Educational Ambition, Marital Sorting, and Inequality*

Frederik Almar† Benjamin Friedrich‡ Ana Reynoso§

Bastian Schulz¶ Rune Vejlin‖

March 27, 2025

Abstract

This paper revisits the link between education-based marriage market sorting and income inequality. Leveraging Danish administrative data, we develop a novel categorization of "ambition types" that is based on starting wages and wage growth trajectories associated with detailed educational programs. We find a substantial increase in assortative matching by educational ambition over time, and the marriage market explains more than 40% of increasing inequality since 1980. In contrast, sorting trends are flat with the commonly-used educational level categorization. We conclude that the mapping from education to types matters crucially for conclusions about how education-based marriage market sorting contributes to 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, University of Copenhagen, Barcelona GSE, the DGPE network, DAEiNA, IAB Nuremberg, University of Michigan, Universidad de San Andres, SEHO 2024, SOLE 2024, 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

§University of Michigan, Department of Economics, and NBER, 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 in recent 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; Chiappori, Salanié and Weiss, 2017; Hryshko, Juhn and McCue, 2017; Ciscato and Weber, 2020; Calvo, Lindenlaub and Reynoso, 2024). 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 *how* educational data 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 to go beyond the usual definitions of education based on levels (e.g., college) or fields of study (e.g., Social Sciences) and study assortative matching and household income inequality based on individuals' *programs* of education (e.g., Economics). We group the more than 1800 educational programs in our data into four types of programs that are similar in the labor market outcomes of their graduates. Our interpretation is that individuals assess the attractiveness of potential partners based on the typical career path of their program's graduates.

Based on these four 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 by commonly-used types constructed based on education *levels* (primary, secondary, bachelor's, and master's/PhD) and the post-secondary *field of study* (STEM, social sciences, business, humanities, health & welfare). We find that commonly used types conceal heterogeneity in both earnings potential and future work-life balance. Moreover, assortative matching by educational levels or fields exhibits no trend since at least 1995 and the changing educational composition of households based on these categorizations can only account for a relatively small fraction of inequality growth.

Our novel definition of types uses detailed data on the labor market outcomes of graduates. We measure these outcomes at the level of the educational program and assign types based on the most advanced educational program that individuals graduate from. This focus on programs builds on the literature that emphasizes heterogeneity in labor market outcomes across education programs (Altonji, Kahn and Speer, 2014, 2016). Programs are defined based on four-digit codes from the education register, which identify more than 1800 unique education programs in Denmark. Examples include the vocational training of *carpenters*, professional degrees held by *nurses* and *pre-school teachers*, and 5-year university degrees in *law* and *business*.

We then group education programs based on their similarity in graduates' average wage dynamics. To do so, we merge the Danish education registers with the labor market histories of all program graduates and compute average *starting wages* and *wage growth* for each program, based on *hourly wages*. We choose these two variables to summarize the relevant features of education programs for marital sorting for three reasons. First, these wage-based measures together capture the potential labor market returns to degree programs over the life cycle. Second, in capturing programs returns based on wages—rather than alternative moments of earnings—we reduce the possibility that education types reflect the role of hours worked, which is an endogenous family choice. Third, the wage dynamics captured by our measures, in turn, influence the opportunity cost of time spent in the household producing the family's public goods—that is, the extent to which individuals balance their careers and family life. Taken together, we view our approach as a parsimonious, yet conceptually crucial step to capture the labor market and marital values of educational programs beyond what is captured by summary measures such as average lifetime earnings. In particular, programs associated with similar levels of lifetime earnings but with different combinations of starting wages and wage growth may have opposite implications for work-life balance. For example, to reach the same expected lifetime income, programs with lower initial wage and faster wage growth may be expected to require stronger career investments over the life cycle and crowd out family-time relative to programs with higher starting wages and slower growth. Hence, these degrees may carry different signals in the marriage market, which we are able to confirm in the data.

Methodologically, we define four types by grouping programs based on similarity in these 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 education-based types to study marriage market matching. An advantage of using k-means is that it allows us to collapse multiple labor market characteristics into a single, one-dimensional trait. This greatly simplifies the analysis of sorting.[1]

The first part of our analysis robustly shows that our novel ambition types convey important

[1] Multidimensional sorting in labor and marriage markets has been explored, for example, by Lindenlaub (2017), Ciscato and Weber (2020), Low (2024), Foerster, Obermeier and Schulz (2024), and Chiappori, Ciscato and Guerriero (2024).

information about both earnings potential (by construction) and work-life balance (by result), which 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 ambition types are clearly distinct in the information they carry about career flexibility and expected time commitments to the family. Using regression analysis, we investigate the link between the program-level labor market outcomes that we use to construct the ambition types and seven proxies for the graduates' work-life balance. These proxies include a measure of the part-time penalty 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 and the field of study. Importantly, the link between, on the one hand, starting wages and wage growth and, on the other hand, time input choices that matter for the family holds even conditional on lifetime earnings. This suggests that different wage dynamics leading to the same lifetime income are associated with different levels of work-life balance.

In the second part of the analysis, we compare trends in the education composition of households and their contribution to the rise in between-household inequality across the different definitions of education categorizations—ambition types, educational levels, and educational fields. To this end, for all categorizations, we consider the role of *trends* along three margins that influence *marriage market sorting patterns*: the distribution of education types by gender (or marginal distributions), the share of singles by gender and education (the marital *extensive margin*), and the frequency of couples by spouses' joint education (who marries whom).

Our first main finding is that the degree of assortative matching 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 assortative matching trends crucially depend on how the categorization of underlying educational programs into types is implemented. We follow Eika et al. (2019), Almar and Schulz (2024), and Chiappori,

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

Costa Dias, Meghir and Zhang (2025) to take into account how the marginal type distributions change over time. We define our aggregate sorting measure as the weighted sum of the matching frequencies of same-type couples relative to the same frequencies under random matching, conditional on marriage (likelihood ratios). Changing weights reflect that the type distributions of women and men change over time. Consequently, trends in assortative matching based on the different categorizations can be compared.

Our second main finding is that changes in the three marriage market margins based on educational ambition explain more than 40% of the overall increase in income inequality (as measured by the Gini coefficient) between 1980 and 2018 in Denmark. In turn, changing sorting patterns based on educational level and field of study contribute much less to explain inequality growth. Methodologically, we compare the observed between-household inequality measure every year to a counterfactual version that we calculate by re-weighting households such that their composition in terms of the three marriage market margins and their degree of assortative matching stay fixed at the 1980 levels. For this counterfactual, we follow DiNardo, Fortin and Lemieux (1996) and Fortin, Lemieux and Firpo (2011). We further assess the separate roles of trends in the marginal distributions of education versus trends in marital choices to explain inequality. To this end, we follow Chiappori et al. (2020a) and combine our counterfactual re-weighting method with a structural marriage market matching model of the Separable Extreme Value (SEV) type (Choo and Siow (2006), Chiappori et al. (2017), Chiappori, Costa Dias and Meghir (2020b), Chiappori, Fiorio, Galichon and Verzillo (2022), Chiappori et al. (2025)), which relates changes in assortative matching to changes in the value of marriage relative to remaining single (marital surplus). We find that the evolution of marital surplus contributes to the rise in inequality regardless of how we construct types. In contrast, changes in the marginal distributions of education types mitigate inequality when the data are seen through the lens of education levels or fields but not based on educational ambition.

Our findings are robust to alternative ways of constructing ambition types as long as they summarize the earnings potential and work-life balance implications of educational attainment. In particular, we show they are robust to constructing gender- and cohort-specific ambition types, different numbers of types, and different levels of aggregation.

By proposing a new definition of education types, we extend the literature on 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 marriage market matching and inequality considering the broad population including couples of mixed levels of education. Furthermore, we build on the literature that underscores the value of (permanent) income for marital sorting (e.g., Chiappori et al. (2022)) and contribute a measure that differentiates various paths (i.e., combinations of starting wages and wage-growth) to achieve a given level of income.

Moreover, we contribute to the debate on the relationship between trends in marriage market matching and inequality by showing that the choice of how to map data on educational attainment into a small number of types affects the conclusions. All three definitions of education types that we consider—ambition, educational levels, and educational fields—use information from the most advanced program that an individual graduates from. Still, conclusions about how marriage market matching has changed and how it influenced inequality differ significantly. First, the previous literature considered 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) and on the field of study (Seiver and Sullivan, 2020; Artmann et al., 2021; Han and Qian, 2022), yielding mixed results. Instead, we consider matching based on our novel ambition types and reveal an upward trend in assortative matching, along with a significant contribution to the increment in inequality. Second, we build on Burtless (1999) and Chiappori et al. (2020a) and consider trends in single-headed households for our analysis of the impact of changes in family composition on income inequality. We extend their work by showing that trends in the share of single individuals by education and gender differ depending on how we define education-based types, and hence contribute to explaining differences in conclusions across categorizations.

Finally, our insight that the categorization of types matters to capture the relationship between marriage market sorting and inequality can potentially be applied in the recent literature that studies the link between assortative matching and inter-generational 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.

The paper is organized as follows. Section 2 introduces our data. In Section 3, we construct 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 marital sorting on ambition has increased substantially since the 1980s while sorting on educational level or field has not. Section 5 presents our analysis of the drivers of inequality trends 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 are a yearly panel and provide information for the entire population of Denmark 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 these data with the Danish Labor Force Survey (LFS) (LFS, Statistics Denmark, 2000–2018), which contains more detailed labor market information on, e.g., hours of work for a subsample of the population. 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 person identifiers. We next describe the key variables we construct for our analysis and provide further details in Online Appendix A.

To construct the link between education and labor market outcomes, we include all individuals living in Denmark between 1980–2018 in the age range of 19-60, irrespective of marital status. On average, we observe 3,031,511 individuals per year. In this sample, we measure individual education as the most advanced *educational program* from which an individual graduates.[3] Educational programs are defined based on four-digit (Danish) ISCED codes from the education register, which uniquely identify 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*.[4] 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.[5] 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. The income registers also give us access to individual and parental wealth, which we use to describe our novel ambition types in the next section.

The representative Danish LFS allows us to investigate how career choices manifest themselves in labor market outcomes beyond wages and participation. The LFS provides unique insights into working conditions because it contains the exact number of hours worked, and information on whether the individual works in the evenings, on weekends, and at home. This

[3] We exclude 2.6% of individuals with missing education information, and we do not consider programs that an individual did not complete. For example, the most advanced program of high school dropouts is "Primary Education."

[4] Some programs are tied to one particular university and thus location, but most programs are spread around the country. We distinguish university degrees by institutions in an extension; see Section 6.

[5] Hourly wages are calculated based on 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.

helps us to assess the work-life balance across educational programs and ambition types. The LFS is available from 2000 onward and we merge it with our main data at the individual level.[6]

Our analysis of household-level income inequality includes both singles and couples aged 19 to 60. To compute couples' household income, we consider individuals who are either married to or cohabiting with another individual of the opposite sex.[7] On average, we observe 1,800,866 individuals in couples per year. There is an upward (downward) trend in cohabitation (legal marriage), but the combined stock of couples is stable over time.[8] Our measure of household income is yearly labor income from both regular employment and self-employment.[9] For couples, we use the sum of both spouses' labor income.

3 Ambition Types

Previous research has established that education is a valuable trait in the marriage market. It predicts both earnings potential and future time investments into career and family (Chiappori et al., 2018; Gayle and Shephard, 2019; Calvo, 2023; Calvo et al., 2024; Reynoso, 2024). 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 depend on the education of the spouses. It is therefore desirable to analyze marital sorting based on educational types that reflect expectations about earnings potential and work-life balance. While educational level or field of study may partially capture these characteristics, we demonstrate below that these established types mask substantial heterogeneity that is crucial to understanding trends in assortative matching and inequality.

3.1 Conceptual Framework

To construct education-based types, suppose that men and women are distinguished by their education program, $i \in \mathcal{P} = \{Program_1, Program_2, ..., Program_I\}$, as defined in Section 2. I is 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

[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.

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

[8]Figure A.1 in Appendix A.4 depicts the evolution of the stocks and the age composition of couples.

[9]It is based on the income register (IND, Statistics Denmark, 1980–2018).

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, programs of education are commonly grouped into a small number of 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 $T << I$ *types* by grouping the I programs of education based on their similarity in the sub-vector of observable characteristics $\tilde{x}$.

Many articles in the literature use the mapping $\mathcal{T}_{Levels}$, which maps programs based on similarity in one characteristic, namely, their level of education (see, e.g., Fernández and Rogerson, 2001; Greenwood et al., 2014, 2016; Eika et al., 2019). One way to implement this mapping is to choose four types, $t_{Levels} = \{Primary, Secondary, Bachelor, Master\&PhD\}$. Another popular one-dimensional mapping is $\mathcal{T}_{Fields}$, which focuses on post-secondary education (see, e.g., Seiver and Sullivan, 2020; Artmann et al., 2021; Kirkeboen et al., 2022; Han and Qian, 2022). It groups educational programs by post-secondary study field, mapping the narrow fields of individual programs, $\tilde{x}_i = field_i$, into broader groups of fields (e.g., STEM, health, humanities, social sciences, business), $t_{Fields} = \{Field_1, Field_2, ..., Field_T\}$.

3.2 Construction of Ambition Types

We take advantage of our rich data and construct the ambition types using average labor market outcomes at the program level. 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.[10] 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.[11] 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 force.[12] We average over years for both w_0 and w_1 to smooth out year-to-year variation that is unrelated to worker productivity.

[10] 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.

[11] 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.

[12] 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.

