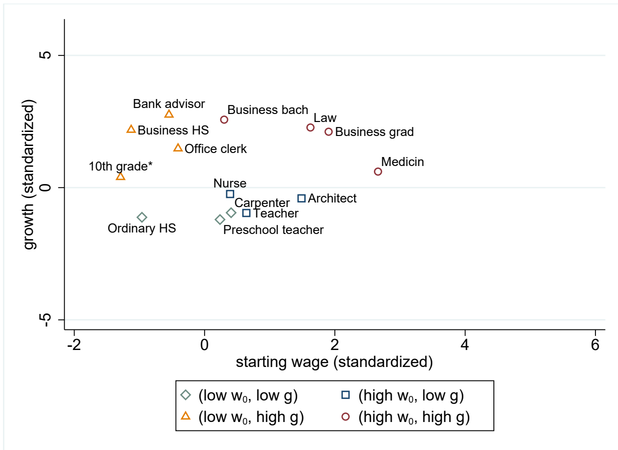
Notes: Marginal distributions for men and women over time by educational level and educational ambition. Sections 2 and 3 explain how the sample, educational levels, educational- fields and the labor market outcome (residualized log hourly wages) underlying the educational ambition types are constructed.

Figure C.3: Educational Program Categorizations by Gender

(a) Men, all programs

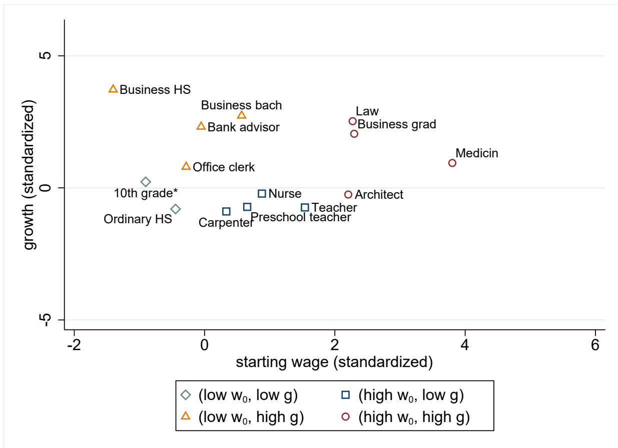


(b) Men, examples



(c) Women, all programs

(d) Women, examples



Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The horizontal axis corresponds to the standardized w_0 and the vertical to the standardized g. Points in the panels locate all educational programs with at least ten graduates in 2018—described in Section 2—along these two dimensions. Colors and markers uniquely assign each program to a marriage market type, depending on the panel's definition.

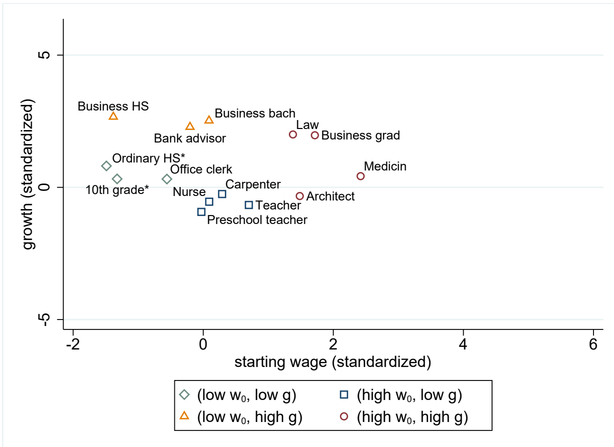
Figure C.4: Educational Program Categorizations by Decade

(a) 1989 and before, all programs

(b) 1989 and before, examples

(c) 1990–1999, all programs

(d) 1990–1999, examples



(e) 2000 and after, all programs

(f) 2000 and after, examples

Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The horizontal axis corresponds to the standardized w_0 and the vertical to the standardized g. Points in the panels locate all educational programs with at least ten graduates in 2018—described in Section 2—along these two dimensions. Colors and markers uniquely assign each program to a marriage market type, depending on the panel's definition.