In our benchmark analysis, we construct four types using the mapping $\mathcal{T}(w_0, g)$. To implement the mapping, we cluster programs based on standardized starting wages and wage 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 wage growth. We denote this mapping $\mathcal{T}_{Ambition}$, which includes the following four types:

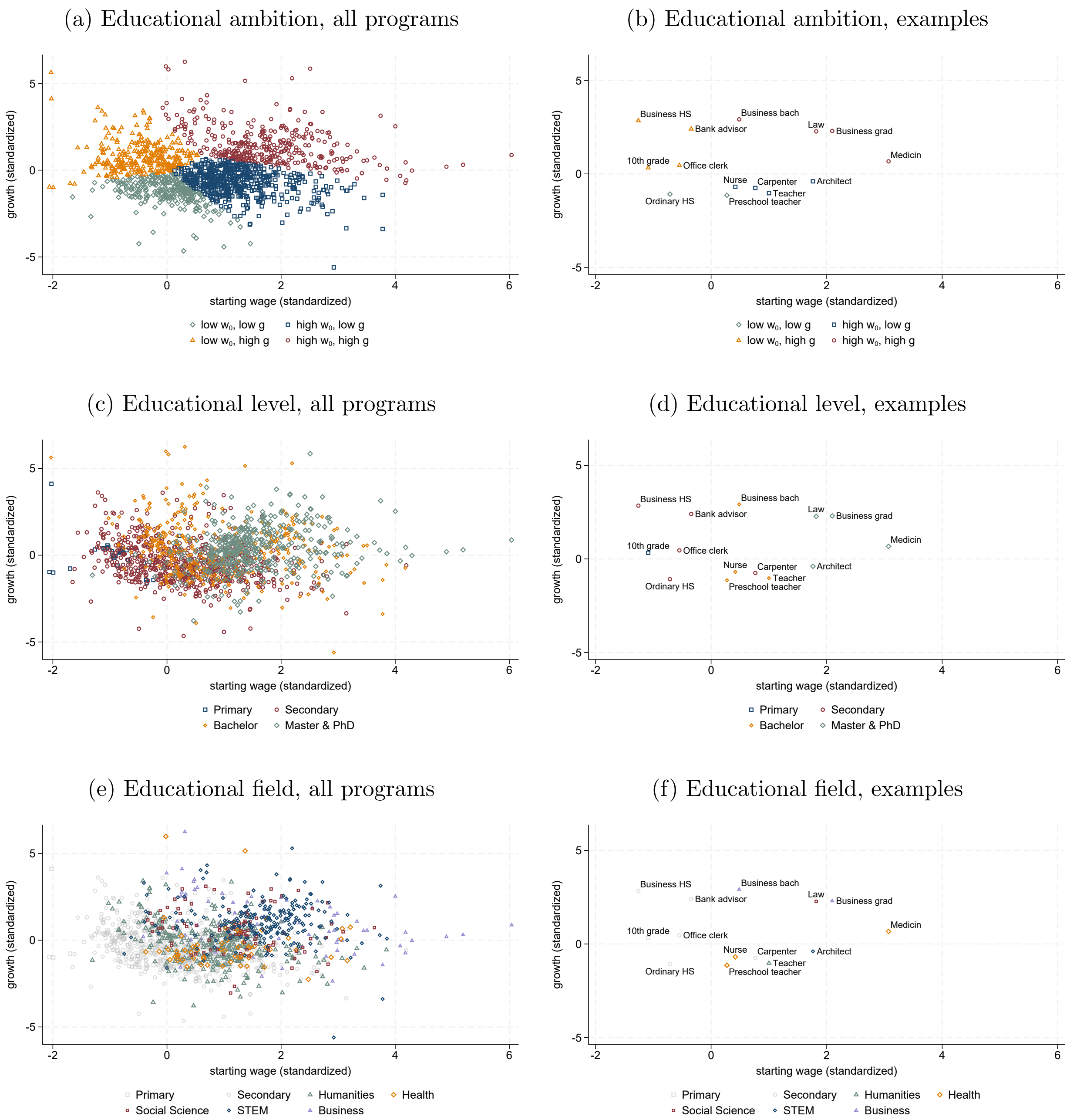
$$t_{Ambition} = \{(low\ w_0, low\ g), (high\ w_0, low\ g), (low\ w_0, high\ g), (high\ w_0, high\ g)\}.$$

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 approximately 1800 programs in the space of standardized starting wages (horizontal axis) and standardized wage growth (vertical axis). We distinguish the resulting four types with different colors and markers. The mapping delivers 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 by their choice of educational program. 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.

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 also predominantly male. Moreover, individuals who graduate from programs with high starting wages tend to have wealthy parents.

For comparison with previous studies, we show the same figure for 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, but Panels (c) and (e) assign those programs to different color-marker groups depending on the level or post-secondary field of eduction, 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 five post-secondary

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



(a) Educational ambition, all programs

(b) Educational ambition, examples

(c) Educational level, all programs

(d) Educational level, examples

(e) Educational field, all programs

(f) Educational field, 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 an education type, depending on the panel's definition.

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, and Business.

The relationship between the types and the two labor market outcomes becomes blurry when the types are constructed based on level or field of education. For example in Panel (c) for level, we see that 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 & 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) using Fields, 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. Thus, the commonly used rankings do not reflect key labor market characteristics very well.

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.[13] 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.

Online Appendix Table A.2 complements these examples with a cross-tabulation of ambition types and levels or fields. 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 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 Marital Value of Ambition Types

Next, we show that our definition of ambition 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 possesses. To show this, we build on the literature and construct seven proxies for the work-life balance of graduates of specific educational programs. The proxies emphasize the trade-off between career investments and time commitments to the family (Wiswall and Zafar, 2021; Goldin, 2014; Calvo et al., 2024). Each row in Panel A of Table 1 corresponds to one of the proxies while the columns label the four ambition types. The table shows the mean (and standard deviation in parentheses) of the proxy conditional on each ambition type—in

[13] These are examples of educational programs that have a large number of graduates within each cluster. We count graduates for the year 2018 in the sample defined in Section 2.

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		
Part-time penalty	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).

the first four columns—and for the pooled sample—in the last column.

Part-time penalty is constructed as the ratio of the w_1 of full-time workers to that of part-time workers.[14] It reflects the additional return to working long hours, a measure of inflexiblilty inspired by Claudia Goldin's analysis of within-occupation gender differences 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 least part-time). *Full-time* captures the fraction working at least 32 hours per week.[15] *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).[16] Finally, *Lifetime earnings* is the sum of deflated annual earnings over 30

[14]Recall that w_1 was the average wage in years 9-11 in the labor force used to construct the ambition types.

[15]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).

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

years after graduation.[17]

Overall, Table 1 documents two key patterns. First, graduates in the most ambitious category $(high, high)$ are more career-focused than graduates in the other categories with lower w_0 or/and lower g. For example, graduates with high wage growth are penalized in terms of hourly wages for part-time work (inflexibility). Types with a high wage level are more likely to participate in the labor market and to work full time, relative to graduates with a low wage level (w_0). Moreover, graduates from the least ambitious programs have their first child on average at the youngest age and two years before the most ambitious graduates. The strong career-focus of graduates from the most ambitious programs is associated with substantially higher life-time earnings and levels of wealth compared to the other three types. 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 standard work week in Denmark), 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 almost always 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 almost all (25 out of 28) pair-wise comparisons of ambition types.

To analyze these patterns further, we use regression analysis at the program level 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 life-time earnings. The table shows the coefficients from regressions of the first six proxies (we control for life-time earnings in the fourth column of each Panel) 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* control for lifetime earnings as defined in Table 1.

[17] 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

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) Part-time penalty				(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.

For all proxies, we find that higher starting wages or higher wage growth (and in most cases both) 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 2.1% *relative* to the mean across programs. Within the same level or field of education, the effects become somewhat smaller (1.9% and 1.5%, respectively) but remain highly significant.

Similarly, we find that higher starting wages and 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. Higher wage growth is associated with

compositional differences by programs in the share of graduates at different life-cycle stages.

higher degrees of inflexibility (part-time penalty) across all specifications, while starting wages are less so. This is expected as wage growth is most likely associated with higher degrees of investment in human capital. Finally, age at first child is strongly associated with starting wages and wage growth in all specifications except when controlling for life-time earnings where the association with wage growth disappears. All 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 lifetime 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, which is captured by different ambition types. Capturing this heterogeneity in the family value of different career types is the main advantage of our new categorization.

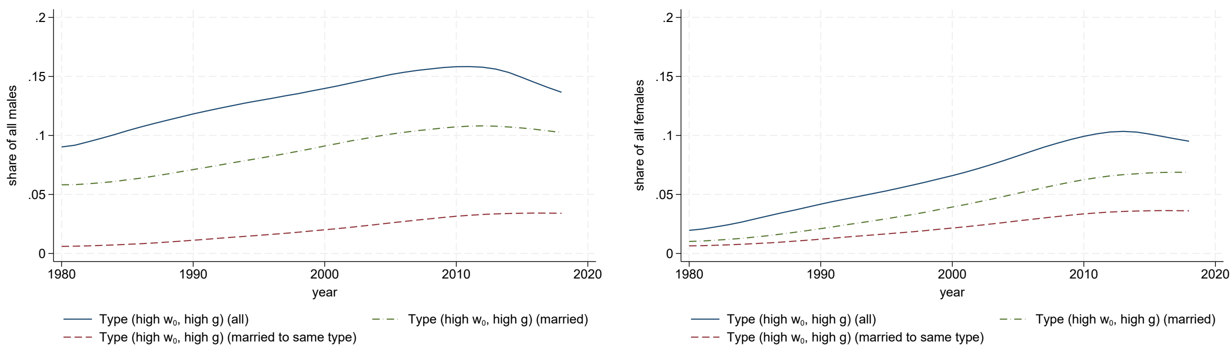
4 Trends in Assortative Matching

In this section, we show that the degree of assortative matching on educational ambition has increased over time, while assortative matching on educational levels or fields has remained constant since the early 1990s.

From an empirical perspective, the degree of positive assortative matching (PAM) manifests itself as a positive association of spousal types in the cross-section, conditional on marriage. This association can be measured based on correlation coefficients, distance measures, or the frequency distribution of spousal types among couples (the contingency table). However, determining whether PAM in the marriage market has *changed over time* has proven elusive in the literature because the educational configuration of households change due to three margins: (i) changes over time in the educational attainment of men and women (i.e., the gender-specific *marginal distributions* of education); (ii) changes over time in selection between singlehood and marriage by gender and education (i.e., the extensive margin decision of *who marries*); and (iii)

Figure 2: Population Shares and Marriage Rates of Highly Ambitious Individuals

(a) Male (b) Female



Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. For each gender, the solid blue line plots the share who graduated from educational programs in the $(\omega_0, g) = (high, high)$ type as defined Section 3.2; the dash-dotted green line plots the share of $(high, high)$-type individuals who marry; and the dashed red line plots the share of $(high, high)$ individuals who marry someone of the same type.

changes over time in the joint distribution of husbands' and wives' education (i.e., the intensive margin of *who marries whom*).

Specifically, we illustrate the relevance of these three adjustment margins in our context in Figure 2 by zooming in on one definition of education (ambition) and one type within that definition that will turn out to play an important role for aggregate sorting trends below: *highly ambitious*—(high w_0, high g)—men (Panel a) and women (Panel b).

Solid blue lines show that the share of high-type individuals[18] increased for both men and women between 1980 and 2018, and, notably, relatively more so for women. For men, the share increased from roughly 9% to around 15%. For women, it quintupled from around 2% to 10%.

In addition to the changes in the population shares, the dash-dotted green lines show that the marriage rates of high-type men and women have not evolved in parallel. Although the share of married high-type men roughly increased in tandem with the share of high-type men, a gap opened up for high-type women.[19] Thus, *who marries*—the extensive margin—has changed differentially for high-type men and women.

Finally, the dashed red lines show the shares of "sorted" high-type men and women, i.e., men and women who have a partner of the same type. For both genders, the shares in "power couples" increased, albeit less so than their population shares and the shares of married individuals, particularly for women. In addition, the evolution of the ratio between the shares of

[18]Expressed as shares of the male and female populations, respectively, so that 100% corresponds to all men (women) across ambition types.

[19]This gap reflects that the relative share of married high-type women has increased between 1980 and 2018. That is, in 1980 the implied single share was around 50% and it decreased to around 30%, see also Appendix Figure B.3.

sorted individuals and the population shares differs across genders. It is relatively stable for men (perhaps even slightly increasing) but decreasing for women, implying that the tendency to sort, i.e., marry an individual of the same type, has evolved differentially across genders.

The increasing number of highly-ambitious individuals who decide to marry, particularly women, mechanically increases the probability that two high-type spouses meet and form a couple. More generally, in Appendix B.1, we find differential trends in these margins by gender and type for all educational categorizations.[20] Therefore, a more refined measurement of assortative matching is necessary. We employ a sorting measure that is based on both the shares of same-type couples and the marginal type distributions conditional on marriage: the likelihood ratio (Eika et al., 2019; Chiappori et al., 2025). This measure compares 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, which is captured by the product of the type-specific shares of men and women in the married population.

Assume that every couple consists of two individuals that are characterized by $t_m = j$ and $t_f = j'$, where m and f indicate gender, so j (j') represents male (female) types. Let $P^M(t_m, t_f)$ denote the fraction of married couples of type (t_m, t_f) and let the fraction of male type t_m who marry be denoted by $P^M(t_m)$ (analogously, we use $P^M(t_f)$ for the fraction of married women of type t_f). For each combination of types, the likelihood ratio is defined as follows:

$$s\left(j, j'\right) = \frac{P^M\left(t_m = j, t_f = j'\right)}{P^M\left(t_m = j\right) P^M\left(t_f = j'\right)}, \tag{1}$$

where the one-dimensional types $t \in \{1, \ldots, j, \ldots, T\}$ are the categories defined in Section 3.1: educational level, field, or ambition. The measure relates the observed frequency of a couple type (the numerator) to the expected frequency under random matching (the denominator). The latter is expressed as the product of the population shares of men and women conditional on marriage in the respective categories, which captures how the marginal distributions conditional on marriage affect the likelihood of matching with a specific type. For couples with two spouses of the same type ($j = j'$), a likelihood ratio greater than one implies that the likelihood of matching in this configuration is higher than what is predicted by random matching, indicating PAM in the cross-section.

To compute an aggregate sorting measure, we sum across the type-specific likelihood ratios of same-type couples ($j = j'$) using the weights $\{\pi_1, \ldots, \pi_j, \ldots, \pi_T\}$, where $\sum_{j=1}^{T} \pi_j = 1$ and T

[20]Specifically, we show heterogeneous trends in the marginal distributions both overall and conditional on marriage (Appendix B.1.1) and single shares (Appendix B.1.2) by gender and type across definitions of education.

is the total number of categories. This yields the measure $\mathcal{S}$, which is a weighted sum of the underlying likelihood ratios:

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

The weights ensure that within-category sorting patterns are properly reflected in the aggregate measure. Chiappori et al. (2020b) suggest to compute the weights as a convex combination[21] of the male and female population shares (conditional on marriage) for type j:

$$\pi_j = \lambda P^M(t_m = j) + (1-\lambda) P^M(t_f = j). \tag{3}$$

Computing the weights based on the marginal type distributions makes the aggregate measure comparable over time. Almar and Schulz (2024) show that there exists a value of the parameter λ that minimizes the distortionary influence of changing marginal distributions on the sorting measure. Specifically, the mechanical effect of changing type distributions can be minimized by placing the weight on the short side of the marriage market.[22] We apply their method to compute λ and use the modal value to ensure comparability across years.[23]

Figure 3 shows the aggregate sorting trends for educational levels (red dashed), fields (orange long dashed), and ambition (green dash-dotted). To facilitate the comparison and because the levels of the three sorting measures are not directly comparable, we index all trends such that $1980 = 100$.[24] Based on educational levels, we find that sorting has increased by around 3% since 1980, consistent with the flat trend for the US documented by Eika et al. (2019). Based on educational-field types, we find that sorting has increased a bit more (around 7% since 1980), but the trend is flat from the mid-1990s onward.

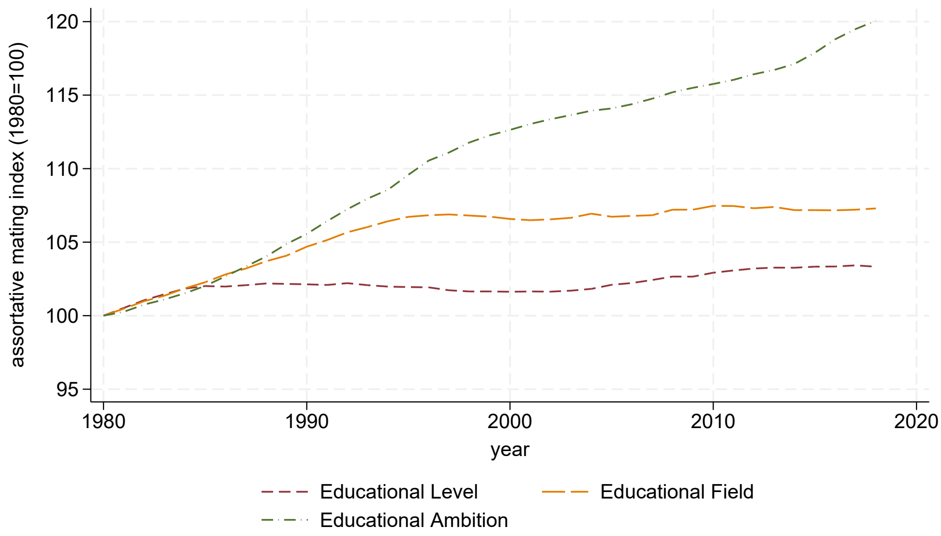
[21]The weighted sum of likelihood ratios using this weighting scheme fulfills the formal criteria for sorting measures discussed in Chiappori et al. (2020b).

[22]Details are provided in Appendix A.3. Intuitively, the short side of the market matters more for matching. For example, if high-type men and women have a preference for matching with one another but there are more high-type men than women, all women are going to find a match while some men are going to be left over. In this example, changes of the female type distribution have a greater mechanical effect on measured sorting. Putting more weight on the female type distribution in (3) minimizes this effect.

[23]We use the mode to avoid jumps in our sorting trend when the short side of the market flips, e.g., when the population share of high-type women becomes greater than the population share of high-type men.

[24]The reason why the values of our sorting measure $\mathcal{S}$ are not comparable in levels is that differences in marginal distributions affect the value of $\mathcal{S}$, even when sorting patterns are the same. For example, consider two marriage markets with different distributions of education but all couples being same-education couples. These two markets exhibit the same sorting patterns but a different value of the weighted sum of likelihood ratios, $\mathcal{S}$. Then, differences in the values are not informative of the differences in the degree of assortative matching. Note that this holds both within and across categorizations. Within market, trends in $\mathcal{S}$ can be compared due to the weighting strategy as we explain in the text. Therefore, in Appendix Figure B.4 we show trends in the non-normalized measure $\mathcal{S}$, and we find evidence for PAM using all three categorizations (the aggregate sorting measure $\mathcal{S}$ is greater than one for all three categorizations in every year).

Figure 3: Increasing Assortative Matching based on Ambition Types



Notes: The figure shows the sorting measure $\mathcal{S}$ derived in 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). Normalized such that $\mathcal{S}$ in 1980 is 100. Types are constructed as explained in Section 3.

Interestingly, we find a strong increase in assortative matching based on educational ambition. Relative to random matching, the likelihood of observing couples with the same ambition type has increased by approximately 20%. The difference in the PAM trends based on educational levels/fields on the one hand and educational ambition on the other is explained by differences in the evolution of the underlying (weighted) likelihood ratios (1) (which we plot in Figure B.5), which in turn are determined by marginal type distributions and the composition of who marries (discussed in detail in Appendices B.1.1 and B.1.2).[25]

For educational levels, the "secondary" category shows a likelihood ratio that is just slightly above one and flat, indicating that PAM among individuals with secondary education is neither pronounced nor increasing over time, see Figure B.5, Panel (a). Panel (b) of the same figure shows trends in the weighted likelihood ratios, which reveals that the flat trend of the "secondary" group dominates the aggregate trend for educational levels. The tendency to sort decreases in the growing tertiary categories and increases in the shrinking primary education category, but these trends largely cancel each other out, so there is almost no change in the aggregate PAM measure.

For educational fields, recall that the primary and secondary categories are shared with the educational-level categorization, so any differences between the educational-levels and the

[25]Panels (b), (d), and (f) of Appendix Figure B.5 show trends in the underlying weighted likelihood ratios of same-type couples for all types and categorizations. These trends add up to the aggregate sorting trends in Figure B.4.

educational-fields sorting trends in Figure 3 are driven by different sorting trends across fields within tertiary education. We show the likelihood ratios for five groups of fields in Figure B.5, Panel (e). The type-specific likelihood ratios are positive and decreasing in all fields, similar to the trends for "Master & PhD" (educational levels).For men, the composition of graduates across fields is relatively stable, see Figure B.1, Panel (e). For women, the fields "Business", "STEM", and "Social Sciences" grew in size over time, see Figure B.1, Panel (f), and the weighted likelihood ratios reveal that these three fields contribute to increasing sorting by educational field, see Figure B.5, Panel (f). However, "Health" and "Humanities" dominate in terms of size and the weighted sorting measures for these fields are decreasing after the year 2000, flattening the aggregate trend.

For educational ambition, the (weighted) sorting trends are depicted in Figure B.5, Panels (c) and (d). The unweighted measures suggest that PAM is pronounced in the category with high starting wages and high wage growth. The unweighted likelihood ratio is decreasing over time, but due to the increasing size of this group, especially for women (Figure B.1, Panels (c) and (d)), the likelihood of a match among highly ambitious individuals becomes more and more important for the aggregate sorting trend over time. The likelihood of a match is lower pronounced but stable in the category with low starting wages and low wage growth. This group also increases in size and therefore contributes to the positive aggregate trend. The other two educational ambition categories exhibit unweighted likelihood ratios close to one. The category with low starting wages but high wage growth shrinks over time, curtailing its influence on the aggregate measure. The category with high starting wages and low wage growth plays the smallest role for the aggregate trend. Taken together, these developments explain the significant 20% increase in aggregate sorting based on educational ambition.[26]

5 Marriage Market Sorting and Inequality

In this section, we show that the evolution of the composition of households based on education affects inequality, but conclusions depend on the categorization of education. To study the link

[26] As a robustness check, we also consider (log) odds ratios to study the evolution of sorting patterns. The odds ratio has a number of desirable properties (Chiappori et al., 2025) but it is local in nature, so interpretation is more involved relative to the aggregate measure $\mathcal{S}$. We calculate it for all the (2×2) submatrices with same-type couples on the diagonal that can be extracted from our contingency tables, as in Table 6 of Chiappori et al. (2020a). For levels of education, all odds ratios decreased between 1980 and 2018. For educational ambition, we see an even split of increasing and decreasing odds ratios. Consistent with the documented increase in the weighted sum of likelihood ratios for educational ambition, we find that the submatrices with increasing odds ratios make up a larger share of the 2018 marriage market than the submatrices with decreasing odds ratios. This supports our conclusion that sorting trends across categorizations differ, pointing toward increasing sorting for educational ambition but not for levels of education.

between the evolution of marital sorting patterns and household-level income inequality for different education-based categorizations of types (levels, fields, and ambition), we first use the semi-parametric decomposition technique of DiNardo, Fortin and Lemieux (1996) (DFL) to quantify the role of the marriage market for inequality. Then, we combine this method with the marriage market matching model of Choo and Siow (2006), which allows us to assess the contribution of the different margins of marriage market matching (education distributions and marital preferences) to raising income inequality between households.

An alternative way of keeping marital sorting patterns fixed is by taking an algorithmic approach as suggested by Mosteller (1968), and, more recently, Greenwood et al. (2014) and Eika et al. (2019). Our decomposition builds on Chiappori et al. (2020a), who propose to take a structural approach to the link between marital sorting and inequality, which allows to incorporate trends in the composition of both married and single households into the analysis.[27]

In Subsection 5.1, we build on the DFL approach to quantify the importance of (i) changing marital sorting patterns and (ii) changes in the returns to educational types for inequality trends. In a nutshell, we compute counterfactual household income inequality measures at any year τ by keeping fixed at a baseline year one of two dimensions: (i) the composition of households (the marriage market dimension); and (ii) the returns to educational types. Note that scenario (i) keeps the whole marriage market fixed (that is, it fixes all the marginal type distributions and the matching patterns along both the extensive and the intensive margin). Our focus in this section is to study how the changing composition of households in terms of education contributes to changes in income inequality based on different categorizations of education. Moreover, by keeping all margins fixed we also guarantee that at any year τ the aggregate sorting measure $\mathcal{S}$ is as in the baseline year, which allows us to link trends in this measure of sorting to trends in inequality. To assess the role of each of the margins that we fix in scenario (i), we require additional structure on how individuals choose partners because we cannot fix one margin and determine the other two.[28]

In Subsection 5.2, we use additional structure to disentangle the effects of changes in the composition of households along the different margins—marginal distributions and marital surplus—for increasing inequality. To this end, we use a structural marriage market matching

[27]Our qualitative results are robust to using the algorithmic approach, which we used in an earlier version of this study (Almar et al., 2023) that did not consider single households.

[28]This is due to feasibility conditions in the marriage market, which only restrict the allocations to be such that the total mass of individuals of any education type equals the sum of married and single individuals of that type. For example, if we fix the marginal distributions of education at the 1980 level, a decision model of the marriage market is needed to assign individuals into counterfactual fractions of couples and singles by educational type.

model of the Separable Extreme Value (SEV) type (Choo and Siow, 2006), which allows us to differentiate between (iii) changing gains from marriage (marital surplus) affecting both the intensive and extensive margin and (iv) changing marginal type distributions.

For each scenario (i) to (iv), we study how changes in the dimension(s) held fixed contribute to rising household-level inequality between 1980 and 2018. To examine how the definition of education affects conclusions about the contribution of changing households' education compositions to inequality growth, we compare results for all three categorizations of types (levels, fields, and ambition) in each counterfactual scenario. 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.

5.1 Fixed Sorting Patterns vs. Fixed Returns

Let $\mu_{\tilde{x}}$ represent the marital frequency table (that is, the frequency matrix of all types of households, married and single) according to the definition of education $\tilde{x}$. Note that $\mu_{\tilde{x}}$ is determined by the marginal distributions of types $\tilde{x}$, who marries (the extensive margin), and who marries whom conditional on marriage (the intensive margin). To reduce notation, in what follows we omit the subscript $\tilde{x}$ when clear from context. Let $\tau \in [1980, 2018]$ denote year, and τ_y and τ_μ denote the year in which household income y and the marital frequency table μ are measured, respectively. Let the distribution of household income conditional on household education types in year τ_y be denoted by $F_{Y|\mu}(y|(t_m, t_f), \tau_y)$. We first consider two counterfactual scenarios, as described below.

(i) Fixed sorting patterns. We analyze how income inequality would have developed had the composition of households remained unchanged. Specifically, we keep μ fixed at base year τ_μ, while letting the household income distribution F_Y change to some year τ_y. The counterfactual income distribution when income is measured in year τ_y and the marriage market is fixed at its composition in year τ_μ is:

$$\widehat{F_Y}(y|\tau_y, \tau_\mu) = \int F_{Y|\mu}(y|(t_m, t_f), \tau_y)\psi_{\tau_y,\tau_\mu} d\mu((t_m, t_f)|\tau_y). \tag{4}$$

Importantly, the factor ψ_{τ_y,τ_μ} allows us to re-weight households in the τ_y data to reflect their frequency had the marriage market stayed at its τ_μ composition. Specifically,

$$\psi_{\tau_y,\tau_\mu} = \frac{d\tilde{\mu}((t_m, t_f)|\tau_\mu)}{d\mu((t_m, t_f)|\tau_y)}.$$

The denominator of the re-weighting factor reflects the composition of households in the observed year τ_y. In practice, we estimate this denominator as the probability of observing household type (t_m, t_f) in year τ_y, which we denote by $P_\mu((t_m, t_f)|\tau_y)$. The numerator, in turn, reflects the *counterfactual* composition of households had the marriage market stayed at its τ_μ configuration. In practice, we follow DiNardo et al. (1996) and Fortin et al. (2011) and estimate it with the probability of observing household type (t_m, t_f) in year τ_μ, which we denote by $P_\mu((t_m, t_f)|\tau_\mu)$. Hence, in expression (4) we use the re-weighting factor

$$\widehat{\psi}_{\tau_y,\tau_\mu} = \frac{P_\mu((t_m, t_f)|\tau_\mu)}{P_\mu((t_m, t_f)|\tau_y)}. \tag{5}$$

in place of ψ_{τ_y,τ_μ}. Intuitively, household types that are more (less) common in τ_y than in τ_μ get a weight lower (greater) than one in the counterfactual income distribution.

In sum, to implement scenario (i), we compute counterfactual income expression (4) and re-weighting factor (5) when $\tau_y = \tau \in [1980, 2018]$ and $\tau_\mu = 1980$: $\widehat{F_Y}(y|\tau_y = \tau, \tau_\mu = 1980)$ and $\widehat{\psi}_{\tau_y=\tau,\tau_\mu=1980}$. In this counterfactual scenario the configuration of households stays at its 1980 level for any year τ in which we measure income. Importantly, our choice of weights further guarantees that those who are married in year τ exhibit the degree of assortative matching $\mathcal{S}$ of the 1980 households.[29]

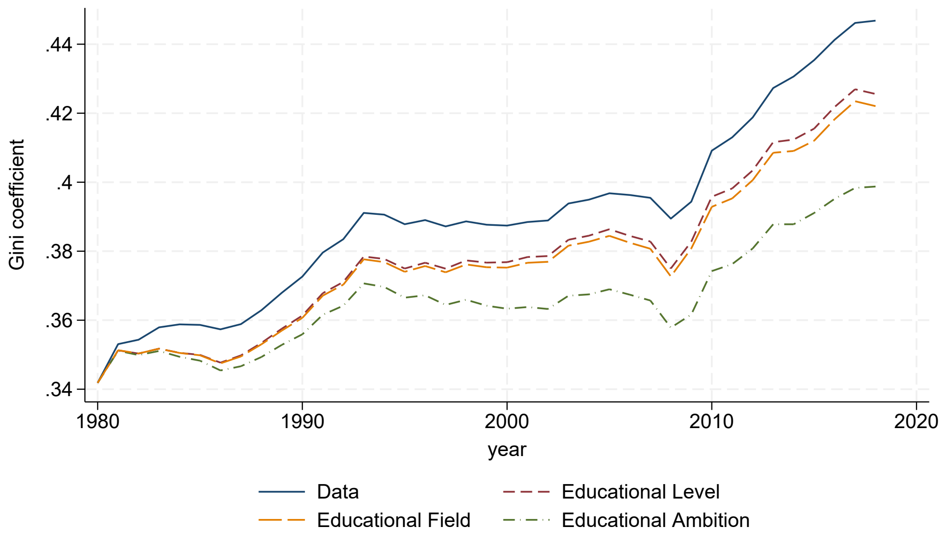
(ii) Fixed returns. To put the contribution of the marriage market to rising inequality into perspective, we again build on the DFL method to analyze how income inequality would have developed had the household income by household education composition (that is, the *returns* to education) remained unchanged while the composition of households evolves as in the data. We interpret this scenario as holding the labor market returns to education-based types fixed. Specifically, we compute the counterfactual income distribution (4) when $\tau_y = 1980$ and $\tau_\mu = \tau \in [1980, 2018]$, that is, $\widehat{F_Y}(y|\tau_y = 1980, \tau_\mu = \tau)$. Once again, in practice, we use a re-weighting factor (5) but when income is held fixed at 1980, that is, $\widehat{\psi}_{\tau_y=1980,\tau_\mu=\tau}$.

Once we have constructed the counterfactual household income based on the definition of education and counterfactual scenario, we create counterfactual moments of income inequality.

In Figure 4 we show that if the marriage market had stayed fixed at its $\tau_\mu = 1980$ composition, household income inequality at any year $\tau > 1980$ would have been mitigated relative to the trend in the data. In the figure, we plot, for all years $\tau_y = \tau \in [1980, 2018]$, both the

[29] In Appendix A.5, we show this result formally. Additionally, our choice of weights implies that the log odds ratios are also held fixed at their 1980 values.

Figure 4: Changes is Sorting Patterns by Ambition Amplified Inequality



Notes: The plot shows the development of the Gini coefficient of household income between 1980–2018 in the data (blue solid line), and in the counterfactual scenario (i) (fixed sorting patterns, explained in Section 5.1) for educational level (red short-dashed line), educational field (orange long-dashed line), and educational ambition (green dash-dotted line) types.

observed Gini coefficient (blue solid line) and the counterfactual scenario (i) Gini coefficients for each of the three categorizations (Educational Level in red short-dashed line, Educational Field in yellow long-dashed line, and Educational Ambition in green dot-and-dashed line). Overall, we see that inequality would have increased less, regardless of the categorization, since all counterfactual trends lie below the observed trend in the data. However, there are important quantitative differences. The red short-dashed and the orange long-dashed trends that hold marriage market matching by educational level and educational field fixed, respectively, are much closer to the data than the green dot-and-dashed line for educational ambition. These results suggest that changes in marital sorting patterns and in assortative matching based on educational ambition can account for a larger share of changes in between-household inequality than sorting based on educational level or field. One reason is that highly ambitious women have increasingly entered marriage (the extensive margin), recall Figure B.3. Another reason is that positive sorting based on educational ambition (the intensive margin) has become more pronounced, while sorting based on levels and fields of education has barely increased, recall Figure 3.

In Table 3, we compare the contributions of the marriage market and returns to education for the different categorizations and extend the analysis by considering additional income distribution moments (percentile ratios). We decompose the total change in between-household income inequality between 1980 and 2018. Column (a) shows the results for the Gini coefficient,

Table 3: Decomposing Changes in Income Inequality: Marriage Market vs. Returns

	(a) Gini		(b) $^{P90}/_{P50}$		(c) $^{P50}/_{P10}$	
Factual change (Δ_{Data})	0.105	100%	0.584	100%	3.760	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 patterns						
Educational Level	0.084	80%	0.385	66%	4.423	118%
Educational Field	0.080	76%	0.370	63%	4.375	116%
Educational Ambition	0.057	54%	0.204	35%	2.825	75%
(ii) Fixed returns						
Educational Level	0.040	38%	0.320	55%	0.400	11%
Educational Field	0.036	35%	0.305	52%	0.451	12%
Educational Ambition	0.048	46%	0.361	62%	0.590	16%

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 Section 5.1. 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 factual inequality changes in the data. For each of the counterfactual scenarios (i)-(ii), 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 within each counterfactual case shows the columns' statistic for one of the three definitions of types (as explained in Section 3): educational level, educational field, and educational ambition types.

which summarizes inequality in the entire distribution; columns (b) and (c) show the 90/50 and 50/10 percentile ratios, respectively. For each inequality measure, the first row contains the inequality changes in the data (Δ_{Data}). Between-household income inequality has increased according to all three measures. The Gini coefficient has increased by 0.105 (from 0.342 to 0.447), the 90/50 percentile ratio has increased by 0.584 (from 1.901 to 2.485), and the 50/10 percentile ratio has even increased by 3.760 (from 3.625 to 7.385). These changes correspond to 100%.

First, we find that changing labor market returns are a major source of increasing income inequality, see Panel (ii) of Table 3. Without increasing returns, the counterfactual increase in the Gini is 38% of the observed increase for educational-level types, 35% for educational-field types, and 46% for educational-ambition types. That is, without the rising income premia that highly-educated individuals receive relative to less-educated individuals, inequality would have increased by less than 50% of the observed increase, and this conclusion holds irrespective of how we construct marital types. Compared to the marriage market, returns to education are the more potent driving factor of inequality in all but one case (the 90/50 percentile ratio for educational ambition).

There are some interesting differences between the effects of fixed returns in the upper and lower halves of the income distribution; see columns (b) and (c) of Panel (ii). In the upper half (90/50 ratio), we see that increasing returns contributed less to the inequality trend. For educational level, field, and ambition types, the 90/50 ratio still increases by 55%, 52%, and 62%, respectively, relative to the data. In the lower part of the income distribution (50/10 ratio), increasing returns can explain almost the entire increment of the percentile ratio in the data. Without increasing returns, the inequality below the median would have barely changed, the counterfactual 50/10 ratios are between 11% and 16% of the observed increases for each categorization. Together, these findings suggest an important role for increasing returns to education for inequality, but differences between categorizations are small.

This is different in the fixed sorting patterns scenario, where the categorization matters greatly for conclusions about the role of the marriage market for the inequality trends, see Panel (i) of Table 3. This is our main result: with a fixed marriage market, the counterfactual 2018 Gini coefficient amounts to 54% of the true coefficient with educational ambition types. That is, more than 40% of the increase since 1980 can be explained with changing matching patterns based on educational ambition. Using the common educational level categorization, we do not arrive at the same conclusion. With 80% for levels and 76% for fields of the observed inequality increment, sorting patterns matter much less according to these categorizations.

Marriage market matching based on educational ambition also stands out by explaining a greater share of increasing inequality for the two other income distribution moments. The 90/50 percentile ratio would have increased by two-thirds of the factual increase with fixed matching based on fields and levels but only by one-third with fixed matching based on educational ambition. That is, the changing composition of households in terms of educational ambition is a more powerful explanatory factor with respect to increasing inequality compared to educational levels and fields. The increase in the 50/10 ratio is even mitigated due to the changing composition of households through the lens of levels and fields, while the compositional changes in terms of ambition amplified household income inequality along the entire income distribution.

The differences in the fixed sorting patterns scenario across education categorizations are driven by differences in the reweighting factors that we apply to the observed distribution of household income in 2018. These weights, in turn, are derived directly from changes in the composition of households over 1980—2018. Because the counterfactual inequality measures based on educational levels or fields are closer to the observed income distribution in 2018, our results suggest that changes in the household composition based on these categorizations were more moderate than for educational ambition. This shows that how we define education significantly affects our conclusions on how the marriage market affects inequality between households.

5.2 Fixed Marital Surplus vs. Fixed Marginal Distributions

We now investigate what particular aspect of the marriage market contributed to amplifying the rise in household income inequality. Because the supply of educations types, the decision to marry, and the decision of whom to marry are endogenously related, disentangling them requires some additional structure. We follow the rapidly evolving literature (Chiappori et al., 2020a,b, 2025) that use the well-known Separable Extreme Value (SEV) model brought to the marriage market literature by Choo and Siow (2006) to guide this decomposition. We closely follow Chiappori et al. (2020a) who apply the SEV structure to link changes in sorting to the rise in inequality in the UK.

Specifically, we decompose the role of trends in marital sorting to explain changes in inequality into the contribution of changing marginal distributions of education versus that of the evolving value from marriage relative to singlehood. To do so, we combine the DFL approach—which we continue to use to construct the re-weighting factors—with the SEV structure—which allows us to isolate the role of changing marginal distributions from that of changes in the value of different types of couples. The framework introduced by Choo and Siow (2006) models individuals' decision on whom to marry or whether to remain single based on the value of their potential matches relative to the value of remaining single—that is, the marital surplus. A distinctive feature of the SEV structure is that the marital surplus is the sum of two components: a *systematic* component that only depends on the type composition of the potential couple, and a random idiosyncratic taste shock for marital alternatives that only depends on the type of the potential partner. The random component is assumed to follow a type-I generalized extreme value (Gumbel) distribution.

Importantly for our purpose, Choo and Siow (2006) exploit the properties of the SEV structure to derive a closed form relationship between the systematic component of the marital surplus generated by couple type (j, j')—which we denote by $\Pi_{jj'}$—and the the ratio of the observed share of couples type (j, j') and the geometric average of the single shares of individuals type j and j':

$$\Pi_{jj'} = \frac{P_\mu\left(t_m = j, t_f = j'\right)}{\sqrt{P_\mu\left(t_m = j, t_f = \emptyset\right) P_\mu\left(t_m = \emptyset, t_f = j'\right)}}, \tag{6}$$

where $P_\mu(t_m = j, t_f = \emptyset)$ $(P_\mu\left(t_m = \emptyset, t_f = j'\right))$ is the probability of observing a man of type j (women of type j') without a partner.

We combine the marital surplus (6) with the feasibility conditions in the marriage market whereby the mass of individuals type t equals the sum of the mass of individuals of that type who marry and who remain single. We denote the mass of single women type t_f and of single men type t_m by $P_\mu(\emptyset, t_f)$ and $P_\mu(t_m, \emptyset)$, respectively. For a particular categorization $\tilde{x}$, we denote the mass of women type t_f and the mass of men type t_m, respectively by

$$\begin{aligned} P_{\tilde{x}}(t_f) &= \sum_{t_m} P_\mu(t_m, t_f) + P_\mu(\emptyset, t_f) \\ P_{\tilde{x}}(t_m) &= \sum_{t_f} P_\mu(t_m, t_f) + P_\mu(t_m, \emptyset) \end{aligned} \tag{7}$$

Second, we closely follow the structural approach by Chiappori et al. (2020a) to construct two counterfactual sorting matrices from equation (6) and two corresponding counterfactual re-weighing factors for the household income distribution of 2018.

Formally, let $\tau_{\tilde{x}}$ be the year in which the marginal distribution of types based on definition $\tilde{x}$ is measured, and let τ_Π be the year in which the marital surplus is measured. We define the income distribution when household income is measured in year τ_y, the marginal distributions of $\tilde{x}$ in $\tau_{\tilde{x}}$, and the marital surplus Π in τ_Π as

$$\widehat{F_Y}(y|\tau_y, \tau_{\tilde{x}}, \tau_\Pi) = \int F_{Y|\mu}(y|(t_m, t_f), \tau_y)\psi_{\tau_y, \tau_{\tilde{x}}, \tau_\Pi} d\mu((t_m, t_f)|\tau_y). \tag{8}$$

As in counterfactual cases (i) and (ii), the factor $\psi_{\tau_y, \tau_{\tilde{x}}, \tau_\Pi}$ allows us to re-weight households in the τ_y data to reflect their frequency had the education marginals and marital surplus stayed at their $\tau_{\tilde{x}}$ and τ_Π compositions, respectively. In this case in which we distinguish between marginals and surplus, we have

$$\psi_{\tau_y, \tau_{\tilde{x}}, \tau_\Pi} = \frac{d\widehat{\mu}((t_m, t_f)|\tau_{\tilde{x}}, \tau_\Pi)}{d\mu((t_m, t_f)|\tau_y)}.$$

Just as in counterfactual scenarios (i) and (ii), the numerator of the re-weighting factor is not observed, so it has to be estimated. However, an additional complication here is that when $\tau_{\tilde{x}} \neq \tau_\Pi$, we do not have a frequency table $P_\mu((t_m, t_f)|\tau_{\tilde{x}}, \tau_\Pi, \tau_{\tilde{x}} \neq \tau_\Pi)$ as a data counterpart anymore, because only marriage markets in which the marginal type distribution and the marital surplus are measured in the same year ($\tau_{\tilde{x}} = \tau_\Pi$) realize. To circumvent this challenge, we retrieve these counterfactual frequency tables, $\widehat{\mu}$, with the help of the SEV model. For example, when we fix marginals (7) at time $\tau_{\tilde{x}}$ and let the marital surplus be as observed in year $\tau > \tau_{\tilde{x}}$, we use the system of equations (6) and (7) in two steps: First, we compute the *counterfactual* fractions of singles that would close the marriage market in τ when the marginal distributions are as in $\tau_{\tilde{x}}$, and, second, we compute the implied *counterfactual* fractions of couples. We end up with a counterfactual matching matrix, $\widehat{\mu}$, specifying the frequency of each type of household—including the singles—with elements denoted by $P^{SEV}_{\widehat{\mu}}((t_m, t_f)|\tau_{\tilde{x}}, \tau_{\Pi})$.

Once we have calculated the counterfactual household compositions, $\widehat{\mu}$, in which the marginals are measured in year $\tau_{\tilde{x}}$ and the surplus in year τ_{Π}, we can estimate the re-weighting factor by

$$\widehat{\psi}^{SEV}_{\tau_y,\tau_{\tilde{x}},\tau_{\Pi}} = \frac{P^{SEV}_{\widehat{\mu}}((t_m, t_f)|\tau_{\tilde{x}}, \tau_{\Pi})}{P_{\mu}((t_m, t_f)|\tau_y)} \tag{9}$$

and use it to compute the counterfactual income distribution (8). With this more general income distribution specification, we solve for two additional counterfactual cases (iii) and (iv), as described next.

(iii) Fixed marital surplus. In one counterfactual, we fix marital surplus at $\tau_{\Pi} = 1980$, while we measure income and education marginals at any year $\tau > 1980$. Specifically, we use counterfactual re-weighting factor $\widehat{\psi}^{SEV}_{\tau_y=\tau,\tau_{\tilde{x}}=\tau,\tau_{\Pi}=1980}$ to obtain the counterfactual income distribution $\widehat{F_Y}(y|\tau_y = \tau, \tau_{\tilde{x}} = \tau, \tau_{\Pi} = 1980)$.

(iv) Fixed marginals. In the last counterfactual scenario, we fix the marginal distributions of education at $\tau_{\tilde{x}} = 1980$, while we measure income and marital surplus at any year $\tau > 1980$. Specifically, we use counterfactual re-weighting factor $\widehat{\psi}^{SEV}_{\tau_y=\tau,\tau_{\tilde{x}}=1980,\tau_{\Pi}=\tau}$ and counterfactual income distribution $\widehat{F_Y}(y|\tau_y = \tau, \tau_{\tilde{x}} = 1980, \tau_{\Pi} = \tau)$. We also keep either the male (case iv.a) or the female (case iv.b) marginal type distributions fixed. In those cases, we further distinguish between marginals by gender in expressions (8) and (9).

Table 4, which is structured similar to Table 3, shows the numerical results for scenarios (iii) and (iv), and variants (iv.a) and (iv.b).

We first keep the surplus fixed for all couple-type combinations, in Panel (iii). When we generate households in 2018 based on the 1980 gains from marriage and the 2018 marginal distributions, we find that inequality would have increased less than in the data. Interestingly, this result holds for all categorizations of education. For levels, fields, and ambition types, holding marital surplus fixed at the 1980 level explains about 30% of the inequality increment based on the Gini coefficient. We reach similar conclusions when we measure income inequality based on the percentile ratios.[30]

[30] This result holds similarly when limiting the analysis to cohorts of individuals aged 40-49 in each year, respectively. Here, fixed marital surplus for ambition types explains 34% of the increase in the Gini between 1980 and 2018, while it captures 29% for educational levels and 32% for fields, respectively. This analysis focuses only on inequality in a subset of the population, but it captures changes in persistent family circumstances over time (Chiappori et al., 2020a).

Table 4: Decomposing Changes in Income Inequality: Marital Surplus vs. Marginals

	(a) Gini		(b) $P90/P50$		(c) $P50/P10$	
Factual change (Δ_{Data})	0.105	100%	0.584	100%	3.760	100%
	Δ_{Gini}	$\frac{\Delta_{Gini}}{\Delta_{Data}}$	$\Delta_{P90/P50}$	$\frac{\Delta_{P90/P50}}{\Delta_{Data}}$	$\Delta_{P50/P10}$	$\frac{\Delta_{P50/P10}}{\Delta_{Data}}$
(iii) Fixed marital surplus						
Educational Level	0.073	69%	0.324	55%	2.629	70%
Educational Field	0.073	70%	0.332	57%	2.741	73%
Educational Ambition	0.072	68%	0.325	56%	2.707	72%
(iv) Fixed marginals						
Educational Level	0.137	130%	0.761	130%	7.432	198%
Educational Field	0.134	128%	0.752	129%	7.419	197%
Educational Ambition	0.110	105%	0.609	104%	4.965	132%
(iv.a) Fixed marginals (male)						
Educational Level	0.108	103%	0.603	103%	4.499	120%
Educational Field	0.108	102%	0.603	103%	4.521	120%
Educational Ambition	0.104	99%	0.581	99%	4.007	107%
(iv.b) Fixed marginals (female)						
Educational Level	0.127	121%	0.652	112%	6.169	164%
Educational Field	0.127	121%	0.664	114%	6.184	164%
Educational Ambition	0.111	106%	0.611	105%	4.574	122%

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 Section 5.2. 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 factual inequality changes in the data. For each of the counterfactual scenarios (iii)-(iv.b), we first report the counterfactual change, Δ_{Gini}, and then the counterfactual change relative to the change in the data, $\Delta_{Gini}/\Delta_{Data}$. Each row within each scenario shows the columns' statistic for one of the three definitions of types (as explained in Section 3): educational level, educational field, and educational ambition types.

In the final marriage market counterfactual, we first keep the marginal type distributions for both genders fixed at the 1980 level, see Panel (iv) in Table 4. We then repeat the exercise keeping either the male or the female marginal type distribution fixed, see Panels (iv.a) and (iv.b). The marginal distributions shifted such that the numbers of individuals in the top categories increased, and this change is more pronounced for women.[31] 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 changing educational attainment had a mitigating effect on inequality. Without the shift, inequality would have increased to 130% (128%) of the actual 2018 value. The mitigating effect manifests itself mainly in the lower half of the income distribution. The 50/10 percentile ratio

[31] The share of men (women) with long-cycle tertiary education increased by a factor of 3 (11) between 1980 and 2018. For educational ambition types, the share of men (women) in the top category doubled (increased five-fold), as shown in Figure B.1 in the Online Appendix.

in column (c) would have been two times higher without changing marginal distributions. For the 90/50 ratio in column (b), we see more modest mitigating effects for educational level and field types (130% and 129%), similar in magnitude to the results for the Gini in column (a). Based on educational ambition types, we do not find a significant mitigation (or amplification) for the Gini or the 90-50 ratio based on fixed marginals (105% and 104%). Only in the lower half of the distribution do we find a noticeable mitigating effect (132%).

If we keep only the female marginal distributions fixed at the 1980 level, overall, the conclusions from this counterfactual exercise hardly change. The results in Panels (iv) and (iv.b) in Table 4 are similar for all categorizations. The mitigating effects for fields and levels are somewhat less pronounced. This similarity implies that changes to the female marginal distributions drive the mitigating effects. If we instead keep the male marginal distributions fixed (iv.a), the mitigating effects on inequality are considerably smaller and the implied household income distributions are overall close to the data, apart from a small mitigation in the lower half of the distribution for levels and fields.

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 did not mitigate the rise in inequality. Differences across categorizations for the male type distributions are much smaller.

6 Robustness

In this section, we assess whether our main results are sensitive to constructing ambition types in alternative ways. 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 the k-means mapping algorithm $\mathcal{T}_{Ambition}(\tilde{x})$. Table 5 shows the trends in the degree of assortative matching (analyzed in Section 4) and the inequality contributions in the fixed sorting patterns counterfactual scenario (analyzed in Section 5, scenario (i)) that we find when constructing the ambition types in various alternative ways. For convenience, the first row repeats the results for our benchmark categorization of ambition types. The columns show: (a) the the numbers of observations in 1980 and 2018, (b) the change in our aggregate sorting measure, equation (2), between those years, (c) the observed Gini coefficients in 1980 and 2018, and (d) the counterfactual Gini in the

Table 5: Assortative Matching and Fixed Sorting Patterns with Alternative Ambition Types

	(a)		(b)			(c)		(d)	
	N (1,000s)		Assortative Matching			Gini, data		Gini, (i)	$\frac{\Delta_{Gini,(i)}}{\Delta_{Gini,data}}$
	1980	2018	1980	2018	Change	1980	2018	2018	
Benchmark types	2,427	2,801	1.29	1.55	20.1%	0.342	0.447	0.399	54.2%
(i) Gender- and cohort-specific ambition types									
Types by gender	2,426	2,798	1.15	1.37	18.6%	0.342	0.447	0.407	61.9%
Types by cohort	2,408	2,799	1.34	1.58	18.2%	0.342	0.447	0.407	61.7%
(ii) Different numbers of clusters									
Three ambition types	2,427	2,801	1.23	1.42	15.2%	0.342	0.447	0.407	62.1%
Five ambition types	2,427	2,801	1.39	1.73	24.9%	0.342	0.447	0.396	51.2%
(iii) Different clustering variables									
Only g	2,427	2,801	1.27	1.40	9.9%	0.342	0.447	0.412	66.4%
Only w_0	2,427	2,801	1.35	1.56	15.7%	0.342	0.447	0.409	62.9%
Male w_0, g	1,191	2,775	1.40	1.60	14.2%	0.375	0.448	0.410	47.7%
w_0, part-time penalty	2,345	2,725	1.37	1.50	9.2%	0.342	0.448	0.404	58.2%
(iv) Different level of aggregation									
Programs × univ.	2,409	2,795	1.28	1.55	20.7%	0.340	0.447	0.399	54.8%
Sub-fields × levels	2,427	2,642	1.32	1.51	14.4%	0.342	0.446	0.403	58.3%
(v) Lifetime earnings sample 1990 – 2010									
	1990	2010	1990	2010	Change	1990	2010	2010	
Lifetime earnings	1,555	1,578	1.52	1.56	3.0%	0.385	0.403	0.396	59.9%
Benchmark types	1,555	1,578	1.40	1.55	11.0%	0.385	0.403	0.396	57.1%
Benchmark, full sample	2,676	2,705	1.37	1.50	9.7%	0.373	0.435	0.390	47.1%

Notes: Columns (a) display the number of observations in each case in thousands of individuals for both 1980 and 2018. Columns (b) show our measure of marriage market sorting $\mathcal{S}$ derived in Section 4 for both years along with the percentage change. Columns (c) show the observed Gini coefficients in each case for 1980 and 2018. Columns (d) show the counterfactual Gini coefficient in the fixed marriage market scenario (i), i.e., had the composition of households stayed fixed at its 1980 level in each case and the fraction of the observed change in Gini that is captured by the change in the counterfactual scenario, $\Delta_{Gini,(i)}/\Delta_{Gini,data}$. Each row shows the columns' statistic for one of the alternative definitions of ambition types.

fixed sorting patterns scenario along with the explained share relative to the trend in the data.

In Panel (i), we analyze the sensitivity of our findings with respect to the assumption that the ambition types do not vary by gender or over time. In the row labeled *Types by gender*, we consider the possibility that education programs may send different signals depending on the gender of the graduate. That is, we construct the four ambition types separately for women and men by using information at the program-gender level i. 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 B.6 (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. We find that the degree of assortative matching (AM) based on gendered ambition types increased slightly less than in our benchmark (18.6%, which is still significantly more than AM by levels or fields) and that the composition of couples explains slightly less of the changes in inequality than our benchmark does (almost 30%, which is significantly more than what we find using levels or fields).

Similarly, for 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). Here, 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 B.7 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. Both our main conclusions regarding the changes in AM and the relationship between trends in sorting patterns and inequality are unchanged when constructing the ambition types by cohort.

In Panel (ii), we analyze the sensitivity of our findings with respect to the assumption that there are exactly four ambition types. To this end, we repeat the analysis but define three and five ambition types instead of four. Once again, our main conclusions remain the same. While more (fewer) categories lead us to detect a slightly stronger (weaker) increase in AM on educational ambition of 24.9% (15.2%), our conclusions on the role of sorting patterns to rising inequality remain.

In Panel (iii), we vary the clustering variables that we use to construct ambition types. In the first two rows, we only use one our benchmark clustering variables, either wage growth g (the sub-vector of characteristics becomes $\tilde{x}_i = (g_i)$) or starting wages w_0 ($\tilde{x}_i = (w_{0i})$), respectively. In the third row, we use both clustering variables, but use only male outcomes to construct ambition types for both men and women. The concern behind this robustness check is that the endogenous labor supply choices of females, e.g., in response to child birth (Kleven et al.,

2019) could bias our way of capturing the signaling value of educational programs by affecting the average starting wages and, especially, wage growth of graduates. Therefore, we only use the starting wages and wage growth of male workers, who have more stable labor supply trajectories, to cluster programs. Finally, in the fourth row, we replace wage growth with one of our direct measures of career flexibility and work-life balance based: the part-time penalty (a measure of inflexibility, recall Section 3.3). In all four robustness checks, we find that ambition types constructed with alternative clustering variables continue to reveal significant increments in terms of assortative matching (9.2%–15.7%) and that changes in households' education composition explain significant shares of increasing between-household inequality (33.6%–52.6%). 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. Thus, our conclusion that AM on career ambition has increased over time and that changes in the marriage market contribute significantly to the rise in between-household income inequality is robust to constructing ambition types differently.

In Panel (iv), we assess the robustness of our findings to constructing ambition types at different levels of aggregation. First, the row labeled *Programs* $\times$ *university* uses a level of observation that is more granular than the education programs used for our benchmark categorization. Specifically, we distinguish between similar programs offered at different Danish universities and define $i' =$ programs $\times$ university as our unit of observation for the categorization. As our clustering variables, we use the same sub-vector as in our benchmark, $\tilde{x}_{i'} = (w_{0,i'}, g_{i'})$. The AM trend and explained share of the inequality increment by sorting patterns in this more disaggregated version of our approach are almost identical to the benchmark. This result is consistent with Appendix Figure B.8, which shows that labor market outcomes w_0 and g are very similar across graduates of the same degree program across universities. Second, the row labeled *Sub-fields* $\times$ *levels* 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.[32] The results only change slightly. AM increased by about 14% and marital sorting by ambition continues to

[32] 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).

explain more than 40% of the increasing inequality between households.

In the last Panel (v), we analyze the robustness of our results when clustering programs based on average lifetime earnings, thus abstracting from the heterogeneity in starting wages and wage growth conditional on lifetime earnings found in Table 2. Yet, we note that this approach comes with significant measurement challenges. To construct lifetime earnings, we need to observe program graduates for at least 30 years from the time of graduation. This leads to lower sample coverage of this alternative classification in early and late years of our analysis period, when a higher share of individuals have outdated or new degrees with missing information on lifetime earnings, respectively. This limitation affects couples more than singles because our analysis requires assigning types to both spouses to be included. As a result, we focus on a shorter time period from 1990 to 2010 for this robustness analysis because over this period the share of singles and couples covered by the lifetime-earnings classification is highest and stable. To interpret the magnitude of the results based on the lifetime-earnings classification, we also provide results for our benchmark ambition type classification for the 1990–2010 period. Specifically, we report results both for the smaller sample of households included in the lifetime-earnings analysis and for the full sample of households in these years.

Column (a) of Panel (v) shows that the number of observations we can use for the lifetime-earnings classification is about 58% in both 1990 and 2010, with similar coverage of singles and couples.[33] Most importantly, we find a small increase in AM by only 3% over 1990–2010 based on lifetime-earnings types. In contrast, the increase implied by our benchmark types is much larger, at 11% in the same sample, or 9.7% in the full sample over this period, amounting to about half of the total increase in AM that we measure over four decades in the main results (20.1%). Turning to inequality trends, the lifetime earnings sample understates the role of the marriage market because households with old or new degrees contribute significantly to the trends but are excluded here. This is evidenced by a larger increase in the observed Gini for the full sample, and a smaller counterfactual increase holding the marriage market fixed (compare the second and third row of columns (c) and (d), respectively). In addition, conditional on the smaller and selected sample, the increase in inequality explained by the household composition is also slightly smaller for lifetime-earnings types than when using our benchmark types.[34] Taken

[33]Specifically, we cover 53% of couples in 1990. For changes compared to baseline year 1980, the sample would be heavily skewed towards singles, with only less than 40% of couples included in the lifetime-earnings analysis.

[34]Over 1980–2018, we find a similar understatement of the AM increase based on lifetime-earnings types compared to our benchmark types. Yet, the substantial sample selection in the early and late years make the inequality trends difficult to compare to the main results, and we find no differences between types for the lifetime-earnings sample.

together, we find qualitatively similar patterns when clustering programs based on lifetime earnings, but the increase in AM and the role of the marriage market for rising inequality are understated because measurement requirements limit the sample and average lifetime earnings hide differences in life-cycle wage profiles across degree programs that matter in the marriage market.

7 Conclusion

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

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. We show that educational ambition better reflects both wage dynamics and work-life balance measures than categorizations based on educational levels and fields of study.

Our first main result shows an increase of about 20% 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 the marriage market 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 households based on educational ambition types stayed at their 1980 levels over the last four decades, between-household inequality growth would have been mitigated by approximately 40%. In contrast, changes in marriage market sorting patterns based on education level and field of study contributed less to income inequality growth. We also assess the contribution of the changing value of marriage and the changing supply of education types separately and find that both margins have contributed to explain inequality when we consider educational levels or fields, but they work in opposite directions. When we consider marital sorting based on ambition, only the evolution of the value of marriage relative to singlehood matters, which implies that changes in the value from marriage between ambition types amplified inequality.

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 the relationship between marriage and labor 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, Benjamin Friedrich, Ana Reynoso, Bastian Schulz, and Rune Vejlin (2023) "Marital Sorting and Inequality: How Educational Categorization Matters," IZA Discussion Paper 15912, IZA Institute of Labor Economics.

Almar, Frederik and Bastian Schulz (2024) "Optimal weights for marital sorting measures," *Economics Letters*, Vol. 234, p. 111497.

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.

Burtless, Gary (1999) "Effects of growing wage disparities and changing family composition on the U.S. income distribution," *European Economic Review*, Vol. 43, pp. 853–865.

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

Calvo, Paula, Ilse Lindenlaub, and Ana Reynoso (2024) "Marriage Market and Labour Market Sorting," *The Review of Economic Studies*, Vol. 91, pp. 3316–3361.

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, Carlo V. Fiorio, Alfred Galichon, and Stefano Verzillo (2022) "Assortative matching on income," JRC Working Papers in Economics and Finance 2022/7, Ispra.

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.

Chiappori, Pierre-André, Monica Costa Dias, Costas Meghir, and Hanzhe Zhang (2025) "Changes in Marital Sorting: Theory and Evidence from the US," Working Paper 33354, National Bureau of Economic Research.

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss (2017) "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, Vol. 107, pp. 2109–67.

Chiappori, Pierre-André, Edoardo Ciscato, and Carla Guerriero (2024) "Analyzing matching patterns in marriage: Theory and application to Italian data," *Quantitative Economics*, Vol. 15, pp. 737–781.

Choo, Eugene and Aloysius Siow (2006) "Who marries whom and why," *Journal of Political Economy*, Vol. 114, pp. 175–201.

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, and Bastian Schulz (2024) "Job displacement, remarriage, and marital sorting," IZA Discussion Paper 17335, IZA Institute of Labor Economics.

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.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019) "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, Vol. 11, p. 181–209.

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 (2024) "The Human Capital–Reproductive Capital Trade-Off in Marriage Market Matching," *Journal of Political Economy*, Vol. 132, pp. 552–576.

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

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

Mosteller, Frederick (1968) "Association and Estimation in Contingency Tables," *Journal of the American Statistical Association*, Vol. 63, pp. 1–28.

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

Reynoso, Ana (2024) "The impact of divorce laws on the equilibrium in the marriage market," *Journal of Political Economy*, Vol. 132, pp. 4155–4204.

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 Data and Measurement

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 (from IDAN).[35] 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). After 2008 we have the exact number of contractual hours.

[35] 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.

We run a regression of log hourly wages on year dummies and educational specific experience 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.[36]

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.[37] 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.

[36]Note that ambition types are assigned based on the *final* program, e.g., time-invariant, whereas educational level and field types are assigned based on the highest degree achieved by a given age. This is to be consistent with the previous literature on education-based marriage market types. In practice, this distinction has a neglectable impact because most individuals finish their studies in their 20s.

[37]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.

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.

A.3 Weights for the Aggregate Sorting Measure

Almar and Schulz (2024) derive a decision rule to minimize the distortion of likelihood-ratio-based sorting measures due to changing marginal distributions. While accounting for changing population shares ensures proper scaling of the likelihood ratios (see equation (1)), the marginal distributions also influence the sorting measure beyond this desired feature. To minimize this distortion, Almar and Schulz (2024) propose to set λ in equation (3) such that the combined distortion from changing male and female marginal distributions is minimized. In our data, the educational attainment of females was initially lower but increased relatively more compared to males. Thus, to the minimize the distortion, it is optimal to put the weight on the "short side" of the market (high-type females). Recall that λ $(1-\lambda)$ is the parameter of the male (female) marginal in the convex combination, and let γ_1 (γ_2) be the total distortional effect for men (women) in the total differential of the aggregate sorting measure (2). The optimal λ is

$$\lambda^* = \begin{cases} 0 & \text{if} \quad \text{sign}(\gamma_1) = \text{sign}(\gamma_2), \quad |\gamma_2| > |\gamma_1| \\ \frac{\gamma_1}{\gamma_1 - \gamma_2} & \text{if} \quad \text{sign}(\gamma_1) \neq \text{sign}(\gamma_2) \\ 1 & \text{if} \quad \text{sign}(\gamma_1) = \text{sign}(\gamma_2), \quad |\gamma_1| > |\gamma_2|. \end{cases} \tag{A.1}$$

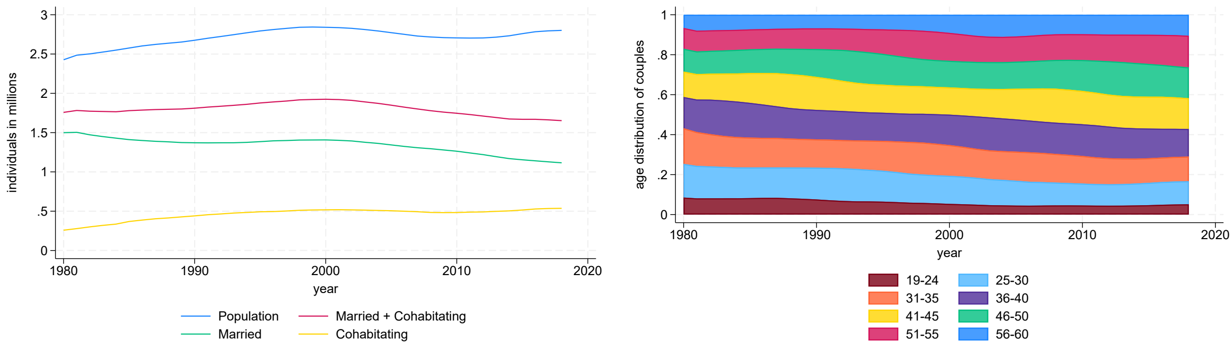
Intuitively, if the signs of the total distortional effect are similar for men and women (e.g., the share of "high-type" individuals could be increasing for both genders), the weight is set such that the bigger change (in absolute terms) is taken into account by the aggregate sorting measure. For example, if $\gamma_2 > \gamma_1$, female type distribution changes are relatively more important. To reflect this, changes of the male marginal distribution are taken out by setting λ equal to 0. The female marginal distribution cancels out in the likelihood ratios (1), removing its distortional impact. For the intermediate case $\text{sign}(\gamma_1) \neq \text{sign}(\gamma_2)$, $\lambda \in (0,1)$ ensures that both population share changes are reflected. The optimal λ^* may change over time, depending on the changing configuration of the marginals. Thus, the optimal weights may flip when, e.g., the population share of high-type women becomes greater than the population share of high-type men. To ensure that our sorting trends are comparable over time, we set λ to the modal value suggested by (A.1) over time. In the vast majority of years and for all categorizations, the rule suggests putting the weights on female type distribution changes.

A.4 Descriptive statistics

Figure A.1: Marriage, Cohabitation, Age Composition

(a) Population and 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	401347.0	664844.4	269760.8	1189937.8	474762.7
at graduation	(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.

A.5 Counterfactual Scenarios and Sorting

In this appendix section, we provide the formal derivation of the result that our choice of weights in Section 5.1, scenario (i), fixed sorting patterns, guarantees that those who are married in year τ exhibit the degree of assortative matching S of the 1980 households. In addition, the log odds ratios are also held fixed at their 1980 values.

To see this result, first note that our choice of weights for $\tau_y = \tau$ and $\tau_\mu = 1980$ is

$$\widehat{\psi}_{\tau,80}(i,j) \;=\; \frac{\frac{N_{80}(i,j)}{N_{80}^{\text{all}}}}{\frac{N_{\tau}(i,j)}{N_{\tau}^{\text{all}}}} \;=\; \frac{N_{80}(i,j)}{N_{\tau}(i,j)} \times \frac{N_{\tau}^{\text{all}}}{N_{80}^{\text{all}}} = \frac{P_{80}^{M}(i,j)}{P_{\tau}^{M}(i,j)} \times \underbrace{\frac{\sum_{r\neq\phi}\sum_{s\neq\phi} N_{80}(r,s)}{\sum_{r\neq\phi}\sum_{s\neq\phi} N_{\tau}(r,s)} \times \frac{N_{\tau}^{\text{all}}}{N_{80}^{\text{all}}}}_{:=A},$$

where $N_y(i,j)$ denotes the number of $(t_m = i, t_f = j)$ households (including singles if one of the types is $\emptyset$) and N_y^{all} is the total number of households in year y, while $P_y^M(i,j)$ denotes the fraction of married couples of type $(t_m = i, t_f = j)$ in year y.

Using equations (1) and (3) to write the aggregate sorting measure S in equation (2) explicitly, and applying the weights $\widehat{\psi}_{\tau,80}(i,j)$, yields the proposed result that $\widetilde{S}_{\tau} = S_{80}$:

$$\begin{aligned}
\widetilde{S}_{\tau} &= \sum_{t=1}^{T} \frac{P_{\tau}^{M}(t,t)\times\widehat{\psi}_{\tau,80}(t,t)}{\left(\sum_{s=1}^{T} P_{\tau}^{M}(t,s)\times\widehat{\psi}_{\tau,80}(t,s)\right)\left(\sum_{r=1}^{T} P_{\tau}^{M}(r,t)\times\widehat{\psi}_{\tau,80}(r,t)\right)} \\
&\qquad \times \left(\lambda \sum_{s=1}^{T} P_{\tau}^{M}(t,s)\times\widehat{\psi}_{\tau,80}(t,s) \;+\; (1-\lambda)\sum_{r=1}^{T} P_{\tau}^{M}(r,t)\times\widehat{\psi}_{\tau,80}(r,t)\right) \\
&= \sum_{t=1}^{T} \frac{P_{\tau}^{M}(t,t)\;\times\;\frac{P_{80}^{M}(t,t)}{P_{\tau}^{M}(t,t)}\;\times\; A}{\left(\sum_{s=1}^{T} P_{\tau}^{M}(t,s)\;\times\;\frac{P_{80}^{M}(t,s)}{P_{\tau}^{M}(t,s)}\;\times\; A\right)\left(\sum_{r=1}^{T} P_{\tau}^{M}(r,t)\;\times\;\frac{P_{80}^{M}(r,t)}{P_{\tau}^{M}(r,t)}\;\times\; A\right)} \\
&\qquad \times\;\left(\lambda(\sum_{s=1}^{T} P_{\tau}^{M}(t,s)\;\times\;\frac{P_{80}^{M}(t,s)}{P_{\tau}^{M}(t,s)}\;\times\; A) \;+\; (1-\lambda)(\sum_{r=1}^{T} P_{\tau}^{M}(r,t)\;\times\;\frac{P_{80}^{M}(r,t)}{P_{\tau}^{M}(r,t)}\;\times\; A)\right) \\
&= S_{80}\;\times\;\frac{A\;\times\;A}{A^2}\;=\;S_{80}
\end{aligned}$$

By a similar argument, we can show that the log odds ratios remain fixed. Consider types $i = \{t(1), ..., t(T-1)\}$ and $j = \{t(2), ..., t(T)\}$ for any categorization t. For any 2×2 submatrix of joint types in which $i \neq j$,

$$\begin{aligned}
\widetilde{I}_{\text{odds},\tau} &= \log\left(\frac{P_{\tau}^{M}(i,i)\;\times\;\widehat{\psi}_{\tau,80}(i,i)\;\times\;P_{\tau}^{M}(j,j)\;\times\;\widehat{\psi}_{\tau,80}(j,j)}{P_{\tau}^{M}(j,i)\;\times\;\widehat{\psi}_{\tau,80}(j,i)\;\times\;P_{\tau}^{M}(i,j)\;\times\;\widehat{\psi}_{\tau,80}(i,j)}\right) \\
&= \log\left(\frac{P_{80}^{M}(i,i)\;\times\;P_{80}^{M}(j,j)}{P_{80}^{M}(j,i)\;\times\;P_{80}^{M}(i,j)}\right) \;=\; I_{\text{odds},80}
\end{aligned}$$

B Additional Results

B.1 Differences in Demographic Changes Across Categorizations

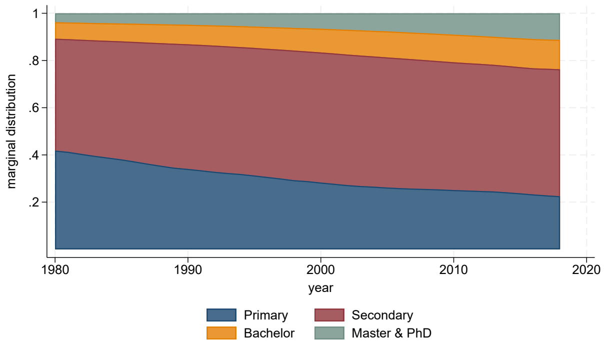
B.1.1 Trends in Education Distributions

How we view educational attainment and its evolution differs depending on which categorization we consider. Figure B.1 shows the fraction of men (left panels) and women (right panels) by education type and year, for our three definitions of education—Level in panels (a) and (b), Ambition in (c) and (d), and Fields in (e) and (f). Comparing the distribution of education levels and ambition types, we first note that for educational levels, the "secondary" category is large and relatively stable for both men and women, while the most frequent ambition types for both men and women are the ones that group education programs with high wage growth. Notably, while over time we observe an increasing representation of individuals, especially women, in Bachelor or more education levels relative to Primary education, we observe an increasing representation of the population in both the low wage growth and the highest ambition types.

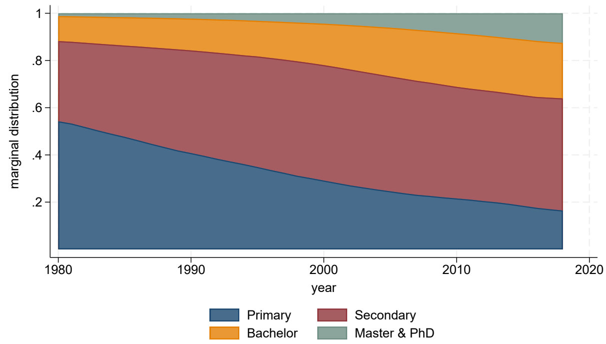
In addition, Figure B.2 shows the educational composition of married individuals over time. We observe that the composition of married individuals has shifted over time towards highly-educated (Master & PhD for both genders and also Bachelor degrees for women) and highly ambitious individuals (especially for women), but there are important differences across categorizations and gender. For example, based on ambition we also observe an increasing representation of low-ambition individuals among the married population.

Figure B.1: Marginal Type Distributions

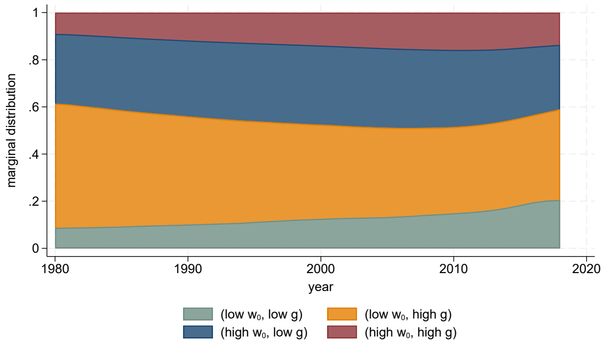
(a) Educational Level, Men



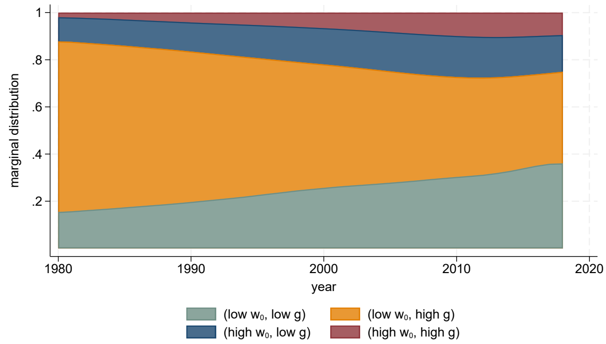
(b) Educational Level, Women



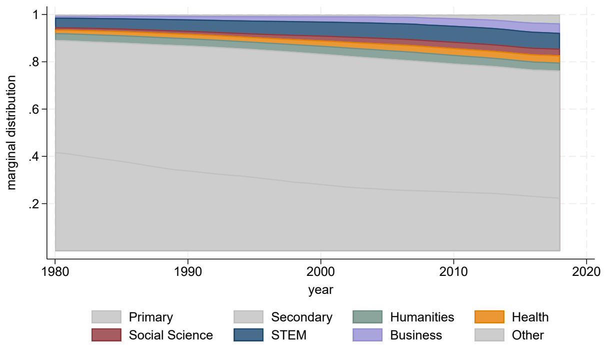
(c) Educational Ambition, Men



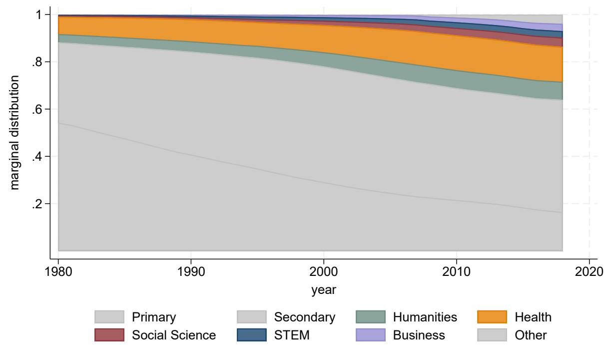
(d) Educational Ambition, Women



(e) Educational Fields, Men



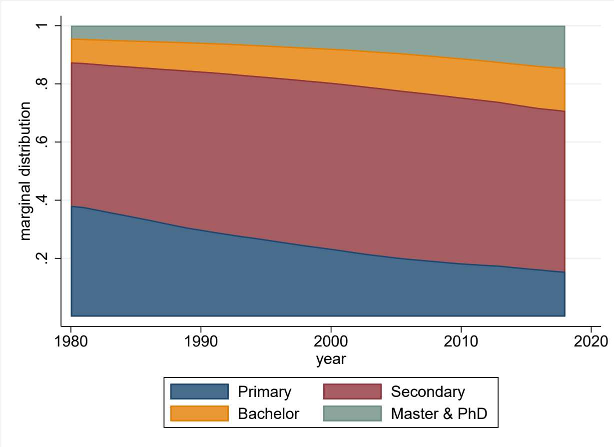
(f) Educational Fields, Women



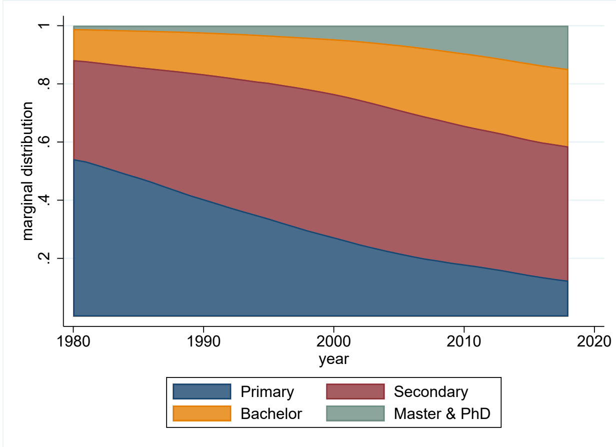
Notes: Marginal distributions for men and women over time by educational level and educational ambition. Sections 2 and 3 explain how the sample, educational levels, educational- fields and the labor market outcome (residualized log hourly wages) underlying the educational ambition types are constructed.

Figure B.2: Marginal Type Distributions, Conditional on Marriage

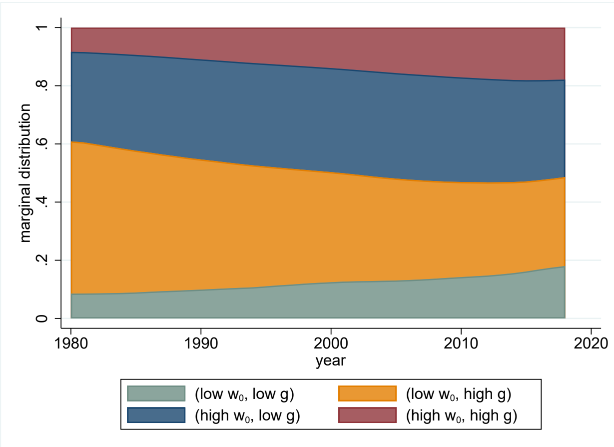
(a) Educational Level, Married Men



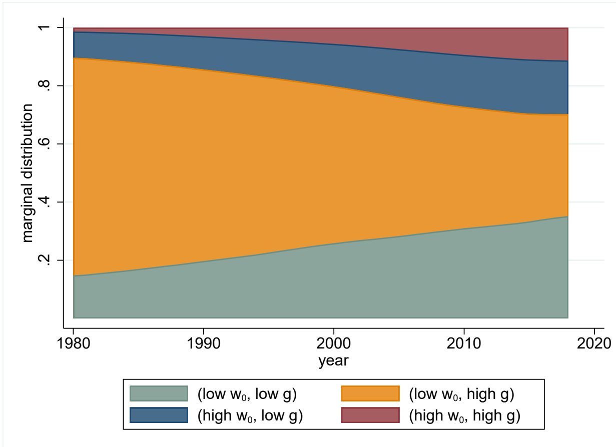
(b) Educational Level, Married Women



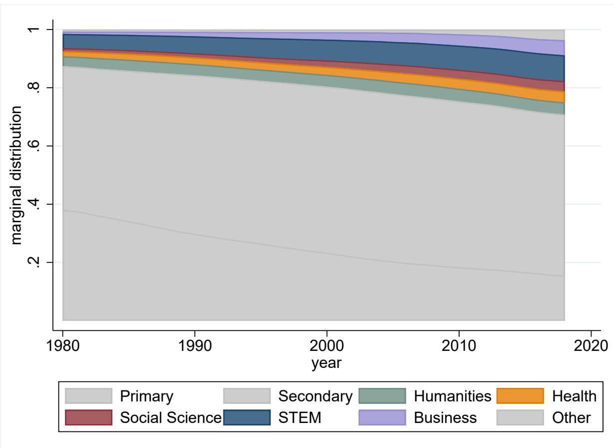
(c) Educational Ambition, Married Men



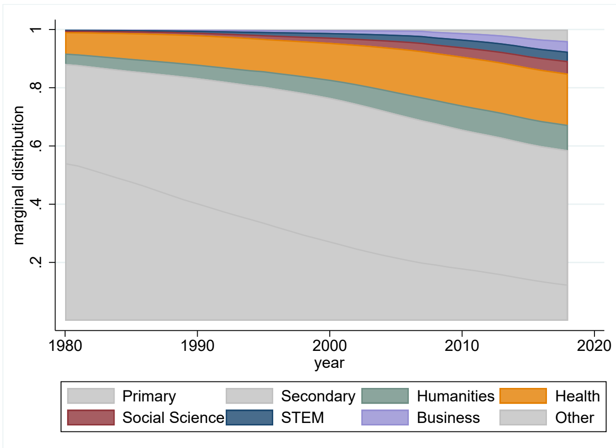
(d) Educational Ambition, Married Women



(e) Educational Fields, Married Men



(f) Educational Fields, Married Women



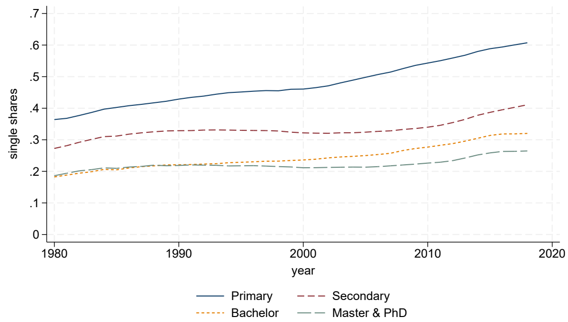
Notes: Marginal distributions for married and cohabiting men and women over time by educational level, educational ambition, and educational field. Sections 2 and 3 explain how the sample, educational levels, educational- fields and the labor market outcome (residualized log hourly wages) underlying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

B.1.2 Trends in Marriage

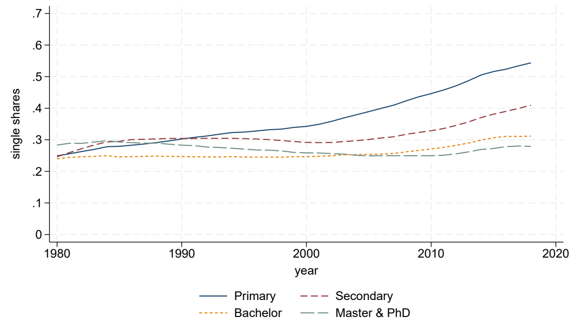
We analyze how individuals sort into being married or single based on their education type and make comparisons across categorizations. In Figure B.3, we show how single shares of men and women have evolved for all the types included in our educational-level, educational-ambition, and educational-field categorizations.

Figure B.3: Single Shares for Educational-Level, Ambition Types, and Fields of Study

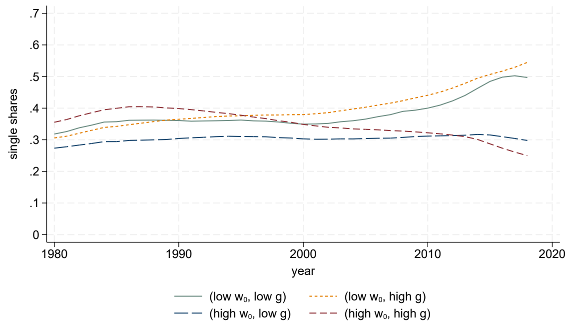
(a) Educational Level (male)



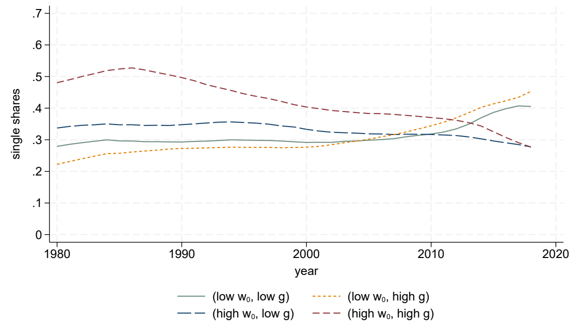
(b) Educational Level (female)



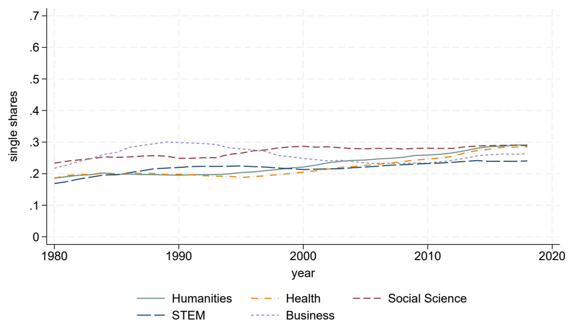
(c) Educational Ambition (male)



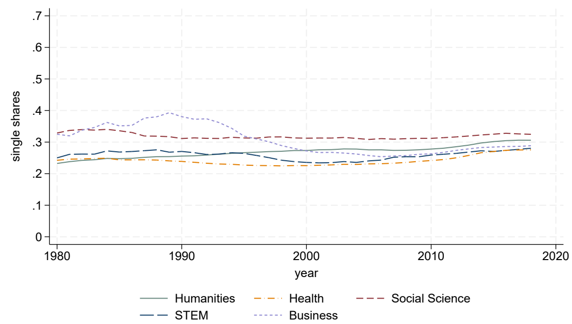
(d) Educational Ambition (female)



(e) Educational Field (male)



(f) Educational Field (female)



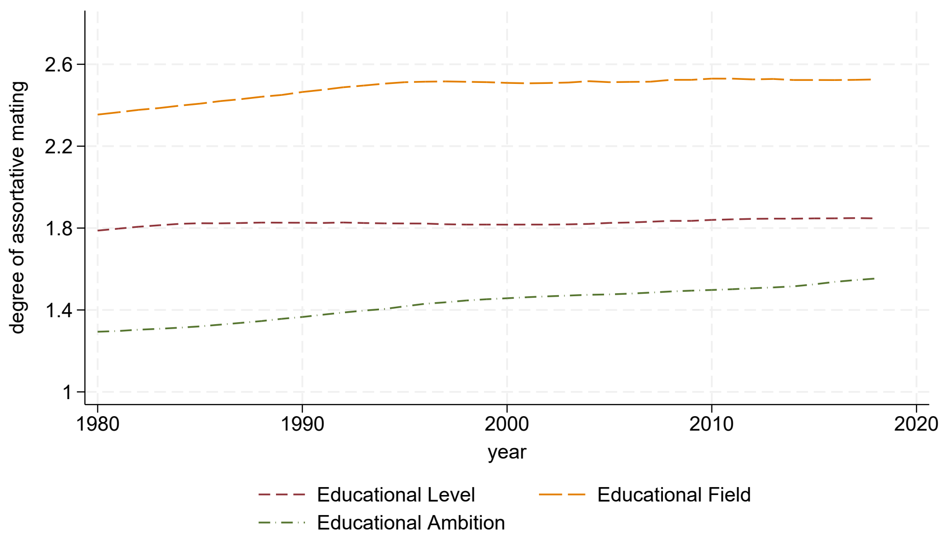
Notes: Single shares by gender for all for types of the educational level (Panels a and b), educational ambition (Panels c and d), and educational field (Panels e and f) categorizations. Types are constructed as explained in Section 3.

For educational-level types, single shares have increased for all types for both men and women, and most significantly so for individuals with primary education (blue solid lines in Panels (a) and (b), Figure B.3). In contrast, if we interpret the data through the lens of our novel ambition types, single shares did not increase for all types. Specifically, men and women who graduate from ambitious educational programs with high starting wages and high wage growth (red dashed lines) exhibit falling single shares. This is especially pronounced for high-type women, who had the highest probability of remaining single in 1980, when their single share was close to 50%. It decreased to around 30% in 2018. If more women with high earnings potential enter marriage, inequality between households might increase, depending on the partner type and couples' labor supply choices.

Single shares increased for both men and women who graduate from educational programs with low starting wages (gray solid and yellow short-dashed lines). The blue dashed lines—the single shares of men and women who hold degrees with high starting wage but low wage growth—changed relatively little but decreased slightly for women. These striking differences in single share trends across categorizations show that the ambition types reveal fundamental changes in marriage market matching that remain undetected using standard educational-level types and potentially matter for inequality trends.[38]

B.2 Trends in Sorting measures

Figure B.4: Aggregate Sorting Measures in Levels

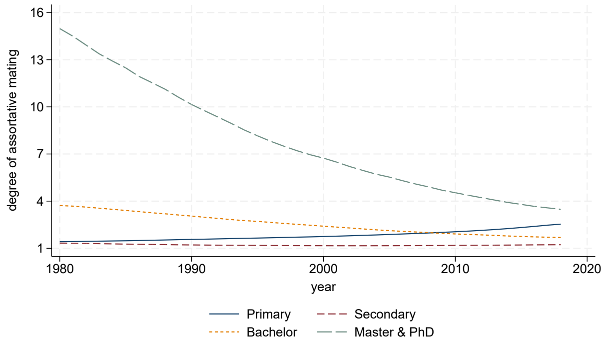


Notes: The figure shows the sorting measure $\mathcal{S}$ derived in Section 4, equation (2) for educational level types (red short-dashed line), educational field types (orange long-dashed line) and educational ambition types (green dash-dotted line). Types are constructed as explained in Section 3.2.

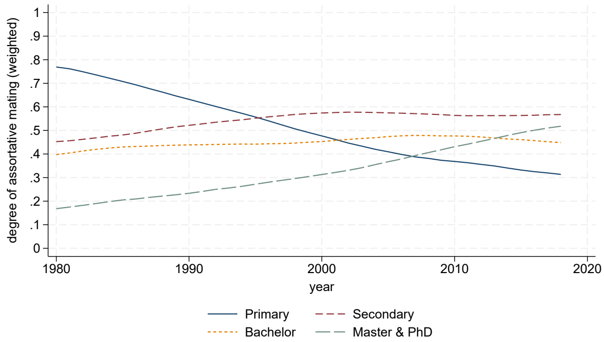
[38]The single-share trends by educational field are rather flat, see Panels (e) and (f). They are different from both educational levels and ambition, which underscores our point that the role that the extensive margin plays depends on the categorization.

Figure B.5: Type-specific Likelihood Indices, unweighted and weighted

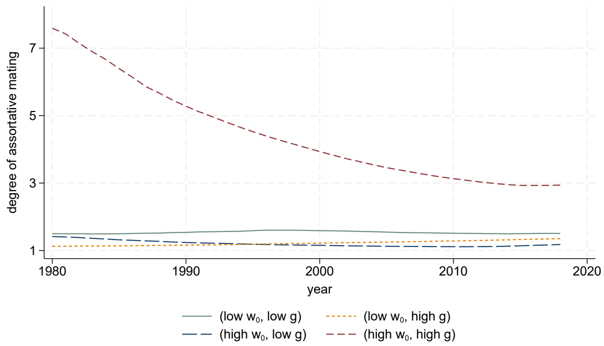
(a) Educational Level



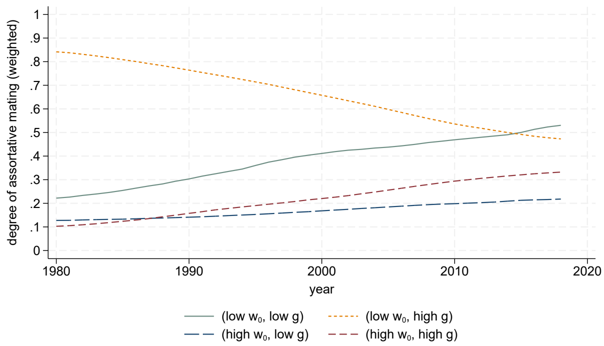
(b) Educational Level (weighted)



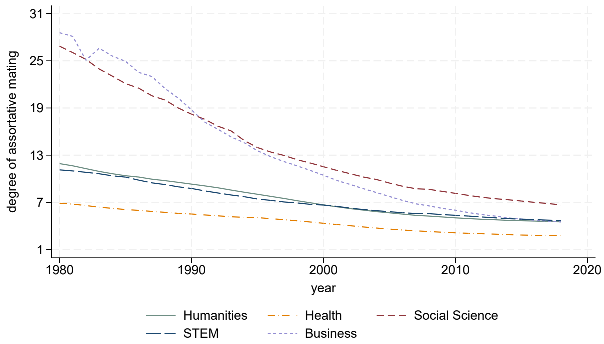
(c) Educational Ambition



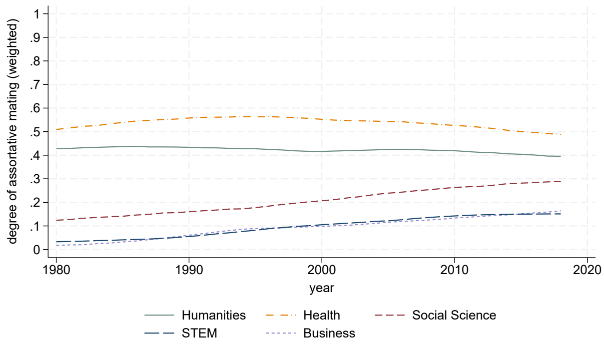
(d) Educational Ambition (weighted)



(e) Educational Field



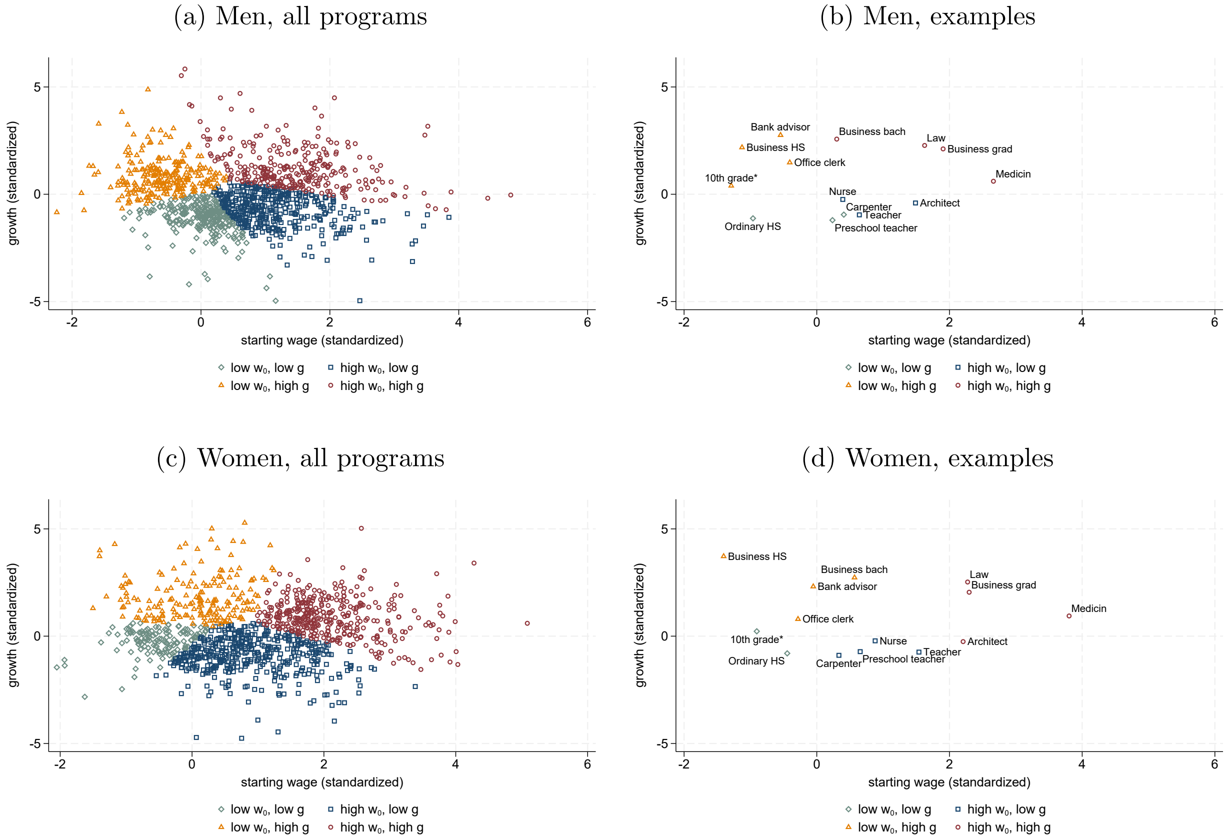
(f) Educational Field (weighted)



Notes: Likelihood indices for equal-education couples (equation (1)) for educational level, educational ambition, and educational field categorizations. Types are constructed as explained in Section 3.2.

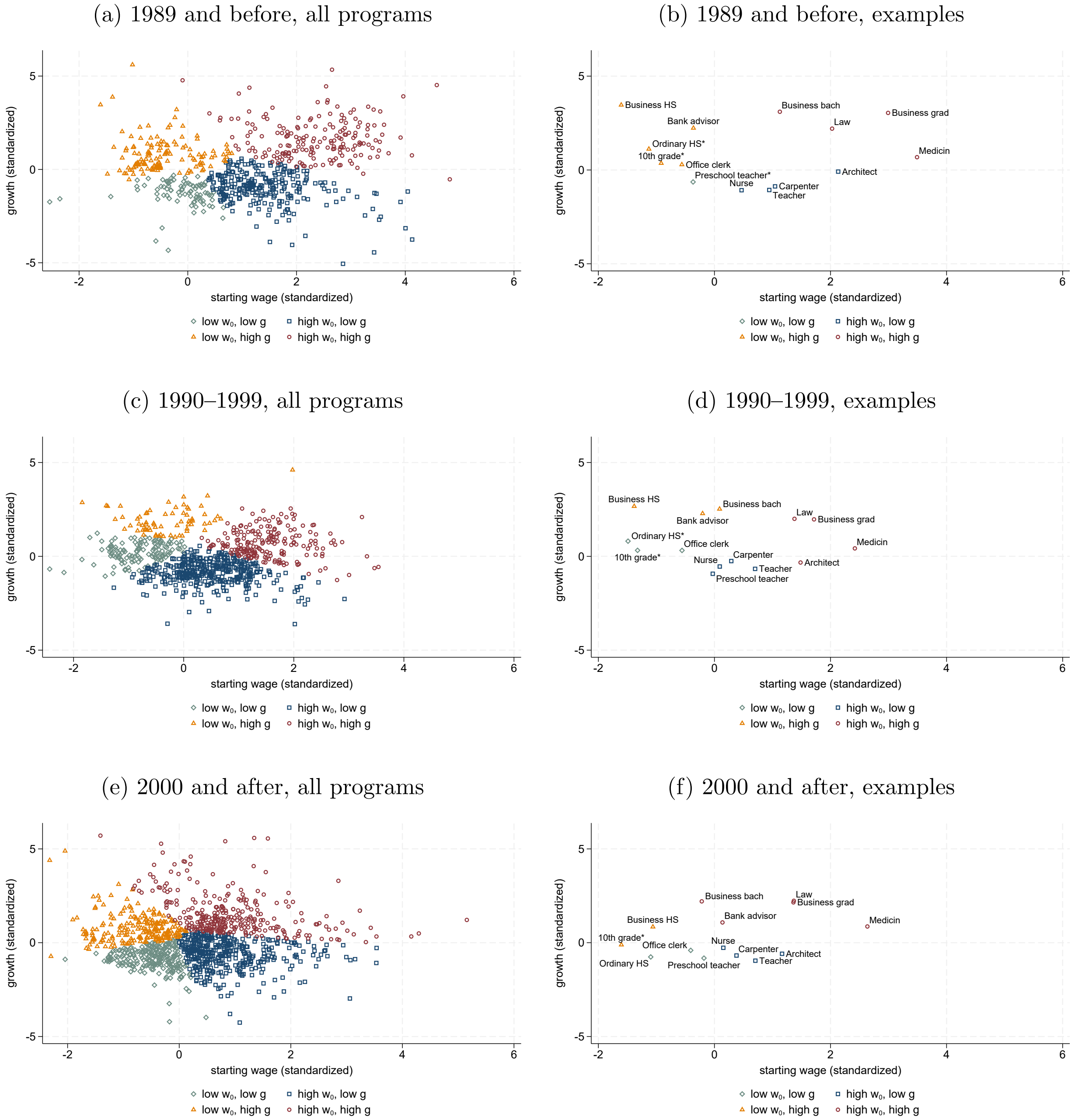
B.3 Supporting figures for robustness analysis

Figure B.6: Educational Program Categorizations by Gender



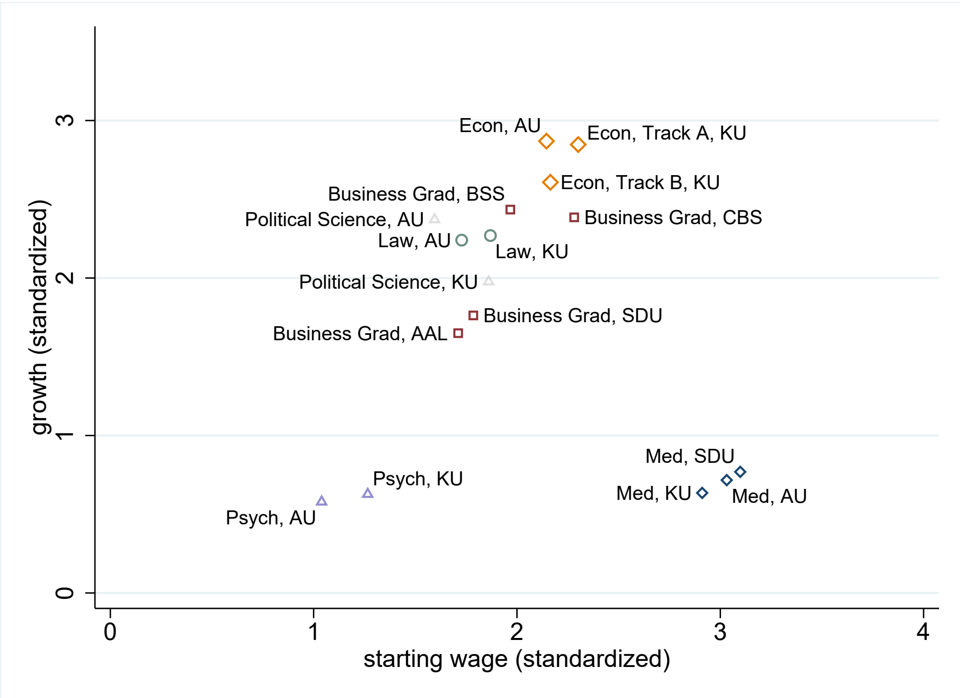
Notes: w_0 stands for *average starting wage* and g for *average wage growth*, defined in Section 3.2. The horizontal axis corresponds to the standardized w_0 and the vertical to the standardized g. Points in the panels locate all educational programs with at least ten graduates in 2018—described in 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 B.7: Educational Program Categorizations by Decade



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 B.8: Starting Wages and Wage Growth for Main Degree Programs by University



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 degree programs by university that have at least 1000 graduates 1980–2018 along these two dimensions. Same symbols and colors indicate the same degree program across different universities.